Research Article

Research on Real-Time Multiple Single Garbage Classification Based on Convolutional Neural Network

Jian-ye Yuan,[1] Xin-yuan Nan [1] Cheng-rong Li,[2] and Le-le Sun[3]

[1]*School of Electrical Engineering, Xinjiang University, Urumqi 830047, China*
[2]*Intelligent Manufacturing Technology and System Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[3]*School of Electrical Engineering, Shandong Huayu University of Technology, Dezhou 253034, China*

Correspondence should be addressed to Xin-yuan Nan; yuanxianshengo@163.com

Considering that the garbage classification is urgent, a 23-layer convolutional neural network (CNN) model is designed in this paper, with the emphasis on the real-time garbage classification, to solve the low accuracy of garbage classification and recycling and difficulty in manual recycling. Firstly, the depthwise separable convolution was used to reduce the Params of the model. Then, the attention mechanism was used to improve the accuracy of the garbage classification model. Finally, the model fine-tuning method was used to further improve the performance of the garbage classification model. Besides, we compared the model with classic image classification models including AlexNet, VGG16, and ResNet18 and lightweight classification models including MobileNetV2 and SuffleNetV2 and found that the model GAF_dense has a higher accuracy rate, fewer Params, and FLOPs. To further check the performance of the model, we tested the CIFAR-10 data set and found the accuracy rates of the model (GAF_dense) are 0.018 and 0.03 higher than ResNet18 and SufflenetV2, respectively. In the ImageNet data set, the accuracy rates of the model (GAF_dense) are 0.225 and 0.146 higher than Resnet18 and SufflenetV2, respectively. Therefore, the garbage classification model proposed in this paper is suitable for garbage classification and other classification tasks to protect the ecological environment, which can be applied to classification tasks such as environmental science, children's education, and environmental protection.

1. Introduction

As early as the 1960s, many countries proposed green treatment schemes to classify garbage and protect the environment as global garbage pollution increased sharply [1]. Among them, Japan was at the forefront of garbage disposal. First, it classified recyclable garbage, and its domestic environment was significantly improved [2]. Considering the tasks in garbage classification and the high labor intensity, low classification accuracy, and poor effects, this paper proposes to use CNN to carry out real-time garbage classification on the conveyor belt to promote garbage recycling and environmental protection. This paper also uses CNN to classify plastic waste based on the research of Bobulski et al. [3, 4]. At the same time, to ensure

the economic value, we classified recyclables into five categories including cardboard, glass, metal, newspaper, and plastic.

CNN was first proposed by LeCun [5] and others. After continuous development, it was used for classification tasks for the first time by KRIZHEVSKY [6] and others with excellent results. Skip connection was proposed by ResNet [7] in 2016 and reached a climax. Based on skip connection, we improved and optimized the design on the DenseNet [8] classification model and used it for real-time garbage classification on the garbage conveyor. We also compared it with the classic image classification networks including AlexNet [6], VGG [9], and ResNet and lightweight classification networks including MobileNetV2 [10] and ShuffleNetV2 [11] in real time. The experimental data show that the

recyclable garbage classification method (garbage classification denseNet, GC-DenseNet) proposed in this paper has good generalization ability, real-time performance, and high accuracy. The network is shown in Figure 1.

2. Algorithm Description

2.1. Depthwise Separable Convolution. Depthwise separable convolution [12] consists of depthwise convolution and pointwise convolution [13]. Depthwise separable convolution was used to replace densenet 3×3 standard convolution to make the number of packets equal to the number of output channels, thereby reducing the number of parameters (Params) and floating point operations (FLOPs) of the model to improve the efficiency and real-time nature of the garbage classification model. The depthwise convolution and standard convolution adopted are shown in Figure 2.

Depthwise separable convolution is used to reduce the number of Params and FLOPs of the model. The formulae of Params and FLOPs are shown in formulae (1) and (2), respectively:

$$Params = C_{in} \times C_{out} \times H_{Kernel} \times W_{Kernel}, \quad (1)$$

$$\begin{aligned} Params = C_{in} \times C_{out} \times H_{Kernel} \times, \\ W_{Kernel} \times H_{out} \times W_{out}, \end{aligned} \quad (2)$$

where H_{out} represents the height of the output feature; W_{out} represents the width of the output feature map; H_{kernel} represents the height of the convolution kernel; W_{kernel} represents the width of the convolution kernel; C_{in} represents the number of input channels, and C_{out} represents the number of output channels.

2.2. Attention Mechanism. The principle of the attention mechanism is to make the model allocate limited resources to the important parts and ignore the unimportant parts, thereby improving the accuracy of the model. At present, the most commonly used attention mechanisms are channel attention mechanism [14] and spatial attention mechanism [15].

The channel attention mechanism makes the model pay more attention to the channel information of the feature map. Firstly, an adaptive maximum pooling and an adaptive average pooling were applied to the feature map. After that, the pooled result was sent to a multilayer perceptron (MLP) [16] for channel splicing, and the weights of the two output feature maps are combined. Finally, we multiply channel-by-channel weighting to the original features on the new channel weights to complete the recalibration of the original picture in the channel of the feature map channel. The network structure of the channel attention mechanism is shown in Figure 3.

The spatial attention mechanism makes the model pay more attention to the spatial information of the feature map. First, the input feature map undergoes an adaptive average pooling and adaptive max pooling in the spatial dimension. Then, a 7×7 convolution kernel is connected to form a new feature map. Finally, the output feature map is obtained by multiplying the attention module feature map by the new feature map through the scale operation. The network structure of the spatial attention mechanism is shown in Figure 4.

This paper combines the features of the channel attention mechanism and the spatial attention mechanism in CBAM [17] and adds the channel and spatial attention mechanisms in series after the 3×3 in DenseNet, to make the network model pay attention to the channel and spatial information while paying attention to the feature map, thereby improving the feature extraction ability of the network model and improving the accuracy of the garbage classification model.

2.3. Model Fine-Tuning. Model fine-tuning is a commonly used method for optimizing network models. It refers to improving the performance of the model by continuously adjusting parameters of the network model. First, because the experiment uses the densenet network model, the pkl parameter file well trained by the garbage is used as the parameter model of transfer learning [18], thereby improving the feature extraction and classification capabilities of the garbage network model to improve the accuracy. After that, since the shallow neural network has a stronger ability to extract garbage information, and to reduce the number of network model parameters to maintain the real-time nature of the garbage classification process, we set the first block of DenseNet to 3, and the others to 1. Finally, given the random garbage information and the unstable pixels of the pictures, to prevent some of the junk data from the explosion during feature extraction, we used hard-Swish [19] instead of RELU [20], which indirectly improved the robustness of the garbage classification model. Among them, hard-Swish is shown as follows:

$$h - swish[x] = x ReLU \frac{(x + 3)}{6}. \quad (3)$$

3. Experiment and Result Analysis

In our experiment, we used the computer of ubuntu20.04 operating system, Intel (R) Core (TM) i-9750 H CPU @ 2.60 GHz 2.59 GHz, and the graphics card model of NVIDIA GeForce RTX 2070 with the memory of 8G. The garbage data set in this experiment was provided by the Institute of Automation, Chinese Academy of Sciences. For each garbage image, the width is 512 pixels, the height is 84 pixels, the horizontal and vertical resolution is 96pi, and the bit depth is 24. DenseNet is used for training. There were 2000 pieces in the training data set, 400 pieces for each category, and DenseNet was used for training. There were 300 pieces in test sets, 60 for each category. To ensure the readability of the experimental data, the following data were all performed with 300 test sets, and only one test set was used. The running time (Time), accuracy rate (Acc, Top1), the number of Params, and FLOPs were used as evaluation indexes.

3.1. Depthwise Separable Convolution. We used this convolution to replace the 3×3 standard convolution in DenseNet121 to reduce the Params and FLOPs of the model. We can see that from Table 1 that, after the depthwise
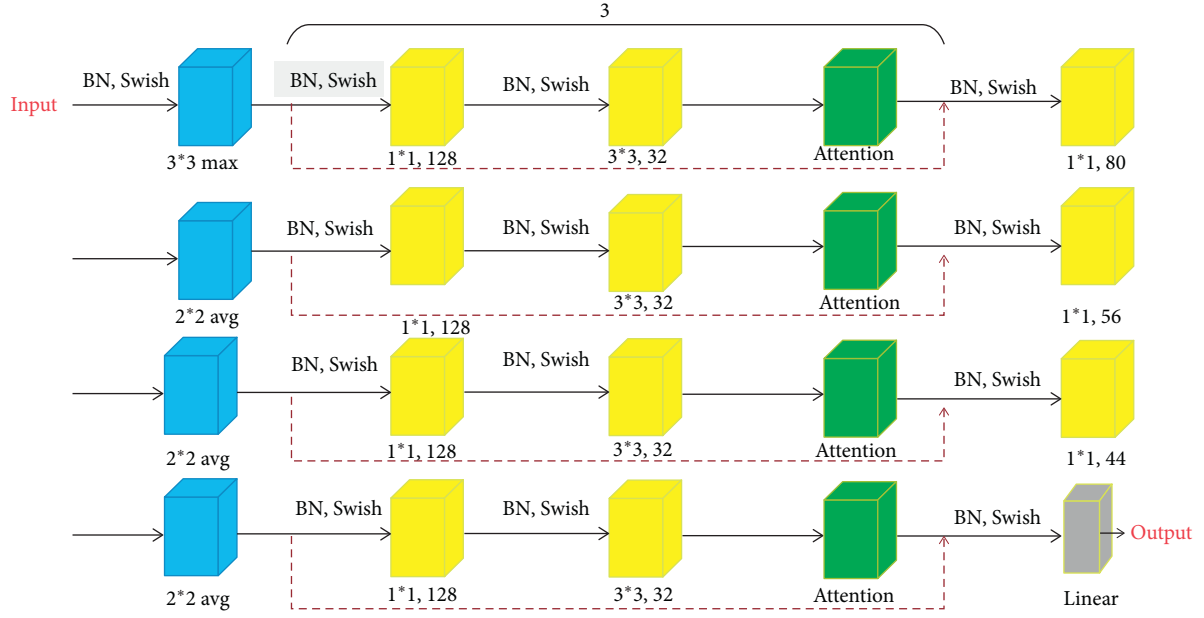
FIGURE 1: GC-DenseNet network model. Note: blue represents the pooling layer; yellow represents the convolution layer; green represents the attention mechanism; and gray represents the full connection layer.
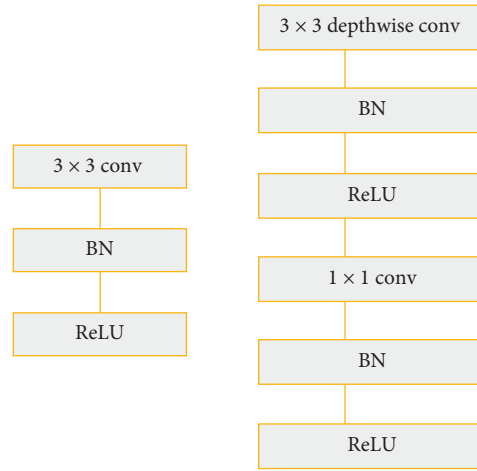

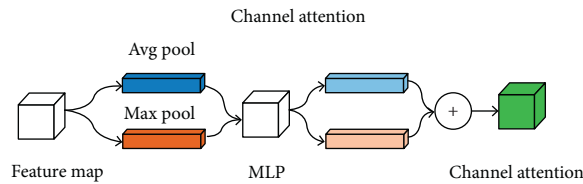FIGURE 2: Depthwise separable convolution and standard convolution.

FIGURE 3: Network structure of channel attention mechanism.

separable convolution (Dense + Group) is used, the accuracy of the garbage classification model is stable, the Params is reduced by 1/3, the FLOPs is reduced by 1.71 times, and the running time is reduced by 3 min, which is conducive to the garbage classification project.

3.2. Attention Mechanism. This paper continues to optimize based on the original and makes the garbage classification model pay more attention to garbage information by attention mechanism to improve the accuracy of the model. From Table 2, we can see that its accuracy has reached
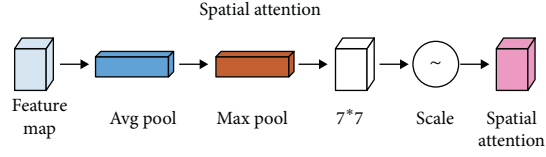
FIGURE 4: Network structure of spatial attention mechanism.

TABLE 1: Depthwise separable data comparison.

Mode	Time	Acc	Params	FLOPs
DenseNet121	56 min48 s	0.884	6.64	2.88
Dense + Group	53 min43 s	0.883	4.66	1.68

TABLE 2: Attention mechanism data.

Mode	Time	Acc	Params	FLOPs
Dense + Group + Atten	55 min12 s	0.922	5.58	2.05

TABLE 3: Data comparison.

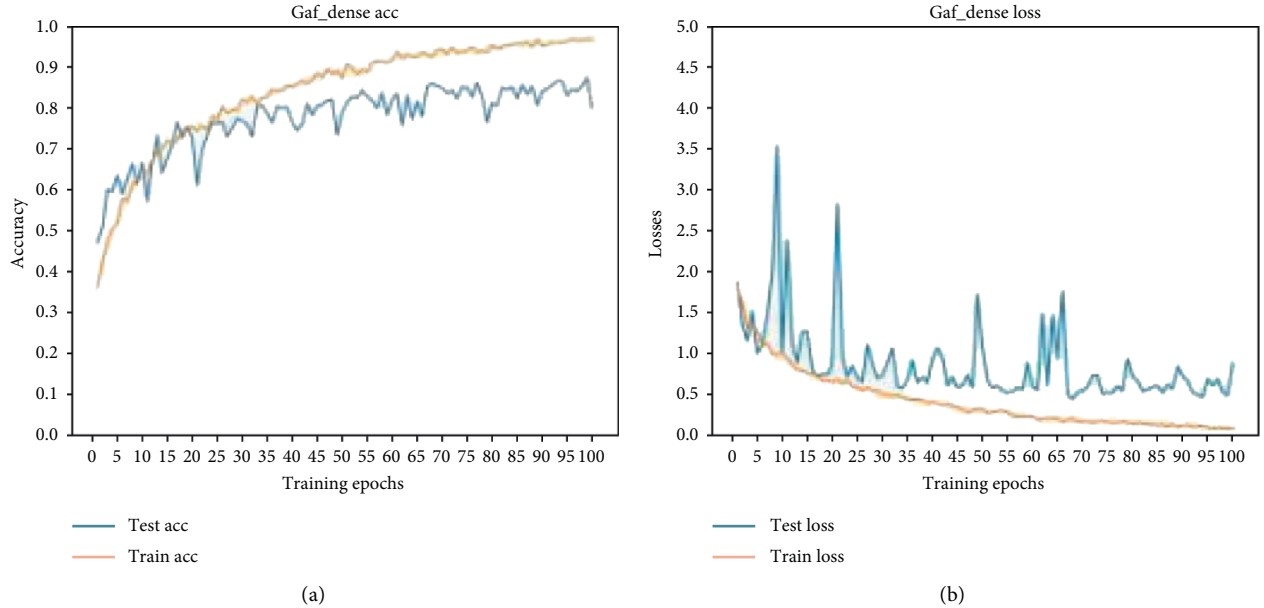| Mode	Time	Acc	Params (M)	FLOPs (G)
AlexNet	19 min 28 s	0.675	54.4	0.71
VGG16	62 min 55 s	0.671	128.1	1.68
ResNet18	26 min 41 s	0.864	10.7	1.82
MobileV2	41 min 60 s	0.857	2.1	0.32
SuffleV2	47 min 11 s	0.874	5.1	0.59
GAF_dense	24 min 41 s	0.921	0.1	0.31

(a)

(b)

FIGURE 5: Curves of loss function and accuracy.

0.9221, although the running time, Params, and FLOPs have been correspondingly improved. So, this garbage classification model is better than DenseNet121 and is preferable for garbage classification tasks.

3.3. Fine-Tuning. To improve the effectiveness of the garbage classification model on the task, we conducted fine-tuning on the model. Then, we compared the fine-tuned model (Dense + Group + Atten + Funing, GAF_dense) with

TABLE 4: CIFAR-10 data comparison.

Mode	Time	Acc
Resnet18	30 min 24 s	0.634
SufflenetV2	28 min 5 s	0.622
GAF_dense	10 min 15 s	0.652

TABLE 5: ImageNet data comparison.

Mode	Acc
Resnet18	0.528
SufflenetV2	0.607
GAF_dense	0.753

AlexNet, VGG16, ResNet18, MobileNetV2, and SufflenetV2. From Table 3, we can see that the FLOPs and Params of GAF_dense are much lower than other models in the same environment. Although the running time is slightly higher than that of AlexNet, AlexNet has a low accuracy rate and is not suitable for garbage classification tasks. At the same time, the accuracy rate is better than other classification network models. We can conclude that the model proposed in this paper is very suitable for garbage classification tasks.

Figure 5 is a curve of the loss function and accuracy during the training and verification of the garbage classification model proposed in this paper. From the graph, we can see that the model converges when the epoch is 100, which is suitable for garbage classification tasks.

To further prove that the classification model has a good generalization capability, we tested it on the CIFAR-10 data set and compared it with the abovementioned well-performed Resnet18 and the lightweight model SufflenetV2. From the data in Table 4, we can see that the model GAF_dense proposed in this paper is superior to Resnet18 and SufflenetV2 in time and accuracy.

At the same time, a test was carried out on ImageNet subdataset [21]. Table 5 shows that the GAF_dense model is still more accurate than Resnet18 and SufflenetV2. Finally, we can conclude that the GAF_dense model not only has a good effect on the garbage classification task but also has a good generalization capability. Especially, it can be transplanted to other classification task models and has a good portability accuracy.

4. Conclusions

This paper designed a model (GAF_dense) on the task of garbage classification and reduced the Params and FLOPs of the network model by using depthwise separable convolution on DenseNet. The attention mechanism was used to improve the accuracy of the classification model, and the model fine-tuning method was used to ensure the real-time nature of the model. We also compared the accuracy of the junk data set, CIFAR-10, and small ImageNet data set with the classic classification model. We compared the real-time nature of garbage classification with the lightweight classification model and found that the garbage classification model proposed in this paper has higher accuracy, better

real-time performance, and generalization capability. We can see that this model is suitable for the classification task on the conveyor belt, which solves the problems of low classification accuracy and high labor intensity in the garbage classification process while protecting the environment and benefiting the country and the people.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study. Some data are internal data of our laboratory, so we cannot share them all at present.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] W. Li and L. Qin, "Green governance: participation, rules, and synergy mechanisms-the experience and enlightenment of Japanese garbage sorting and disposal," *Modern Japanese Economy*, no. 1, pp. 52–67, 2020.

[2] J. Zhong and X. Zhong, "Japanese garbage disposal: policy evolution, influencing factors, and successful experience," *Modern Japan Economy*, no. 1, pp. 68–80, 2020.

[3] J. Bobulski and M. Kubanek, "Waste classification system using image processing and convolutional neural networks," *Advances in Computational Intelligence*, LNCS, vol. 11507, pp. 350–361, 2019.

[4] J. Bobulski and M. Kubanek, "The triple histogram method for garbage classification," in *Proceedings of the 18th International Conference of Numerical Analysis and Applied Mathematics ICNAAM 2019*, Rhodes, Greece, September 2019.

[5] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097–1105, 2012.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, Honolulu, HI, USA, July 2017.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Salt Lake City, UT, USA, June 2018.

[11] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856, Salt Lake City, UT, USA, June 2018.

[12] A. Howard, M. Zhu, B Chen et al., "MobileNets: efficient convolutional neural networks for mobile vision applications," 2017, https://arxiv.org/abs/1704.04861.

[13] W. Wu, Z. Qi, and L. Fuxin, "PointConv: deep convolutional networks on 3D point clouds. in Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition," pp. 9621–9630, Long Beach, CA, USA, 2019.

[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, UT, USA, June 2018.

[15] X. Zhu, D. Cheng, Z. Zhang et al., "An empirical study of spatial attention mechanisms in deep networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6688–6697, Seoul, Korea, November 2019.

[16] F. Rosenblatt, *The Perceptron-a Perceiving and Recognizing Automaton*, Cornell Aeronautical Laboratory Report, Cheektowaga, NY, USA, 1957.

[17] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, Munich, Germany, September 2018.

[18] P. Zhao and S. C. H. Hoi, "OTL: a framework of online transfer learning," in *Proceedings of the ICML 2010-27th International Conference on Machine Learning*, pp. 1231–1238, Haifa, Israel, June 2010.

[19] A. Howard, M. Sandler, C. Bo et al., "Searching for mobilenetv3," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1314–1324, Seoul, Korea, November 2019.

[20] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines vinod nair," in *Proceedings of the 27th International Conference on Machine Learning ICML-10*, vol. 27, pp. 807–814, Haifa, Israel, June 2010.

[21] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.