Research on Lightweight Convolutional Neural Network in Garbage Classification

Yaqing Gu 1, Bing Ge 2,*

1 College of Electrical Engineering, Shandong Huayu University of Technology, Dezhou 253034, China
2 Institute of Information Engineering, Shandong Huayu University of Technology, Dezhou 253034, China

*Corresponding author e-mail: gebing@cnooc.com.cn

Abstract. With the rapid development of society, more and more garbage consumables are produced. How to better recycle and use "garbage" has become a widespread concern. In this paper, a lightweight garbage classification model (GroupAtten-MobileNet, GA_MobileNet) is designed based on the convolutional neural network ResNet-50. Reduce the amount of Params, FLOPs and memory consumption by using deep convolution and group convolution; use the channel attention mechanism to increase the accuracy of the model; Use the model fine-tuning method to further reduce the model's Params, FLOPs, and memory consumption. This paper divides the recyclable garbage into 6 categories. Through an experimental comparison on the garbage data set, it is found that the amount of GA_MobileNet Params proposed in this paper is 280 times less than ResNet-50; the FLOPs amount is reduced by 31.7 times; the accuracy rate increased by about 3.1%; operating time reduced by 50%; At the same time, compared with the same type of lightweight models MobileNetv2, SqueezeNet and shuffleNetV1 also has a clear advantage. Finally, it is concluded that GA_MobileNet is a garbage classification model suitable for lightweight convolutional neural networks, which is suitable for recycling of garbage, protecting the environment and benefiting the country and the people.

Keywords: garbage classification; lightweight; group convolution; attention mechanism; model fine-tuning.

1. Introduction
With the continuous development of society, the amount of garbage generated is also increasing. In traditional machine vision, support vector machine (SVM) is mainly used to classify garbage. With the continuous development of deep learning, convolutional neural networks have been used in image classification for better and better results. At the same time, many garbage classification methods based on machine vision have appeared in garbage classification. For example: the automatic garbage collection device designed by Zhou [1] and others has realized the automatic screening of recyclable garbage; Wang [2] and others designed a compressible garbage classification and recycling bin; Wang [3] and others designed a medical sub-format garbage bin, specifically for medical waste classification.
and recycling; Liang [4] and others invented an automatic sorting sweeper, which realized the functions of sweeping and recycling. In summary, although many garbage classification systems have achieved the effect of garbage classification, but the memory and FLOPs amount occupied by them are large, and they cannot run on embedded devices. Therefore, the model is not suitable for lightweight classification. In this regard, this paper designs and proposes a lightweight network model that can run on embedded devices as a lightweight garbage classification model.

In July 2019, Shanghai issued a compulsory garbage classification policy to classify garbage into dry garbage, wet garbage, recyclable garbage, and toxic and hazardous garbage, which received positive response and attention from all walks of life. This paper classifies six categories of recyclable garbage: cardboard, glass, metal, paper, plastic and cigarette butts, and implements a garbage classification task by building a convolutional neural network.

2. Lightweight Model
Since AlexNet [5] was proposed in 2012, various classification models have been continuously proposed such as: GoogleNet [6], VGG [7], ResNet [8] and mobileNet [9]. Many research efforts are devoted to improving the accuracy of the model on the ImageNet data set, which makes the model's layers continue to deepen and the structure becomes larger and larger, resulting in higher and higher performance requirements for the machine during model training and testing. Therefore, the application is not suitable for automatic driving, mobile phones and small embedded devices. This article is dedicated to the development of a lightweight garbage classification model that can be applied to lightweight devices.

In the process of designing lightweight models, common methods are included: fine model, model pruning[10], network decomposition[11-12], weight sharing[13-14], and model fine-tuning. The lightweight network models proposed by the above methods mainly include mobileNet, ShuffleNet [15] and SqueezeNet [16], where mobileNet uses depthwise separable convolution to reduce the amount of parameters (Params) and floating point of operations (FLOPs); ShuffleNet uses Group Convolution to reduce the amount of Params and FLOPs, and uses channel confusion to improve the accuracy of the model (Accuracy, top1); SqueezeNet proposed to use Fire Module to reduce the amount of Params. In this paper, based on ResNet-50, the fine model and model fine-tuning methods are used to reduce the amount of Params, and use the channel attention mechanism to increase the accuracy of the model.

3. Algorithm Design
3.1. Grouped Convolution
Grouped convolution is to divide the convolution operation into different groups according to the number of convolution channels, which reduces many links in the process of convolution operations. Setting the number of groups to the number of input channels is the depthwise convolution. The running time of the network model affected by the convolution operation is mainly determined by the amount of Params and FLOPs. In general, the amount of Params and FLOPs is inversely proportional to the time required to run the network model. The amount of Params and the amount of FLOPs during the convolution operation are shown in Equation 1 and Equation 2 respectively.

\[
\text{Params} = H_{\text{Kernel}} \times W_{\text{Kernel}} \times C_{\text{in}} \times C_{\text{out}}
\]

\[
\text{FLOPs} = H_{\text{out}} \times W_{\text{out}} \times H_{\text{Kernel}} \times W_{\text{Kernel}} \times C_{\text{in}} \times C_{\text{out}}
\]

In the formula, Hout represents the height of the output feature map; Wout represents the width of the output feature map; HKernel represents the height of the convolution kernel; WKernel represents the width of the convolution kernel; Cin represents the number of input channels; Cout represents the number of output channels.
This paper refers to the lightweight network model MobileNet[7] architecture, optimizes on the basis of the existing ResNet-50 residual block, and uses the grouped convolution[5] to integrate the standard 1×1 volume into 8 groups for volume product; The use of depthwise convolution [17] instead of a 3×3 convolution kernel greatly reduces the amount of Params and FLOPs of the garbage classification model, while keeping the accuracy rate unchanged. The model structure is shown in Figure 1, the dotted line is the skip connection.

![Fig. 1 Group convolution and standard convolution network structure](image)

3.2. Channel Attention Mechanism

CBAM[18] is a lightweight module that calculates feature map information through two dimensions of space and channel. In this paper, in order to improve the accuracy of the garbage classification model while keeping the increase in the amount of Params and FLOPs as small as possible, a channel attention mechanism is added after the ResNet50 residual block. The channel attention mechanism is mainly used to increase the effective channel weight in the feature map and reduce the invalid channel weight, so the model pays more attention to the effective channel, thereby improving the accuracy of the model. The channel attention mechanism is shown in Equation 3.

\[
M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))
\]

In the formula: AvgPool (F) represents the average pooling operation of the Feature Map; MaxPool (F) represents the maximum pooling operation of the Feature Map; MLP() represents the channel combination of AvgPool(F) and MaxPool(F) by using Multi-Layer Perceptron (MLP); σ() Represents the Sigmoid function activation of the combined data; Mc(F) represents the output value through the channel attention mechanism. The network structure of the channel attention mechanism is shown in Figure 2.

![Fig. 2 Network structure of channel attention mechanism](image)
3.3. Model Fine-Tuning
As the model structure becomes deeper, there will be a lot of data redundancy in the model, resulting in a huge waste of computer computing resources and memory. This paper further optimizes and reduces the redundancy by fine-tuning the model. First, this paper sets the number of residual blocks in ResNet-50 to 1, and then uses the activation function $h$-swish\[^{19}\] instead of ReLU to prevent some garbage data from exploding. The activation function $h$-swish is shown in Equation 4, where $h$-swish represents the output value after the activation function.

$$h - swish = x \times \text{ReLU}(x + 3) / 6$$  \hspace{1cm} (4)

After that, the number of model input channels is reduced and designed as [16, 32, 64, 128], while maintaining the accuracy. Among them, the reduction of the number of channels is the phenomenon of redundancy of garbage data found through the experiment. After the optimization in this paper, each residual module in this paper is named GA-Bottleneck, the model structure is shown in Figure 3, and the broken line is the skip connection.

![GA-Bottleneck network structure](image)

Fig. 3 GA-Bottleneck network structure

4. Experimental Process
The deep learning development platform used in the experiment is Pytorch, the experimental training environment is Windows10 64 operating system, Intel(R) Core(TM) i-9750 HCPU@2.60GHz 2.59GHz, the graphics card model is NVIDIA GeForce RTX2070, the video memory is 8G computer to complete training and testing. The data set is provided by the Intelligent Manufacturing Technology and System Research Center of the Institute of Automation, Chinese Academy of Sciences. A total of 2346 pieces of garbage image information have been collected, including 1866 training sets and 480 test sets. The categories of training and test sets are evenly distributed and each The resolution of the picture is 512×384.

First, randomly use image processing methods such as rotation, flipping, zooming, cropping, contrast adjustment, brightness adjustment, and saturation adjustment to enhance the input training pictures. In this experiment, the number of input garbage images (batch size) is set to 16 each time; the initial learning rate is set to 0.001, and the learning rate decayed to 95% every 5 iterations (epoch); the number of training completion iterations is 100. In this paper, there are 480 test sets, and only one test set is actually used. The calculation time is 480 garbage images for testing, which is more conducive to the experiment. In this paper, the Params, FLOPs, multiply-add operation (Madds), accuracy and running time (Time) are selected as the lightweight indicators to measure the model.

1) Group convolution test and data comparison. This paper first uses experiments to use depth convolution (ResNetDepth) in the ResNet-50 residual block. It can be seen from Table 1: that the Params, FLOPs and Madds size of the model are reduced by half, the accuracy rate remains unchanged, and the running time is reduced by about 8min; Then perform group convolution at ResNetDepth, integrate 2
1×1 volumes in ResNet-50 residual block into 8 groups (ResNetGroup8). From the table data, it can be seen that the Params, FLOPs and Madds further decrease, and the Params decreases by 4.5 times. The amount of FLOPs and Madds are reduced by about 3 times. The accuracy is slightly reduced, and the running time is reduced by about 3 minutes. Depthwise convolution is a special form of group convolution. It can be seen that the deep convolution and group convolution used in this paper compress the network model, so that the Params, FLOPs and Madds is significantly reduced, running time is reduced, and lightweight goals are achieved.

Table 1. Group convolution data table

| model            | Params(MB) | FLOPs(GB) | Madds(GB) | Accuracy | Time         |
|------------------|------------|-----------|-----------|----------|--------------|
| ResNet-50        | 22.43      | 4.12      | 8.22      | 0.8528   | 47min27s    |
| ResNetDepth      | 11.67      | 2.28      | 4.54      | 0.8471   | 39min46s    |
| ResNetGroup8     | 3.86       | 0.74      | 1.46      | 0.8405   | 36min49s    |

2) Channel attention test and data comparison. In order to improve the accuracy of the model, this paper continues to optimize on the original basis, adding a channel attention mechanism (ResNetGroup8_Atten) after the ResNet-50 residual block. From the test data table 2, we can see that although ResNetGroup8_Atten has doubled the amount of Params compared to ResNetGroup8, but the FLOPs amount and Madds are basically the same as ResNetGroup8, the accuracy rate is indeed increased by 4% compared to ResNetGroup8. Therefore, the channel attention mechanism increases the accuracy of the model to a certain extent, and is suitable for the garbage classification task in this paper.

Table 2. Channel attention mechanism data table

| model            | Params(MB) | FLOPs(GB) | Madds(GB) | Accuracy |
|------------------|------------|-----------|-----------|----------|
| ResNet-50        | 22.43      | 4.12      | 8.22      | 0.8528   |
| ResNetGroup8     | 3.86       | 0.74      | 1.46      | 0.8405   |
| ResNetGroup8_Atten| 8.66      | 0.75      | 1.47      | 0.8937   |

3) Model fine-tuning test and data comparison. In order to further design a more lightweight garbage classification model, this paper fine-tunes the original model. It can be seen from the data in Table 3 that the Params of the lightweight classification model GroupAtten_MobileNet (GA_MobileNet) proposed in this paper is 280 times less than ResNet-50; the amount of FLOPs is reduced by 31.7 times; Madds is reduced by 30.4 times compared to ResNet-50; the accuracy is improved about 3.1%; the running time is reduced by 50%. In order to further observe the network model proposed in this paper, this paper compares the loss function (Loss) and finds that GA_MobileNet has a loss of 0.9 compared to ResNet-50.

Compared with the lightweight models MobileNetv2[20], SqueezeNet and shuffleNetV1, it can be seen that the amount of Params, FLOPs, Madds and running time are lower than other lightweight models. Although the running time of SqueezeNet is lower than GA_MobileNet, its accuracy is much lower than GA_MobileNet, which is not suitable for the classification of garbage images; The accuracy of MobileNetv2 is comparable to GA_MobileNet, but it has a long running time and is not suitable for lightweight garbage classification tasks. Therefore, the lightweight classification model GA_MobileNet proposed in this paper is suitable for garbage classification tasks.

Table 3. Data comparison

| model            | Params(MB) | FLOPs(GB) | Madds(GB) | Accuracy | Time         | Loss  |
|------------------|------------|-----------|-----------|----------|--------------|-------|
| ResNet-50        | 22.43      | 4.12      | 8.22      | 0.8528   | 47min27s    | 1.3487|
| ResNetGroup8_Atten| 8.66      | 0.75      | 1.47      | 0.8937   | 47min14s    | 0.5109|
| MobileNetv2      | 2.13       | 0.32      | 0.63      | 0.8753   | 33min35s    | 0.7208|
| SqueezeNet       | 0.69       | 0.27      | 0.53      | 0.4479   | 21min41s    | 1.4112|
| shuffleNetV1     | 3.38       | 0.54      | 1.07      | 0.8589   | 34min5s     | 0.7606|
| GA_MobileNet     | 0.08       | 0.13      | 0.27      | 0.8834   | 25min48s    | 0.4408|
4) Training and loss value curve. Figure 5 is a graph of the accuracy and loss of classification training and test sets, of which Figure 4 (a) (b) is the accuracy and loss curve of ResNet-50 in the garbage classification task; (c) (d) is the graph of the accuracy and loss of GA_MobileNet in the garbage classification task proposed in this paper. The comparison between the accuracy curve and the loss curve shows that GA_MobileNet is significantly better than ResNet-50 in the first 10 epochs. At the same time, GA_MobileNet is better in terms of stability, and ResNet-50 is more volatile, but in terms of accuracy, it can also be approximated that GA_MobileNet is higher than ResNet-50. Therefore, GA_MobileNet proposed in this paper is more stable, more robust and more accurate than ResNet-50.

(a) ResNet-50 accuracy curve    (b) ResNet-50 loss value curve

(c) GA_MobileNet accuracy curve    (d) GA_MobileNet loss curve

Fig. 4 Training, test accuracy and loss curve

5) Model structure. Table 4 shows the structure of the GA_MobileNet garbage classification model proposed in this paper. It can be seen from the table that there are 6 modules in GA_MobileNet. Since a GA-Bottleneck5 layer is added with ConV1 and ConV6 at the same time, the garbage classification network model GA_MobileNet proposed in this paper has a total of 22 layers.

| ConV   | LayerName             | OutputSize | Params (85502) | FLOPs (134385276) | Madds (267430394) |
|--------|-----------------------|------------|----------------|------------------|------------------|
| ConV1  | ConV/BN/h-swith       | 112×112    | 9536           | 119619584        | 238436352        |
| ConV2  | 3×3, MaxPool, stride=2| 56×56      | 2728           | 8555268          | 16809408         |
| ConV3  | GA-Bottleneck1        | 28×28      | 3952           | 2296584          | 4455040          |
| ConV4  | GA-Bottleneck2        | 14×14      | 14816          | 2105232          | 4141184          |
| ConV5  | GA-Bottleneck3        | 7×7        | 52928          | 1807072          | 3585344          |
| ConV6  | AdpAvgPool, FC, SoftMax, Linear | 1×1 | 1,542 | 1,536 | 3,066 |
5. Conclusion
This paper designed a lightweight garbage classification model for the problem of recyclable garbage classification, optimized on the existing ResNet-50, using depth convolution and grouping convolution to reduce the amount of calculation and Params, thereby reducing memory consumption and running time; use the channel attention mechanism to improve the accuracy of the model; use the model fine-tuning method to make the model classification accuracy higher and occupy less memory consumption. Therefore, the garbage classification model GA_MobileNet proposed in this paper has the characteristics of small FLOPs, small Params, memory occupation and short running time. Compared with the lightweight models MobileNetv2, SqueezeNet and shuffleNetV1, the GA_MobileNet Params, FLOPs, Madds and running time are lower than other lightweight models. Finally, it is concluded that GA_MobileNet is a lightweight garbage classification model, which solves the application problem of garbage classification on embedded devices and classifies the garbage to protect the ecological environment.

Acknowledgements
The authors would like to acknowledge the financial support provided by Shandong Higher Education Undergraduate Education Reform Project "Research and Practice of Automation Major Upgrading and Transformation to Serve the Transformation of Old and New Driving Forces" (Grant No. Z2018S027).

References
[1] Y.M.ZHOU. Design and Implementation of Intelligent Sorting System for Domestic Waste Based on Machine Vision [D]. Harbin Institute of Technology, 2018.
[2] Z.N.WANG, Zhang P.Y.ZHANG, Zhao Y.ZHAO. A sortable recycling bin for compressible garbage [P]. Tianjin: CN111115059A, 2020-5-8.
[3] C.M.WANG, D.N.WU, Y.W.LIU. Design and application of a medical format trash can[J]. Chinese Journal of Modern Nursing, 2017, 23(15): 2059.
[4] J.F.LIANG, Liang J.H.LIANG. An automatic garbage sweeper [P]. Zhejiang Province: CN110863465A, 2020-3-6.
[5] KRIZHEVSKY, A; SUTSKEVER, I; HINTON, G. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems. 2012. 25(2). 1097-1105.
[6] Szegedy C,Liu W,Jia Y,et al.Going Deeper with Convolutions [J].Proceedings of the IEEE conference on computer vision and pattern recognition.2014:1-9.
[7] SIMONYAN K; ZISSERMAN,Andrew. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:2014:1409.
[8] HE, K, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision & pattern recognition.2016:770-778.
[9] HOWARD, A., et al. Mobilenet: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.2017:04861.
[10] HAN, S, et al. Learning both weights and connections for efficient neural network[J].Advances in neural information processing systems. 2015:1135-1143.
[11] DENTON E, ZAREMBA W, BRUNA J, et al. Exploiting Linear Structure within Convolutional Networks for Efficient Evaluation[C]//Proc of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2014:1269-1277.
[12] SIRONI A, TEKIN B, RIGAMONTI R, et al. Learning Separable Filters[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1):94-116.
[13] HAN S, MAO H, DALLYW J. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding [EB/ OL]. [2018-10-24] https://arxiv.org/pdf/1510.00149.pdf.
[14] CHEN W L, WILSON J, TYREE S, et al.Compressing Neural Networks with the Hashing Trick[C]//Proc of the 32nd International Conference on Machine Learning. Berlin, Germany:
[15] ZHANG, Xiangyu, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[J]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:6848-6856.

[16] IANDOLA, Forrest N., et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv.2016:1602.

[17] CHOLLET, F. Xception: Deep learning with depthwise separable convolutions [J]. Proceedings of the IEEE conference on computer vision and pattern recognition.2017:1251-1258.

[18] WOO S, et al. Cbam: Convolutional block attention module[J]. Proceedings of the European conference on computer vision (ECCV). 2018:3-19.

[19] HOWARD, A, et al. Searching for mobilenetv3[J]. Proceedings of the IEEE International Conference on Computer Vision. 2019:1314-1324.

[20] Sandler, M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:4510-4520.