Received 26 February 2024, accepted 21 March 2024, date of publication 25 March 2024, date of current version 29 March 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3381533

RESEARCH ARTICLE

Garbage Classification Algorithm Based on Improved MobileNetV3

XUEYONG TIAN, LIWEI SHI, YUANQING LUO, AND XINLONG ZHANG
College of Environmental and Chemical Engineering, Shenyang University of Technology, Shenyang 110870, China
Corresponding author: Xueyong Tian (tianxueyong@sut.edu.cn)

This work was supported by Liaoning Provincial Science and Technology Tackling Program under Grant 2021JH1/10400031.

ABSTRACT In recent years, the amount of household waste has increased sharply, and there is an urgent need to use intelligent waste classification equipment to assist in completing waste classification tasks. However, existing garbage classification algorithms have large parameter counts and poor real-time performance, which are not suitable for embedded garbage classification devices. Therefore, in order to achieve a lightweight and efficient classification model, this article uses a self-built garbage dataset and a pre-trained MobileNetV3 model in the PyTorch framework for garbage recognition and classification. We introduce the CBAM attention mechanism in the network model to enhance spatial feature perception. Utilize the Mish activation function to fully utilize the extracted depth image information. Using global average pooling instead of the fully connected layer of the original model reduces the number of model parameters while improving the recognition accuracy of the model. Finally, we propose an improved lightweight model called GMC-MobileNetV3. The experimental results show that the recognition accuracy of the improved MobileNetV3 model on self-built data set reaches 96.55%, which is 3.6% higher than that of the original model, the number of model parameters is 0.64M, the memory resource consumption is reduced by 56.6%, and the recognition time of a single garbage image is only 26.4ms. The network proposed in this article can achieve low consumption and high accuracy in garbage recognition and classification, providing reference for future academic research and engineering practice.

INDEX TERMS Classification algorithms, convolutional networks, deep learning, image recognition, transfer learning.

I. INTRODUCTION

In recent years, with the development of urbanization and the improvement of people's living standards, the amount of domestic waste has increased sharply. According to the data of the National Bureau of Statistics of China, the amount of national domestic waste removal reached 248.692 million tons in 2021 [1], which increased by 51.6% compared with 2012, so how to properly dispose of domestic waste has become an urgent problem to be solved. At present, waste incineration is the main way of waste treatment, and waste classification before incineration is an important means to achieve resource utilization. The traditional way of garbage sorting is to collect domestic garbage to the garbage treatment plant first, and then sort it manually, with staff standing on both sides of the conveyor belt and grabbing the garbage with their hands or tools to complete the sorting. However, this way of garbage sorting has high manpower costs and low sorting efficiency, and the pollution and odor from the garbage treatment plant may have a certain impact on the health of garbage-sorting staff [2]. If sorting can be completed when the garbage is put out and recycled separately according to different types, the pressure of post-recycling treatment can be greatly reduced.

With the development of artificial intelligence technology, machine vision based intelligent garbage classification and recycling bins and intelligent garbage bins [3] have entered the public's vision. The core is deep learning algorithms, and the quality of recognition algorithms directly determines the

The associate editor coordinating the review of this manuscript and approving it for publication was Zeev Zalevsky.

classification effect of the equipment. Due to the excellent performance of convolutional neural networks in the field of computer vision, more and more scholars use them to carry out research on garbage classification and have made some progress. Currently, most research uses large-scale convolutional neural network models, such as VGG [4], ResNet [5], DenseNet [6], and InceptionNet [7]. However, large-scale models contain a large number of parameters, which places high demands on the computational capabilities of garbage classification equipment. Limited by the level of hardware, these devices not only have a large size, but also have a high cost. Therefore, for embedded devices or platforms with limited computing resources, researchers are exploring the use of lightweight CNN(Convolutional Neural Networks) models for garbage image classification, such as MobileNet [8], [9], [10], ShuffleNet [11], and EfficientNet [12]. The lightweight CNN model has fewer parameters and higher efficiency, making it very suitable for small devices. However, due to the significant reduction in parameter count, the lightweight CNN model is prone to low recognition accuracy when directly used for garbage sorting work. Therefore, adjustments and improvements need to be made for specific usage scenarios.

This study further explores low-cost and high-precision CNN models based on lightweight CNN, and proposes an improved MobileNetV3 network, namely GMC-MobileNetV3, to achieve better performance in garbage classification tasks. The main contributions of this article include:

(1) A dataset including four common types of household waste was constructed to fine tune the pre trained MobileNetV3 model and improve its recognition performance in garbage classification tasks.
(2) CBAM (Convolutional Block Attention Module) is used instead of SE-Net to enhance the model's spatial perception of features, so that it can adaptive emphasize and suppress different feature information according to the spatial distribution of feature maps.
(3) Mish activation function is used in the convolution layer to improve the generalization performance and information representation ability of deep networks.
(4) The classifier chooses global average pooling instead of full connection layer to reduce the number of parameters in the model and mitigate overfitting.

The remainder of this paper is structured as follows: In the second section, the dataset used in the experiment is first introduced, followed by the principle and improvement methods of MobileNetV3, and an improved model is constructed. Finally, the training method of transfer learning and the evaluation indicators of the model are introduced. The third section presents the comparison results between our algorithm and other mainstream algorithms, and analyzes the improvement effect of our algorithm on the original algorithm. The fourth section elaborates on the main conclusions and presents prospects.

II. MODELS AND TRAINING METHODS

A. DATA

The original MobileNet pre-trained model is trained on the ImageNet dataset [13], but the data in ImageNet does not fully contain the images we need, which is difficult to ensure the efficiency of transfer learning. Therefore, we need a dataset containing garbage images to fine tune the model. At present, there is no standard dataset for garbage classification tasks. Yang and Thung created the TrashNet [14] dataset for garbage classification, but it contains very few categories, which is not in line with the actual situation of domestic waste classification in China. Therefore, this paper constructs a data set specially used for visual garbage sorting through network retrieval and laboratory real scene shooting, including multiple scenes such as single object, multiple similar objects, complex background, and various interference situations such as lighting and motion blur. It is divided into four categories: kitchen waste, recyclable waste, hazardous waste, and other waste, with a total of 4152 images in JPG format. The specific quantity is shown in Table 1. In order to expand the number of samples and enhance the diversity of samples, data enhancement [15] was carried out on the experimental data.

TABLE 1. Self built garbage dataset distribution table.

Types of garbage	Training set	Validation set	Test set
food waste	952	190	95
recyclable waste	1750	350	175
hazardous waste	924	184	92
other waste	526	104	52

Before training the model using image data, we use image processing algorithms to perform image transformation operations such as random rotation, random resizedcrop, contrast transformation and Gaussian Blur on the image data. As shown in Figure 1, the data enhancement can generate mineral water bottle images in four different states, so that the training data is as close as possible to the real distribution of data, thus improving the generalization ability and robustness of the model. We divide the dataset into training set, validation set and test set in 7:2:1. The training set is used to train the model's parameters, the validation set is used to evaluate the model's performance and adjust hyperparameters during the training process, and the test set is used to evaluate the model's performance and generalization ability.

FIGURE 1. Example of data enhanced images.

B. IMPROVEMENTS TO THE MobileNetV3 MODEL

The MobileNet [16] network utilizes depthwise separable convolution to lightweight convolutional layers. Depthwise separable convolution is composed of depthwise convolution and pointwise convolution. Unlike standard convolution, depthwise convolution performs independent convolution operations on each input channel, with the same number of convolution kernels as the input channel. Pointwise convolution is a convolution operation between the feature graph obtained by deep convolution and a 1 × 1 convolution kernel for linear combination and fusion of feature graphs of different channels. Compared with standard convolution, depthwise separable convolution significantly reduces the number of model parameters. The comparison of their computational complexity is as follows:

$$\frac{P_1}{P_2} = \frac{D_K \times D_K \times D_F \times D_F \times M + M \times N \times D_F \times D_F}{D_K \times D_K \times D_F \times D_F \times M \times N} = \frac{1}{N} + \frac{1}{D_K^2} \tag{1}$$

where P_1 and P_2 are the computations of depthwise separable convolution and standard convolution respectively, D_K is the size of the convolution kernel, D_F is the size of the feature map, N is the number of channels of the feature map and convolution kernel, and M is the number of convolution kernels.

In the MobileNetV3 model, convolution kernels of size 3 × 3 and 5 × 5 are used. From the above formula, we can see that the number of parameters of depth-separable convolution is about 1/9 and 1/25 of the standard convolution. Overall, MobileNetV3 [17] networks have fewer parameters compared to large networks, which can significantly reduce model training time and provide direction for optimizing subsequent network structures. However, the MobileNetV3 network still has the problem of low model classification accuracy due to insufficient image feature information extraction ability, and the excessive number of parameters in the classification layer affects inference speed. To solve these problems, we optimize the MobileNetV3 model in three aspects: the attention mechanism, the activation function, and the classification layer structure.

1) CBAM ATTENTIONAL MECHANISMS

The CBAM [18] module combines the channel attention mechanism and the spatial attention mechanism. Given an intermediate feature map, it will sequentially infer the attention weights along two separate dimensions, channel and spatial, and then multiply the attention weights by the input feature maps for adaptive feature tuning, thus making the convolutional neural network pay attention to more important information. Compared with SE-Net, CBAM not only considers the relationship and importance between different channels, but also the relationship between different spatial positions in the feature map, so that it can better capture the correlation and context information in the feature map. As a lightweight module, the model performance can be improved appreciably with an increase in the number of slight parameters, and it can be easily embedded into existing network architectures. Therefore, this paper adopts the CBAM attention module as an alternative to the SE-Net module in the MobileNetV3 network.

2) MISH FUNCTION

The Mish function is a self-regularizing non-monotonic activation function which has the expression:

$$Mish(x) = x \times \tanh(\ln(1 + e^x)) \tag{2}$$

where $Mish(x)$ is the output value of the activation function. x is the input value and can be viewed as a weighted sum of neurons.

Mish is derivable in all real number ranges and has continuous first-order derivatives, avoiding the problem of ReLU neuron death and helping to better handle gradients. At the same time, the Mish activation function can converge faster on some complex tasks, and as the network layers deepen, the accuracy of ReLU and Swish decreases rapidly, while Mish can better maintain accuracy [19]. Therefore, in this study, the Mish function is used instead of the ReLU function, and the H-Swish function of the deeper network is replaced by the Mish function to improve the generalization ability of the model.

3) GLOBAL AVERAGE POOLING

Global Average [20] Pooling layer is a network structure used for pooling, For the input feature map, the pixel values of each channel are averaged to obtain a single value, which is used to represent the corresponding feature map. It can efficiently reduce the number of parameters of the model and reduce the overfitting phenomenon, and it can integrate the global spatial information to make the model more robust to the spatial translation of the input image. As shown in Figure 2, this study moved the global average pooling layer of the original model's classification layer from the middle to the last. After optimization, the number of channels in the feature map input to the classification layer is reduced to the number of categories through two convolutional layers. After passing through the global average pooling layer, a feature vector with a dimension equal to the number of categories is obtained, and then directly input to the Softmax layer for final classification. This structure strengthens the connection between the feature map and categories, improving the model's expressive ability.

4) STRUCTURE OF THE GMC-MobileNetV3 MODEL

The network structure of GMC-MobileNetV3 is shown in Figure 3, Bneck_s1 represents a convolutional layer with a stride of 1, Bneck_s2 represents a convolutional layer with a stride of 2, h represents the activation function used in the convolutional layer is H-Swish, and m represents that the activation function is Mish. When the step size is 1 and 2, use two different convolutional layer structures. As shown in Figure 4, when the step size is 1, the convolutional layer consists of DW Convolution (Depthwise Convolution) and two PW Convolution (Pointwise Convolution). When the step

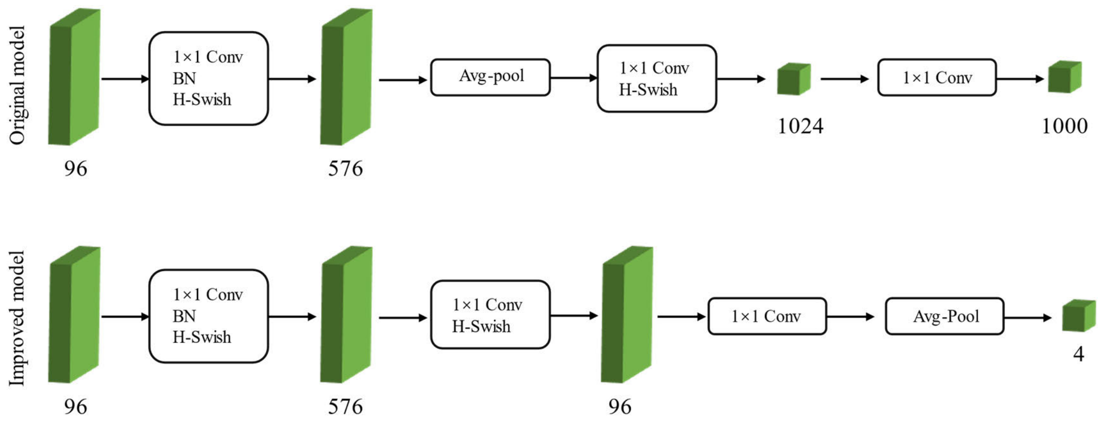


FIGURE 2. Network structure of improved model classification layer.

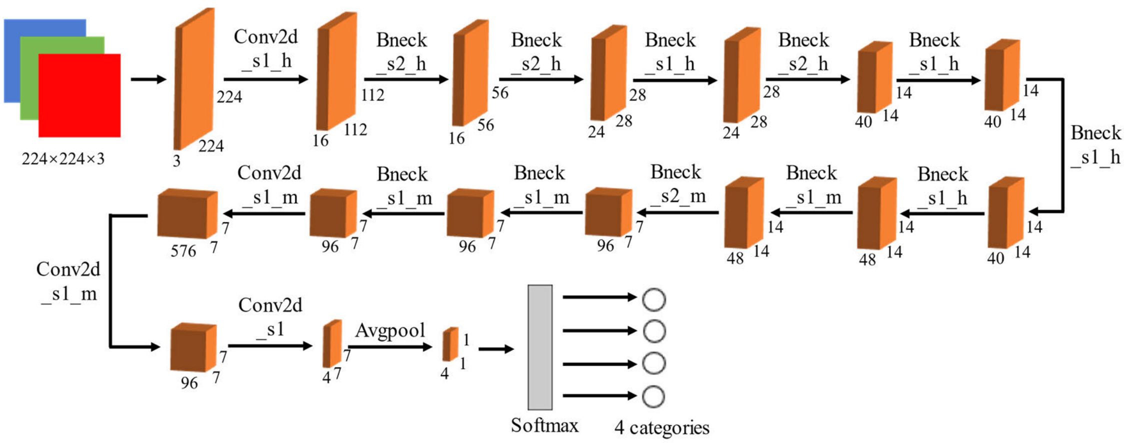


FIGURE 3. Overall structure diagram of GMC-MobileNetV3.

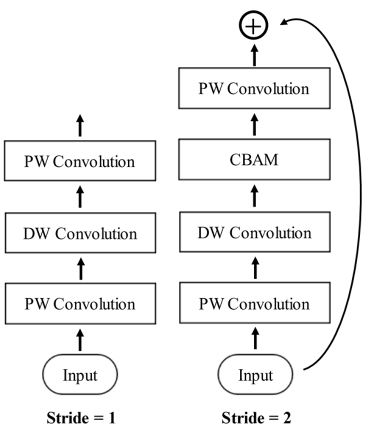


FIGURE 4. Bottleneck structure under two strides.

size is 2, a CBAM attention mechanism is added after the DW Convolution, and the convolutional layer module uses residual connections.

C. TRANSFER LEARNING METHODS

In the field of garbage recognition, there is no dataset dedicated to training garbage classification models, and it is difficult to train a deep learning network using a self-built dataset with small samples. Therefore, the GMC-MobileNetV3 model proposed in this paper combines the transfer learning method with the MobileNetV3 network to further improve the classification accuracy and generalization ability of the network.

Transfer Learning [21] refers to the application of already learned knowledge or experience from one task to another to facilitate machine learning. The use of transfer learning can help to train an accurate and generalization capable model on a limited dataset [22].

Firstly, we obtained a pre-trained model of MobileNetV3 in the PyTorch framework and made improvements on this model to obtain our GMC-MobileNetV3 model. At this point, the parameters and weights of the model were pre-trained. Then, we trained it using a self built garbage dataset and fine-tuned [23] the weights and parameters of the model. In the process of transfer learning, we froze the underlying network used to extract common features, so that its parameters and weights are not updated during the training process. Then, we retrained the top-level network using garbage images and removed the output layer, so that the entire network can be regarded as a feature extractor. Finally, the extracted features are input into an improved classifier to achieve accurate

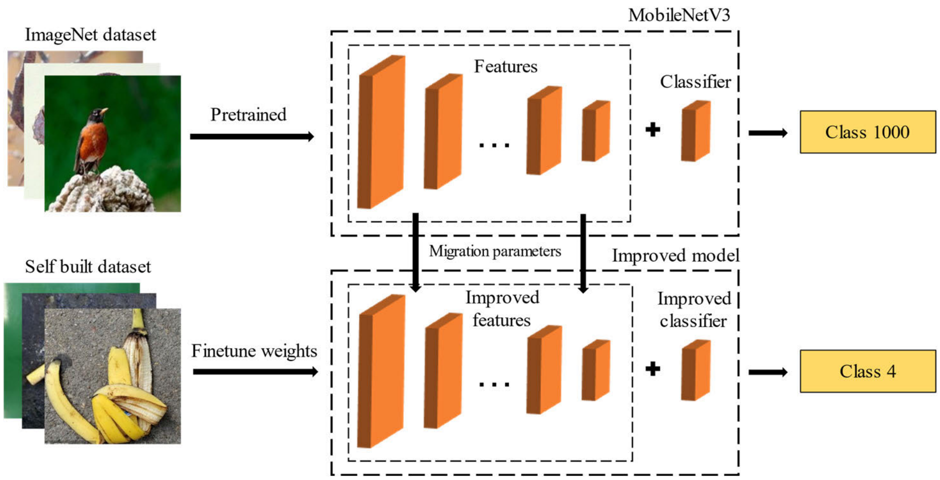


FIGURE 5. Flowchart of transfer learning.

prediction of garbage images, The process of transfer learning for the model is shown in Figure 5.

D. EXPERIMENTAL IMPLEMENTS

After building the GMC MobileNetV3 model, we trained it using the PyTorch deep learning framework. The hardware configuration used in the experiment is Intel(R) Xeon(R) Silver 4314 CPU @ 2.40GHZ 3.40GHZ processor; 128GB RAM; NVIDIA A30 graphics card. The operating system used for the experiments is Windows Server 2019 Standard. 50 rounds of training are used for each experiment, and the Batch Size is set to 32. The learning rate is reduced to 10% of the previous rate every 5 rounds using the auto-adjustment strategy. The optimizer uses Adam optimization algorithm.

E. EVALUATION INDICATORS

According to the experimental purpose, this model adopts the number of parameters (P), model size (S) and single image detection time (T) as the indexes for evaluating the performance of the model, and evaluates the model's effect of recognizing the garbage by Precision, Recall and F1-Score, and the F1-Score integrally takes into account the two indexes of Precision and Recall, which can provide a more comprehensive evaluation for this model. The expressions of Precision, Recall and F1-Score are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{5}$$

where TP refers to the number of positive samples that the model correctly predicts as positive; FP refers to the number of negative samples that the model incorrectly predicts as positive; FN refers to the number of positive samples that the model incorrectly predicts as negative.

III. EXPERIMENT AND ANALYSIS

A. RESULTS AND DISCUSSION

1) PERFORMANCE EVALUATION OF THE GMC-MobileNetV3 MODEL

We compare the GMC-MobileNetV3 model with the ResNet50, VGG16, and ShuffleNetV2 models to evaluate the performance of the models performed in the garbage classification task. They all use self built datasets for training and testing. As shown in Figure 6, we are able to find that the F1-Score of the GMC-MobileNetV3 model test is much higher than the other models, and the accuracy improvement is smoother, which indicates that the model has better stability in recognition and classification. At the same time, it can be seen that the loss rate of GMC-MobileNetV3 is also lower than the other three models.

Table 2 shows the model parameters and garbage image recognition performance of the four transfer learning models. From Table 2, it can be seen that the classification accuracy, recall and F1-Score of the GMC-MobileNetV3 model are higher than those of the two large-scale models, VGG16 and ResNet50, and the number of model parameters and the model size are only 2.72% and 28.7% of ResNet50, 0.48% and 0.5% of VGG16, so that the recognition time of a single image is also shortened to 27.6% of ResNet50 and 16.2% of VGG16, respectively. Compared with ShuffleNetV2, which is also a lightweight network, it can achieve a smaller number of parameters and model size with higher recognition accuracy, and a shorter recognition time, so the improved MobileNetV3 model is more suitable for embedded devices with limited computing power to carry out garbage sorting tasks.

The comparison between GMC-MobileNetV3 and other improved models is shown in Table 3, and the testing of these models was completed on the TrashNet public dataset. The methods used for comparison include large networks such as ResNet50, DenseNet121, VGG19, and lightweight networks such as MobileNet and Xception. From the table,

TABLE 2. Comparison of test results for each model in self-built dataset.

Model	Precision/%	Recall/%	F1-Score/%	P/M	S/MB	T/ms
ResNet50	92.91	93.36	93.05	23.52	90	95.7
VGG16	92.27	94.14	93.14	134.28	512	162.6
ShuffleNetV2	92.78	92.18	92.48	5.35	20.6	35.8
GMC-MobileNetV3	96.89	96.24	96.55	0.64	2.58	26.4

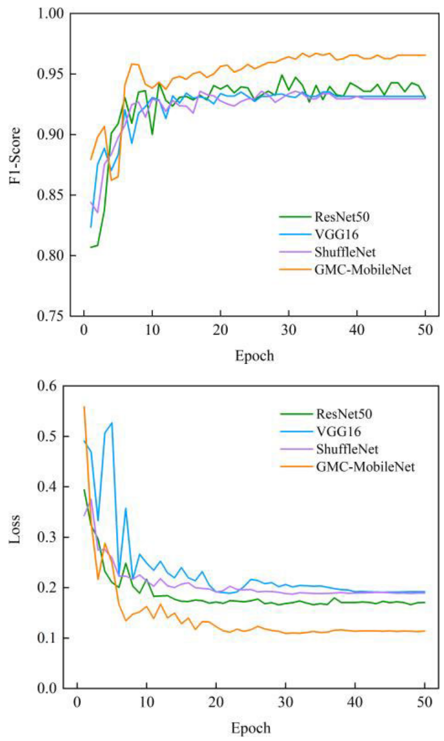


FIGURE 6. Test curves for each model in the self-built dataset.

TABLE 3. Comparison of GMC-MobileNetV3 with existing improved networks.

Methodology	Precision/%	References
GMC-MobileNetV3(our proposed)	96.31	
WasNet	96.10	[3]
DenseNet121	95.00	[24]
M-b Xception	93.25	[25]
ResNet50	92.08	[5]
Inception-ResNet	88.66	[26]
OscarNet (based on VGG-19)	88.42	[27]
MobileNet	87.20	[28]

it can be seen that our proposed GMC-MobileNetV3 network has better recognition performance on the TrashNet dataset than existing improved networks. Only the WasNet network proposed by Yang et al. is close to our GMC-MobileNetV3 network, but as a lightweight network, the number of parameters of GMC-MobileNetV3 is only 40% of WasNet.

2) ANALYSIS OF THE GMC-MobileNetV3 MODEL

In order to evaluate the extent to which each improvement point of the model contributes to the model performance improvement, ablation experiments were performed on the model. The experiment includes Mish activation function, CBAM attention mechanism and improved classification layer, starting from the MobileNetV3 model of transfer learning, and the improvement points were added sequentially, and at the same time, the model's accuracy, recall, F1-Score, and model size metrics were recorded. The Top-1 accuracy of the model is used as the performance metric for this experiment. As shown in Figure 7, the MobileNetV3 curve represents the recognition curve of the MobileNetV3 model of transfer learning, the curve marked by RLS represents the recognition curve after the model changes the classification layer, Mish represents the change in model recognition rate after using the Mish activation function on the basis of RLS-Mobilenetv3, and CBAM represents that the model further changes the attention mechanism. It can be seen that the recognition accuracy of the improved model is improved when the number of training rounds is the same.

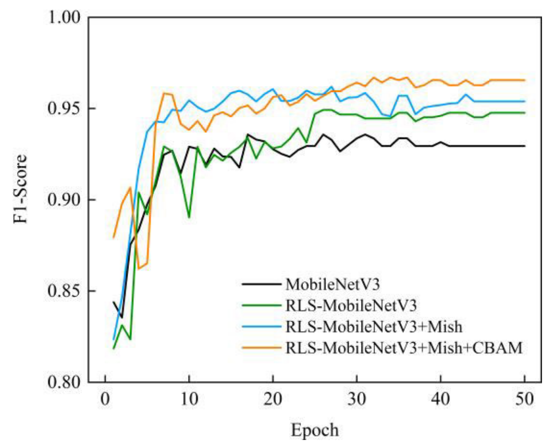


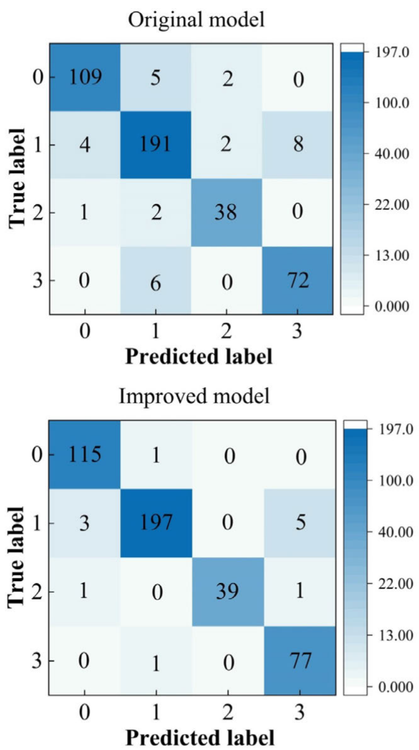
FIGURE 7. Comparison curves for ablation experiments.

As can be seen from Table 4, the F1-Score is improved by 1.81% in the case that the model changes the classification layer, and at the same time, the number of parameters and the size of the model are reduced by 35.5% and 34.3%, respectively, and the recognition time of a single image is shortened by 2.3ms. Using Mish activation function in the deep network can improve the generalization ability of the convolutional layer and increase the F1-Score of the model by 0.63%, but it will slightly increase the amount of computation. We increase the reduce divider of CBAM attention mechanism in the model, so it will reduce the size of the model by 33.8%, while the F1-Score of the model will be improved by 1.16% to 96.55%. Eventually, compared with the original model, the improved MobileNetV3 model's F1-Score is improved

TABLE 4. Comparison of results of ablation experiments.

Model	Precision/%	Recall/%	F1-score/%	P/m	S/mb	T/ms
Mobilenetv3-transfer	93.24	92.70	92.95	1.52	5.94	22.9
RLS-mobilenetv3	94.44	95.08	94.76	0.98	3.9	20.6
RLS-mobilenetv3-+mish	95.52	95.26	95.39	0.98	3.9	24.8
RLS-mobilenetv3+mish+cbam	96.86	96.24	96.55	0.64	2.58	26.4

by 3.6%, the model size is reduced by 56.6%, and the recognition time of a single image is increased by 3.5ms, which achieves the improvement of the prediction accuracy and the optimization of the model size at the cost of a slight increase in the recognition time.



Note: 0: food waste; 1: recyclable waste; 2: other waste; 3: hazardous waste

FIGURE 8. Confusion matrix.

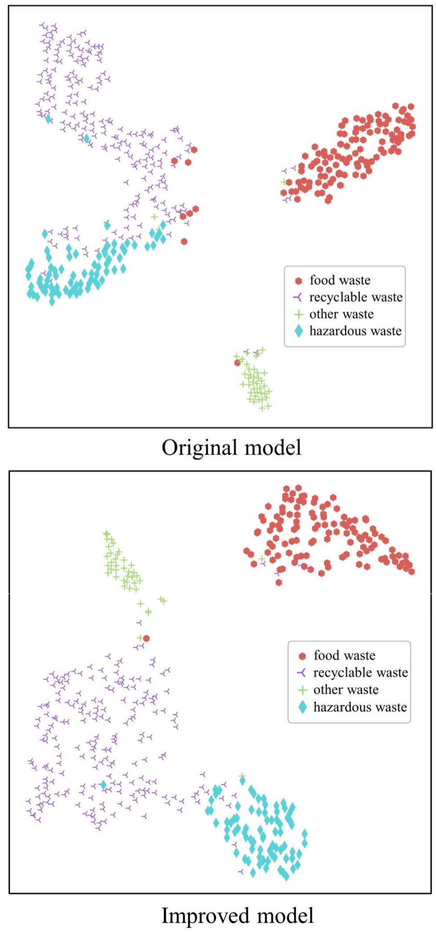


FIGURE 9. Visualization plots of t-SNE downscaling.

In order to verify the performance improvement of the GMC-MobileNetV3 model compared to the original model, the classification effects of the two models were tested separately using a test set, and the confusion matrix of the test is shown in Figure 8. By comparing the number of samples predicted correctly by the confusion matrix, it can be seen that the improved model improves the recognition accuracy of food waste, recyclable waste, other waste and hazardous waste by 5.5%, 3.05%, 2.63% and 6.94%, respectively, compared to the original model. From the confusion matrix of the original model, it can be found that the main reason for the low accuracy is that recyclable and hazardous wastes are easily confused during the classification process. It is speculated that some recyclable waste and other waste have similar characteristics, and the improved model effectively improves the performance of distinguishing the two types of garbage.

We use the t-SNE dimensionality reduction method for visualization to calculate the similarity measure between pairs of instances in the high-dimensional space and the low-dimensional space. As shown in Figure 9, the original model considers that some of the recyclable and hazardous wastes have higher similarity of features, which explains that the original model was prone to misclassification of recyclable and hazardous wastes when performing classification. For the improved model, the similarity between the four categories of waste is lower, and each category has unique features that are

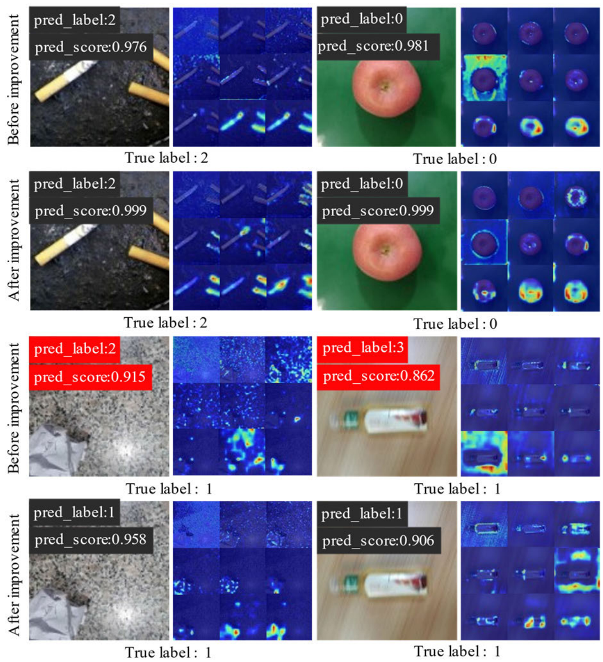


FIGURE 10. Example of prediction and feature maps for garbage image.

less likely to cause confusion when classifying, which gives the improved model greater stability when classifying.

Figure 10 shows the Grad-CAM visualization results of garbage prediction before and after the model improvement. The nine feature maps of garbage are the heat maps of eight convolutional layers and one model classification layer in the model structure, which can reflect the location of the model's attention in the recognition process and the degree of attention to the target object. From the figure, it can be seen that the extraction ability of the improved model for texture features has been improved, and it can better recognize the edges of the target. Meanwhile, for the recognition of fuzzy images and complex images, the improved model pays better attention to the target than the original model, so the Top-1 classification accuracy of the improved model in recognizing garbage is higher than that of the original model.

IV. CONCLUSION

In this paper, based on the lightweight neural network MobileNetV3, the attention mechanism, activation function and classification layer mechanism are improved to construct a garbage recognition and classification model. On the self-constructed garbage dataset, the recognition accuracy of the improved model reaches 96.55%. Experimental analysis of the garbage recognition model proposed in this paper shows that replacing the fully connected layer with global average pooling can reduce the number of parameters in the model and improve the recognition effect of the model at the same time. Using the Mish activation function and the CBAM attention mechanism, the features of the target can be extracted more accurately and the extracted image information is better utilized, which improves the recognition accuracy of the model. Compared with the original model and other mainstream convolutional neural network models trained using the same method and dataset, the improved model in this paper achieves high accuracy and low cost in garbage image classification, which is of great value and significance for helping small embedded devices complete garbage classification tasks.

We believe that this work has achieved certain research results, providing insights for similar visual recognition problems. However, in this study, we only used image data of common household waste, so the model may exhibit insufficient recognition ability for types of waste that do not appear in the training set. In the future, we will use more different types of garbage image data in the training process to make the model more suitable for practical application scenarios.

REFERENCES

[1] National Bureau of Statistics of the People's Republic of China, *China Statistical Yearbook*. Beijing, China: China Statistics Press, 2022.

[2] X. Cai, F. Shuang, X. Sun, Y. Duan, and G. Cheng, "Towards lightweight neural networks for garbage object detection," *Sensors*, vol. 22, no. 19, p. 7455, Sep. 2022.

[3] Z. Yang and D. Li, "WasNet: A neural network-based garbage collection management system," *IEEE Access*, vol. 8, pp. 103984–103993, 2020.

[4] W. Yanyan, W. Yajie, W. Chenglei, and S. Yinghao, "A novel garbage images classification method based on improved VGG," in *Proc. 34th Chin. Control Decis. Conf. (CCDC)*, Hefei, China, Aug. 2022, pp. 1571–1575.

[5] X. Ma, Z. Li, and L. Zhang, "An improved ResNet-50 for garbage image classification," *Tehnički vjesnik*, vol. 29, no. 5, pp. 1552–1559, 2022.

[6] W.-L. Mao, W.-C. Chen, C.-T. Wang, and Y.-H. Lin, "Recycling waste classification using optimized convolutional neural network," *Resour., Conservation Recycling*, vol. 164, Jan. 2021, Art. no. 105132.

[7] T. Gupta, R. Joshi, D. Mukhopadhyay, K. Sachdeva, N. Jain, D. Virmani, and L. Garcia-Hernandez, "A deep learning approach based hardware solution to categorise garbage in environment," *Complex Intell. Syst.*, vol. 8, no. 2, pp. 1129–1152, Apr. 2022.

[8] G. Yaqing and B. Ge, "Research on lightweight convolutional neural network in garbage classification," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 781, no. 3, May 2021, Art. no. 032011.

[9] C. Wang, J. Qin, C. Qu, X. Ran, C. Liu, and B. Chen, "A smart municipal waste management system based on deep-learning and Internet of Things," *Waste Manage.*, vol. 135, pp. 20–29, Nov. 2021.

[10] X. Liu, Z. Wu, Z. Wu, L. Zou, L. Xu, and X. Wang, "Lightweight neural network based garbage image classification using a deep mutual learning," in *Proc. Int. Symp. Parallel Archit., Algorithms Program.*, 2021, pp. 212–223.

[11] Z. Chen, J. Yang, L. Chen, and H. Jiao, "Garbage classification system based on improved ShuffleNet V2," *Resour., Conservation Recycling*, vol. 178, Mar. 2022, Art. no. 106090.

[12] Z. Feng, J. Yang, L. Chen, Z. Chen, and L. Li, "An intelligent waste-sorting and recycling device based on improved EfficientNet," *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, p. 15987, Nov. 2022.

[13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[14] M. Yang, and G. Thung, "Classification of trash for recyclability status," CS229 Project Rep., 2016, vol. 1, p. 3.

[15] Z. Mushtaq, S.-F. Su, and Q.-V. Tran, "Spectral images based environmental sound classification using CNN with meaningful data augmentation," *Appl. Acoust.*, vol. 172, Jan. 2021, Art. no. 107581.

[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[17] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.

[18] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[19] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.

[20] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[22] W. Xie, S. Li, W. Xu, H. Deng, W. Liao, X. Duan, and X. Wang, "Study on the CNN model optimization for household garbage classification based on machine learning," *J. Ambient Intell. Smart Environ.*, vol. 14, no. 6, pp. 439–454, Nov. 2022.

[23] H. Pan, Z. Pang, Y. Wang, Y. Wang, and L. Chen, "A new image recognition and classification method combining transfer learning algorithm and MobileNet model for welding defects," *IEEE Access*, vol. 8, pp. 119951–119960, 2020.

[24] R. A. Aral, S. R. Keskin, M. Kaya, and M. Haciömeroglu, "Classification of TrashNet dataset based on deep learning models," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2058–2062.

[25] C. Shi, R. Xia, and L. Wang, "A novel multi-branch channel expansion network for garbage image classification," *IEEE Access*, vol. 8, pp. 154436–154452, 2020.

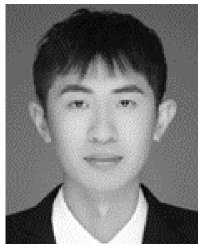
[26] V. Ruiz, A. Sánchez, J. F. Vélez, and B. Raducanu, "Automatic image-based waste classification," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.*, 2019, pp. 422–431.

[27] T. Kennedy, "OscarNet: Using transfer learning to classify disposable waste," Dept. Deep Learn., Stanford Univ., Stanford, CA, USA, Tech. Rep. CS230, 2018.

[28] S. L. Rabano, M. K. Cabatuan, E. Sybingco, E. P. Dadios, and E. J. Calilung, "Common garbage classification using MobileNet," in *Proc. IEEE 10th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ. Manage. (HNICEM)*, Nov. 2018, pp. 1–4.

XUEYONG TIAN was born in Linyi, Shandong, China, in 1979. He received the Ph.D. degree in navigation, guidance, and control from the Armored Army Engineering College, Beijing, China, in 2011.

He is currently the Director of the Institute of Intelligent Environmental Protection Technology, Shenyang University of Technology. His research interests include artificial intelligence technology, intelligent environmental protection technology, intelligent control theory, and intelligent robots.



LIWEI SHI was born in Tangshan, Hebei, China, in 1998. He received the bachelor's degree in vehicle engineering from Shenyang University of Technology, Shenyang, China, in 2020. His research interest includes the application of machine vision in the field of environment.

YUANQING LUO was born in Harbin, Heilongjiang, China, in 1990. He received the Ph.D. degree in mechanical engineering from the Shenyang University of Technology in 2021. His main research directions include signal processing, intelligent diagnosis of rotating machinery, and acoustic metamaterials.

XINLONG ZHANG was born in Jinzhou, Liaoning, in 2001. He received the bachelor's degree in mechanical design, manufacturing, and automation from Shenyang University of Technology, in 2022, where he is currently pursuing the master's degree. His research interest includes intelligent disassembly system for new energy vehicle batteries.