Resources, Conservation & Recycling 178 (2022) 106090

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

journal homepage: www.elsevier.com/locate/resconrec

Full length article

Garbage classification system based on improved ShuffleNet v2

Zhichao Chen [a,b], Jie Yang [a,b,*], Lifang Chen [c], Haining Jiao [a,b,*]

[a] School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China
[b] Jiangxi Provincial Key Laboratory of Maglev Technology, Ganzhou 341000, Jiangxi, China
[c] School of Science,Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China

A R T I C L E I N F O

Keywords:
Garbage image classification
ShuffleNet v2
Attention mechanism
Activation function
Transfer learning
Recycling

A B S T R A C T

Garbage classification technology is not only an important basis for the harmless treatment of waste and resource recovery, but also the inevitable trend of social development. The current garbage classification methods rely on manual classification in the garbage collection stage, and it is difficult to achieve satisfying results in consistency, stability, and sanitary conditions. For this reason, this study designs and develops a garbage classification system based on deep learning that can recognize and recycle domestic garbage. Focusing on the problems of low accuracy and poor real-time performance, a lightweight garbage classification model GCNet (Garbage Classification Network) is proposed. GCNet contains three improvements on ShuffleNet v2, including the design of parallel mixed attention mechanism (PMAM), the use of new activation functions, and transfer learning. The experimental results show that the average accuracy of GCNet on the self-built dataset is 97.9%, the amount of model parameters is only 1.3M, the single inference time on Raspberry Pi 4B is about 105ms, and the classification system needs only 0.88 seconds to complete the classification and collection of a single object. The method proposed in this paper is an effective attempt at machine vision in garbage classification and resource recovery. With the improvement of technology, it will effectively promote academic exploration and engineering application in the field of resources and environment.

1. Introduction

In recent years, as the living standard of residents continues to improve and the consumption structure becomes richer, the amount of domestic garbage is increasing dramatically, and many cities are suffering from the pain of "garbage siege" (Chen et al., 2019; Xu et al., 2015). Garbage classification is considered an effective way to improve resource efficiency and protect the environment, which has been actively promoted as a management measure everywhere (Meng et al., 2019; Yang et al., 2021). However, the implementation of garbage classification has not been optimistic due to the variety of garbage categories, the low awareness of garbage classification among residents, and the imperfection of relevant policies. Therefore, it is of great academic value and practical significance to study for an effective automatic garbage classification method.

The development of artificial intelligence provides a new solution to this problem. Many scholars have designed intelligent garbage classification algorithm based on deep learning technology (Azhaguramyaa et al., 2021; Kang et al., 2020; Mao et al., 2021; Zhang et al., 2021a,b). These methods can be directly applied to intelligent garbage sorting equipment, such as intelligent trash can, garbage sorter, intelligent garbage station, and mobile app, which will effectively improve the efficiency of garbage sorting. But there are still problems such as complex model structures, long inference times and high computational costs. The general application object of garbage classification technology is classification devices. For garbage classification devices, identification is usually performed before the action, i.e. the controller can send instructions to the actuator only after the identification result is obtained. If the real-time recognition of garbage is not guaranteed, the control process will have a significant delay or even a system crash, resulting in classification failure. Combined with the application, it may be impractical to equip every device with a high-performance graphics processing unit (GPU) device, which results in a huge chip cost. The speed of the executive motor is relatively easy to control, and even a faster motor can be selected, which will not significantly increase the cost. Therefore, the main factor restricting the speed of the whole system is the computing platform. At present, significant progress has been made in lightweight deep learning models, which are smaller and faster

* Corresponding author.
E-mail addresses: YJ_yccf@163.com (J. Yang), jiaohaining@yeah.net (H. Jiao).

https://doi.org/10.1016/j.resconrec.2021.106090
Received 22 August 2021; Received in revised form 24 November 2021; Accepted 27 November 2021
Available online 4 December 2021

while ensuring accuracy, providing theoretical and technical support for embedded devices to achieve garbage classification.

The achievements of large network models in computer vision tasks are remarkable, such as AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016). However, the bulk and structural complexity of large network models, which can often only be run in high-computing GPU devices, has prompted a shift to lightweight research in deep learning. Lightweight neural networks have been developed so far, and there are Xception Chollet (2017), MobileNet (Sandler et al., 2018), ShuffleNet (Zhang et al., 2018), and a host of other lightweight networks that are smaller and faster while ensuring accuracy. The advent of these models has made it possible to run deep learning models directly on edge computing devices.

Given the advantages of lightweight neural networks in terms of inference speed, a large number of scholars have used lightweight networks for image classification tasks. Gfa et al. (Fu et al., 2019) designed a fast and accurate classification model for steel surface defects based on the pre-trained backbone SqueezeNet for transfer learning, which could reach 100 frames per second (FPS)on the TITAN X platform. Tang et al. (2020) proposed a lightweight grape disease classification model with ShuffleNet selected as the model backbone and optimized using a channel attention mechanism, with a recognition accuracy of 99.14% and only 1.1M parameters; Jia et al. (2021) proposed a lightweight model for chest x ray (CXR) image classification of neocrown pneumonia, using MobileNet v3 as the backbone network and weighted combination of the output of point-by-point convolutionwith an accuracy of 99.6% and only 1.86M model parameters; Gao et al. (2021) designed a lightweight network backbone based on a novel spatial attention mechanism and transfer learning to classify garbage images, and achieved good results in the garbage classification task on Huawei cloud platform with an accuracy of 96.17% and about 450M FLOPs (floating-point operations per second); Yang and Li (2020) drew on current mainstream network structures and used a deeply separable convolutional lightweight model, while introducing an attention mechanism, to design a lightweight garbage classification network, WasNet, with an accuracy of 82.5% on a self-built household garbage dataset and a model parameter count of 1.5M.

The above research has improved the lightweight network from several aspects, which improves the accuracy while maintaining its lightweight nature. Therefore, this study applies the speed advantage of lightweight neural networks to garbage image classification and designs an efficient garbage classification algorithm for application to embedded devices. Based on the lightweight ShuffleNet v2 (Ma et al., 2018) network backbone, a fast and accurate garbage image classification method is designed in the following three aspects:

1. Embedding a parallel mixed attention mechanism in the model to enhance important features and suppress useless features;
2. Using FReLU (Ma et al., 2020) instead of ReLU as activation function to achieve pixel-level spatial information modeling;
3. Training based on transfer learning (Tan et al., 2018) to optimize the initialization weights of the network backbone and improve the model performance.

2. Related work

As mentioned above, usually deep learning models are large in the number of parameters and computation, demanding high computational resources, and are difficult to run directly in edge devices. In this case, scholars have proposed various lightweight models. The following is a brief review of the current mainstream lightweight models and an explanation of why ShuffleNet v2 is chosen.

In 2017, Chollet (2017) proposed Xception based on Inception and deep separable convolution. Deep separable convolution improves the efficiency of convolution operations and reduces the FLOPs of convolution layers and parameters. Therefore, while maintaining the same complexity, Xception used deep separable convolution to design a more complex network structure, and its accuracy on the ImageNet dataset was better than Inception v3. However, the branch structure of the model is complex, and the calculation process is fragmented, resulting in low calculation efficiency.

In 2018, Sandler et al. (2018) proposed MobileNet v2 on the basis of MobileNet v1. The main innovations are as follows: using deep separable convolution to reduce the number of model parameters; proposing a reverse residual structure to enlarge the number of network layers; using a linear bottleneck structure to reduce the loss of low-dimensional features. In the end, the accuracy of the ImageNet dataset reached 72%, and it only spent 75ms using the central processing unit (CPU) of the Google Pixel.

In the same year after MobileNet v2 was proposed, Ma et al. (2018) proposed ShuffleNet v2 on the basis of ShuffleNet v1. The contribution of ShuffleNet v2 is to propose four design guidelines for efficient networks, which are as follows: keeping the number of input and output channels of the convolutional layer equal to minimize the memory access cost in the convolution operation; reducing the use of group convolution to reduce memory access cost; minimizing branch structure when designing the network to enhance parallel operation ability; reducing element-level operations to improve the network operation speed. Their experimental results show that under the same model complexity, ShuffleNet v2 has higher accuracy than MobileNet v2 and Xception.

The main focus of this study is to apply the model to resource-limited edge devices, where both the recognition accuracy and speed of the model need to be considered. Based on the analysis of various lightweight networks, ShuffleNet v2 is an efficient and elegant lightweight model with better recognition accuracy than other lightweight models with the same complexity. Therefore, we choose ShuffleNet v2 as the base model for this study.

3. Materials and methods

3.1. Garbage sorting machine

3.1.1. Introduction of the sorting machine

We used Unigraphics(UG) software to model the structure of the garbage classification system. The overall device structure is shown in Fig. 1 and is divided into four parts: garbage delivery platform, machine vision recognition, actuating motor, and collection box. The moving belt speed range is 0–50 cm/s, which is adjustable. The distance between garbage is set to at least 40 cm to ensure the stability of the system. The collection box is a four-box structure corresponding to the set garbage categories, including recyclable garbage, wet garbage, hazardous garbage, and dry garbage. The specific devices are described as follows.

Photoelectric sensor: as shown in Fig. 1(a), the component selected is the E18-D80NA photoelectric switch with an operating voltage of 5V and a detection distance of 3-80cm.

Industrial camera: as shown in Fig. 1(b), the camera of choice is LRCP1080P; the focal length of the camera is 8.0mm; the output image is in RGB pixel format with a resolution of 1920 × 1080; the interface of the camera is USB, and the image transmission rate is 30 FPS.

Actuating motor: as shown in Fig. 1(c), the motor is a stepper motor, the specific model is 86BYG250D, providing stable and precise control of the rotation angle of the motor.

According to the modeling structure design diagram of the garbage classification system, the physical production is conducted, as shown in Fig. 2. The device works on the mechanism that the recognition area, near the end of the conveyor belt, has an industrial camera and photoelectric sensor installed; when the photoelectric sensor detects the garbage, it triggers the industrial camera and obtains the image of the current frame; the system calls the image recognition algorithm to quickly recognize the garbage and classify it according to the recognition result. The executor drives the classification baffle to rotate according to the classification situation so that the garbage falls into the

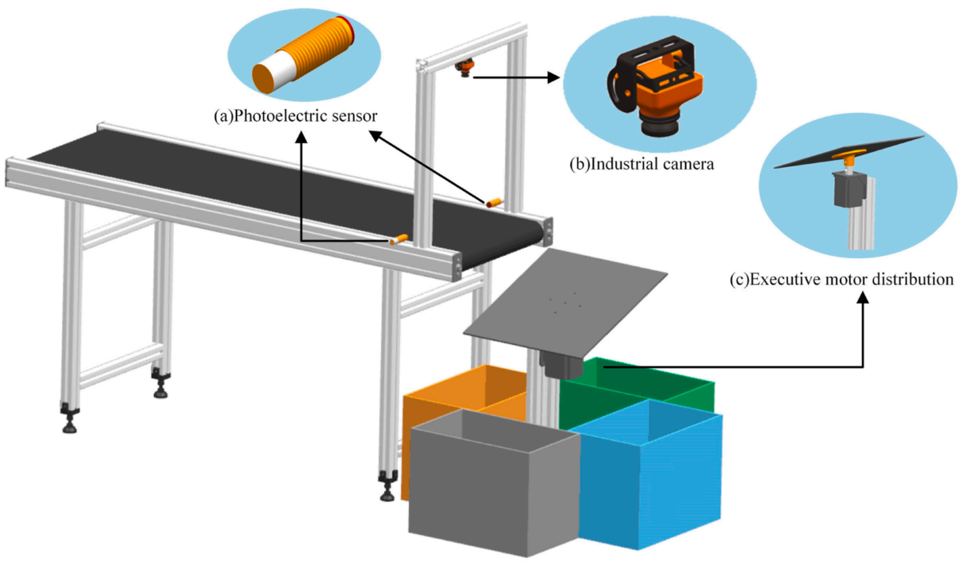

Fig. 1. Structure of the garbage classification system.

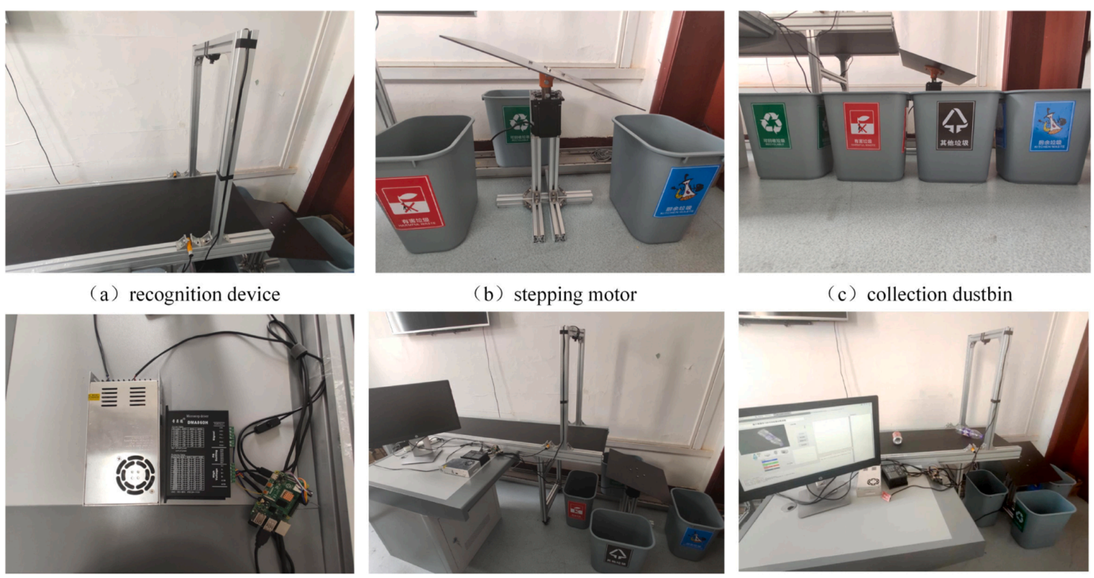

Fig. 2. Production of the garbage classification system.

corresponding box along the inclined direction of the classification baffle.

3.1.2. *Control circuit*

Raspberry Pi (Chheda et al., 2014; Khan et al., 2020; Nikolaidis and Refanidis, 2021) was chosen as the central control chip of the classification system. Raspberry Pi is a microcomputer motherboard based on ARM architecture and is widely used in artificial intelligence, IoT, and other applications. Its system supports programming languages such as Python, C, and C++ with Linux-based, rich programmable circuit pins, and has the advantages of low cost and low power consumption. To ensure the stability and real-time of the system, Raspberry Pi 4B is selected with a CPU frequency of 1.5GHZ, RAM size of 4G, support for peripherals such as cameras and displays, and 40 circuit pins, 8 of which are programmable. Finally, the hardware circuit is designed as shown in Fig. 3.

3.2. *Dataset description*

This study constructed a self-built garbage image dataset with rich scenes. These images were collected from web pictures and real-life shots, including single objects, complex backgrounds, and many other scenes, involving various interferences such as lighting, motion blur, and object deformation. According to the garbage classification standards of Shanghai (TON, 2020), this article divided all household garbage into four categories: recyclable garbage, wet garbage, hazardous garbage, and dry garbage. The collection boxes were also set according to the standards. A total of 14 sub-categories were collected in the dataset, with a total of 4256 images. The specific relationship and quantity are shown in Table 1.

3.3. *Algorithm design*

To address the problems of high hardware resource consumption,

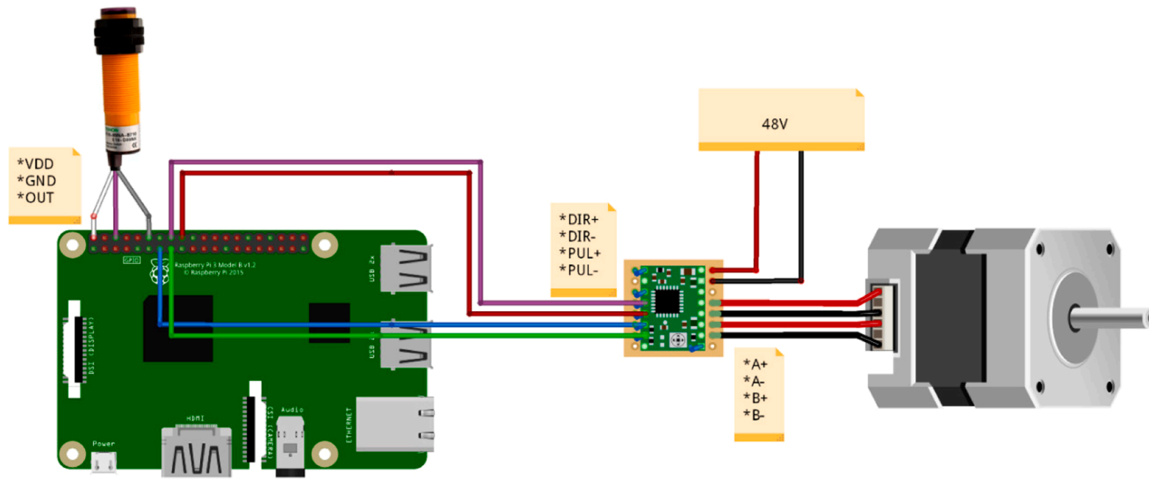

Fig. 3. Control circuit.

Table 1
Detail of the self-built garbage dataset .

Category	Name	Number	Category	Name	Number
recyclable garbage	towel	302	wet garbage	orange	287
	spitball	264		leaf	341
	packing bag	291		banana	317
	metal	365	hazardous garbage	bulb	321
	paper box	260		battery	300
	bottle	357	dry garbage	plastic bag	270
	book	298		glass cullet	283

background interference, and low classification accuracy in garbage image classification, a lightweight network ShuffleNet v2 is selected and improved by transfer learning, introducing attention mechanism and optimizing activation function to design a low consumption and high accuracy end-to-end garbage classification model, named GCNet, and the model architecture is shown in Fig. 4. The model is described as follows: build a backbone feature extraction network with ShuffleNet v2 as the core; initialize the backbone network weights by transfer learning to enhance the training starting point of the model; a parallel mixed attention mechanism is introduced to boost the network's ability to refine features, so that useful features are enhanced and useless features are suppressed; the new activation function FReLU is adopted, which makes the activation function stage with spatial context feature extraction capability and also avoids the deactivation of neurons caused by using ReLU when the input has negative values.

3.3.1. *ShuffleNet v2*

ShuffleNet v2 (Ma et al., 2018) is an upgraded version of ShuffleNet v1 proposed by MEGVII, which is based on channel shuffle and four design criteria and outperforms both ShuffleNet v1 and MobileNet v2 in terms of accuracy at the same complexity. ***Channel shuffle*** Grouped convolution (Krizhevsky et al., 2012; Zhang et al., 2019) allows each convolution kernel to operate only on the corresponding channel grouping, significantly reducing the computational cost. However, grouped convolution causes the output features of a given channel to be related only to the inputs within that group, preventing the exchange of information between channel groups and weakening the expressiveness of the output features. To solve the above problem, ShuffleNet (Zhang et al., 2018) proposed the method of channel shuffle, which ensures that the feature maps of different channel groupings exchange information without increasing the computation. The application of channel shuffle in group convolution is shown in Figure 5. The rule is to randomly and uniformly disrupt the output features of the previous group convolution in the channel dimension. After channel shuffle, each group of input features of the next grouped convolution contains the output features of different groups of the previous group convolution, so that the input and output are correlated on the channel.

Design guidelines Previous to ShuffleNet v2, the common measure of complexity of lightweight models was the number of FLOPs. ShuffleNet v2 proves that FLOPs can only be used as an indirect indicator of model

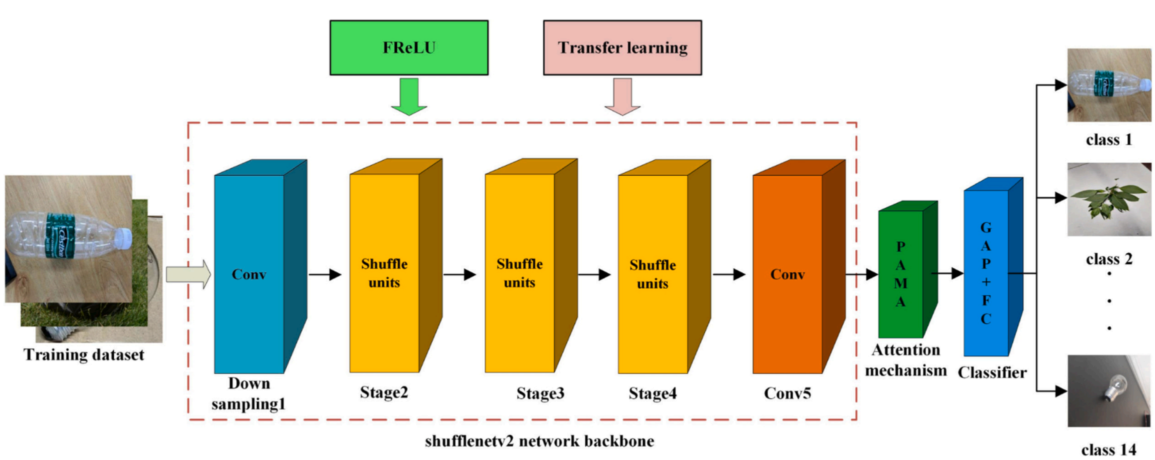

Fig. 4. The framework of the garbage classification model.

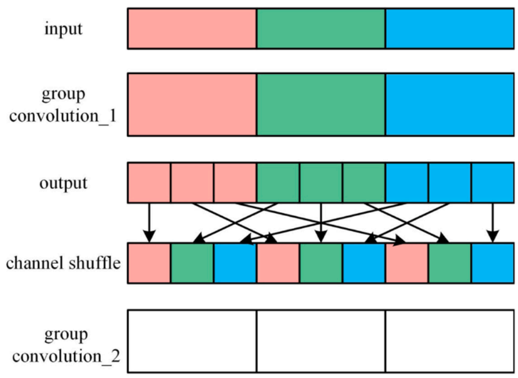

Fig. 5. Application of channel shuffle in group convolution.

complexity. Factors such as memory access costs, parallelism levels, and tensor operations can cause large differences in model inference speeds under the same FLOPs. For this reason, ShuffleNet v2 is designed based on four criteria, and the constructed cell structure is shown in Fig. 6. Channel separation is the division of input features into two parts by channel; DW convolution is depthwise convolution (Chollet, 2017). The basic unit of ShuffleNet v2 is designed utilizing four guidelines. The corresponding guidelines are as follows:

(1) Minimizing the memory access cost by keeping the number of channels of input features and output features in the convolutional layer as equal as possible. For an example of 1×1 convolution, the input feature size is $c_i \times h \times w$, c_o is the number of channels of output, B is FLOPs,then:

$$B = hwc_i c_o \tag{1}$$

$$MAC = hw(c_i + c_o) + c_i c_o \tag{2}$$

When B is kept fixed, according to the mean inequality:

$$MAC \geq 2\sqrt{hwB} + \frac{B}{hw} \tag{3}$$

When $c_i = c_o$, the inequality sign holds and MAC consumption is minimum.

(2) Reducing the use of group convolution, too many of which can increase the cost of memory access. Assuming that the number of groups in the group convolution is g, then:

$$B = \frac{hwc_i c_o}{g} \tag{4}$$

$$\begin{aligned} MAC &= hw(c_i + c_o) + \frac{c_i c_o}{g} \\ &= hwc_i + \frac{Bg}{c_i} + \frac{B}{hw} \end{aligned} \tag{5}$$

It can be shown that the MAC increases as the number of groups g increases when the amount of floating-point operations B is certain.

(3) Reducing network branches: The more branches in the network design, the slower the speed. In the Inception architecture, a large number of multi-branch structures are used as the basic module of the network, which affects the parallel computing ability of the computer.

(4) Reduction of tensor operations: e.g. ReLU activation function and feature summation operations, although the FLOPs are both small, the MAC consumption is significant.

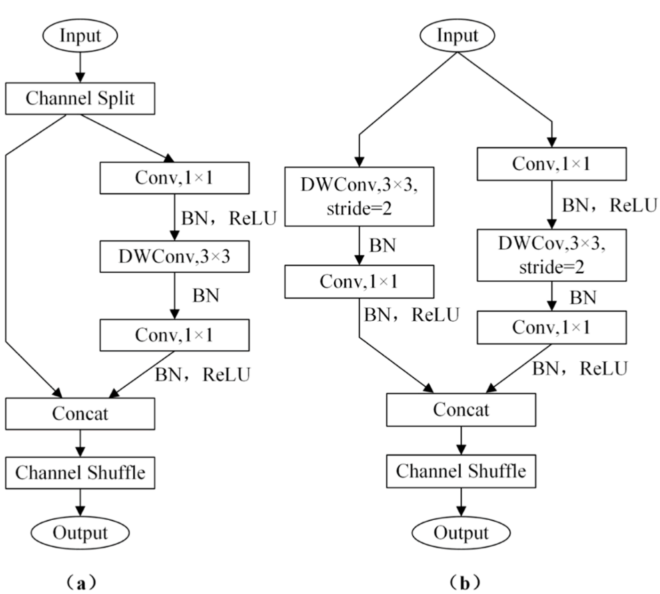

Fig. 6. The basic unit of Shufflenet v2.

3.3.2. Attention mechanism

CBAM (Woo et al., 2018) is an integrated channel and spatial convolutional attention mechanism module that can be embedded in the network for end-to-end learning with CNN (Convolutional Neural Network) models, effectively enhancing important features of the feature map and suppressing useless features. CBAM is composed of channel attention and spatial attention sequence sub-modules. The input feature $\mathbf{F}$ is weighted by the channel attention module to obtain $\mathbf{F}'$, and then the feature $\mathbf{F}''$ is output through the spatial attention mechanism. The calculation process is as follows:

$$\begin{aligned} \mathbf{F}' &= \mathbf{M_c}(\mathbf{F}) \otimes \mathbf{F} \\ \mathbf{F}'' &= \mathbf{M_s}(\mathbf{F}') \otimes \mathbf{F}' \end{aligned} \tag{6}$$

Where $\mathbf{M_c}$ represents the channel attention mapping, $\boldsymbol{M_s}$ represents the spatial attention mapping. $\otimes$ represents the element multiplication with the broadcast mechanism.

As can be observed, CBAM is a serial connection between the channel and spatial attention mechanisms, which can lead to the latter's input being derived from the features modified by the former's attention mechanism, affecting the latter's feature learning to a certain extent. At the same time, the channel attention mechanism uses the MLP layer to perform dimension reduction and dimension increasing operations, which not only increases the number of parameters but also easily causes feature loss. In response to the above problems, this paper proposes a parallel mixed attention mechanism module (PMAM), which is designed to be connected in parallel, so that the input of channel attention and spatial attention mechanism are original features and independent of each other; the MLP layer is replaced by one-dimensional convolution to avoid feature degradation dimensions to strengthen the correlation between channels. The designed parallel mixed attention mechanism module is shown in Fig. 7.

As can be seen from Fig. 7, the PMAM module enables features to enhance important features in both the channel and spatial dimensions through parallel channel attention and spatial attention mechanisms, with the following overall formulation:

$$\mathbf{F}' = \mathbf{F} \otimes \mathbf{M_s}(\mathbf{F}) \otimes \mathbf{M_c}(\mathbf{F}) \tag{7}$$

For the channel attention mechanism, the feature map $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$ is simultaneously pooled by the global maximum pooling and the global average pooling to output the corresponding results $\mathbf{F_{max}} \in \mathbb{R}^{C \times 1 \times 1}$,$\boldsymbol{F_{avg}} \in \mathbb{R}^{C \times 1 \times 1}$; the two feature vectors are added after one-dimensional convolution for feature extraction, and the final result is activated using a nonlinear activation function to obtain $\mathbf{M_c} \in \mathbb{R}^{C \times 1 \times 1}$. The expressions are as follows:

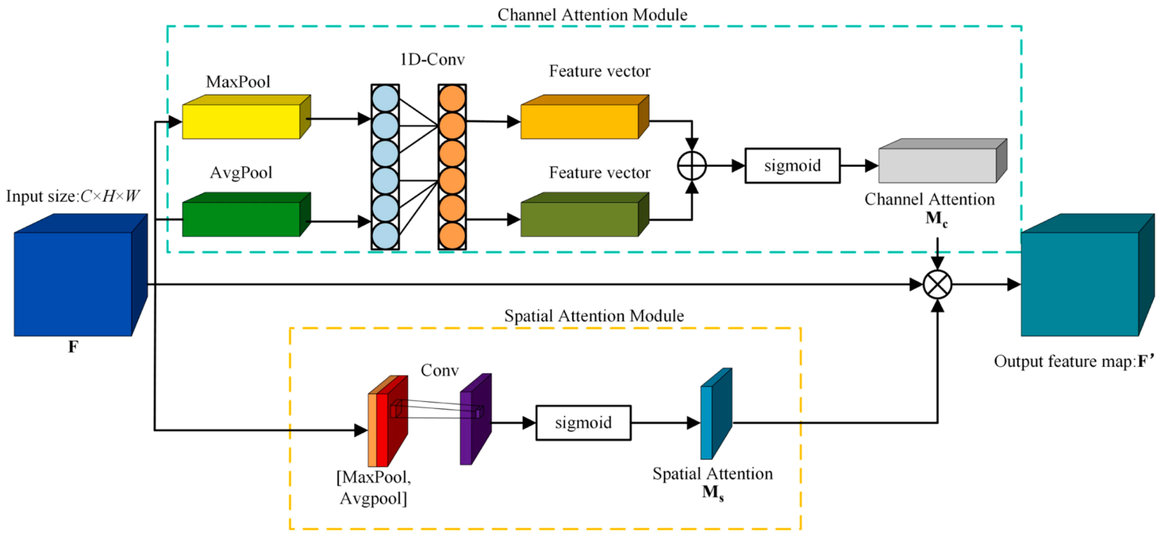

Fig. 7. Parallel mixed attention mechanism module (PMAM).

$$\mathbf{M_c}(\mathbf{F}) = \sigma\left(f_{1d}(\mathbf{F}_{\max}) + f_{1d}(\mathbf{F}_{\mathbf{avg}})\right) \tag{8}$$

Where σ is the nonlinear function sigmoid and f_{1d} is the one-dimensional convolution function.

For the spatial attention mechanism, the feature map is $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$ subjected to maximum pooling and mean pooling according to channel dimensions, and the results are concat to produce a feature description of size $2 \times H \times W$. The above feature description is fed into the convolution layer for computation, and the result is activated nonlinearly using the sigmoid function to obtain the spatial attention map $\mathbf{M_s} \in \mathbb{R}^{1\times H\times W}$. The expressions are as follows:

$$\mathbf{M_s}(\mathbf{F}) = \sigma\left(f_{2d}\left(f_c\left(\mathbf{F}_{\mathbf{max}}, \mathbf{F}_{\mathbf{avg}}\right)\right)\right) \tag{9}$$

Where f_{2d} is a two-dimensional convolution operation and f_c is a concat operation.

3.3.3. Activation function

In convolution neural networks, a linear convolution layer is usually followed by a nonlinear activation function to enhance the nonlinear ability of the model. The activation function usually uses ReLU (Glorot et al., 2011), which is computationally fast and speeds up the training of the network. However, the ReLU activation function is a unified operation for all pixels and lacks spatial modeling capabilities. In addition, when a negative value is input, the neuron will die and the gradient cannot be updated.

The FReLU activation function is a new visual activation function proposed by Ma et al. in the ECCV2020 paper (Ma et al., 2020). The author compared FReLU and ReLU in a number of experiments such as classification, detection, and segmentation, and proved that FReLU is more effective than ReLU. FReLU is a simple and effective activation function, which is innovative in its ability to model spatial information at the pixel level in the activation function stage. The spatial conditions based on convolution are used, and manual feature selection is not required. The schematic diagram is shown in Fig. 8, and the expression is as follows:

$$\mathrm{FReLU}\left(x_{c,i,j}\right) = max\left(x_{c,i,j}, T\left(x_{c,i,j}\right)\right) \tag{10}$$

$$T\left(x_{c,i,j}\right) = x^{\omega}_{c,i,j} \cdot p^{\omega}_{c} \tag{11}$$

where $T(x)$ represents the spatial condition, implemented through the depthwise separable convolution and BN layers. p^{ω}_{c} represents the shared weights of the convolution kernels in the same channel in the depthwise separable convolutions.

3.3.4. Transfer learning

Domain and task are the core concepts of transfer learning (Tan et al.,

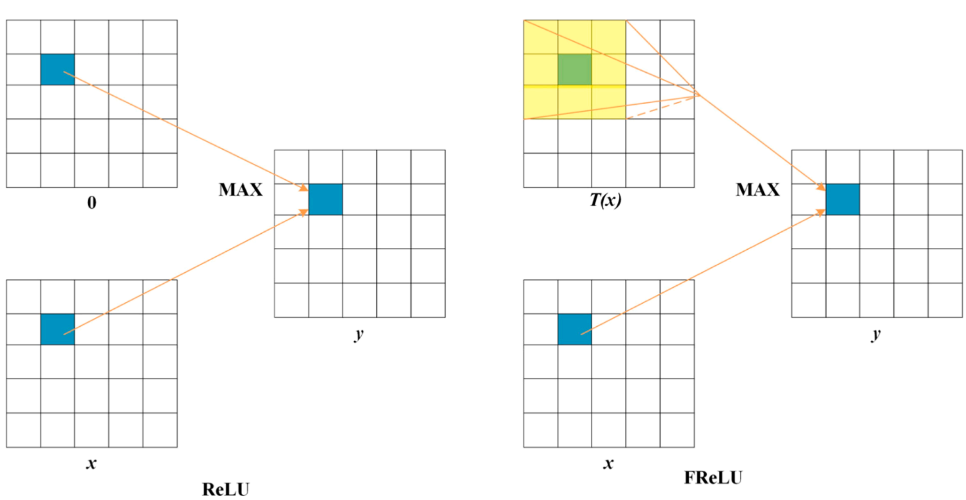

Fig. 8. Working principle comparison of ReLU and FReLU activation functions.

2018). Domain $\mathscr{D}$ is composed of feature space $\mathscr{X}$ with dimension d and marginal probability distribution $P(X)$, its mathematical expression is $\mathscr{D} = \{\mathscr{X}, P(X)\}$, where X represents the sample set. Task $\mathscr{T}$ is composed of the label space Y of the domain and the prediction model $f(x)$, which is $\mathscr{T} = \{Y, f(x)\}$. Transfer learning is a given source domain $\mathscr{D}_s$ and source domain learning task $\mathscr{T}_s$, as well as the target domain $\mathscr{D}_t$ and target learning task $\mathscr{T}_t$, using the knowledge in the source domain $\mathscr{D}_s$ and source task $\mathscr{T}_s$ to improve the prediction model in the target domain $\mathscr{D}_t$ and target domain task $\mathscr{T}_t$ performance.

ImageNet (Krizhevsky et al., 2012) is a definitive public dataset in computer vision that has grown to include more than 14 million images covering more than 20,000 categories. Due to the huge size and wide range of categories of ImageNet, many researchers Díaz-Romero et al. (2021); Relekar and Shanmugam (2021); Talo (2019) have adopted the dataset as a source domain for transfer learning to improve the accuracy of their models. Therefore, this study uses the training weights of ShuffleNet v2 in ImageNet to design the ShuffleNet v2 network backbone based on transfer learning. The specific steps are as follows:

1) Downloading the official PyTorch release of ShuffleNet v2 model files trained on the ImageNet dataset;

2) Using the pre-training weights to initialize the GCNet model weights. When loading the network layer parameters, only the backbone part is loaded, and the classifier layer is skipped.

3.4. Experimental setup

In this study, GCNet is selected as the research object. Firstly, conducting an ablation experiment on the self-built garbage dataset to verify the contribution of the improvements proposed in this paper to the model. Secondly, a comparative experiment of different CNN models is carried out in the self-built garbage dataset to compare the performance differences between GCNet and other CNN models. Finally, to further test the advantages of GCNet and the universality of transform to other tasks, this study set up a comparative experiment to compare the performance of GCNet and other CNN models on the large public datasets CIFAR-100 (Krizhevsky and Hinton, 2009) and tiny-ImageNet (Vinyals et al., 2016).

In addition to the lightweight models GCNet, MobileNet v2, ShuffleNet v2, three classic models, VGG16, ResNet50, ResNet101, were considered for comparison. The experimental platform used in our evaluation is the Nvidia TITAN RTX GPU, Pytorch and the CuDNN library. In the experiments, the data are divided into two groups: training set and test set, with the percentages of 80% and 20%, correspondingly. The test set is used for the prediction and evaluation of the model. We follow the spit proportions used in Mohanty et al. (2016).

The hyperparameters were standardized on all networks. During the training of the model, a data enhancement strategy was performed on the dataset to enhance the generalization ability of the model, including random brightness variations, random center cropping of images, and random flipping of images. Stochastic gradient descent (SGD) was chosen as the model optimizer, the loss function was cross-entropy, and the learning rate was set using the stochastic gradient descent strategy SGDR (Loshchilov and Hutter, 2017) with hot restart, with a maximum learning rate of 0.01; the training period was set to 200, and the input images per batch were 32. The application of this study is for an embedded device, therefore, the inference time consumption of the model is tested in the embedded device Raspberry Pi 4B.

4. Results and discussion

4.1. Results of ablation experiment

In order to demonstrate the contribution of each improvement point to the performance improvement of the network, the improvement points proposed in this study were introduced into ShuffleNet v2 separately and ablation experiments were conducted. The ablation experiments include the introduction of a mixed attention mechanism PMAM module, FReLU activation function, and transfer learning. The training curve of the ablation experiment on the self-built dataset is shown in Fig. 9, and the numerical comparison of the final training results is shown in Table 2. Where A is the top1 accuracy of the model, P represents the number of parameters of the model, and FLOPs is the number of floating-point operations per second. It can be seen that ShuffleNet v2 itself has good performance, and even though it is a lightweight network, it still achieves high recognition accuracy, with a recognition accuracy of 93.4%; after the introduction of the mixed attention mechanism PMAM, the cost of the model remains essentially the same, and the accuracy is improved by 0.9%; FReLU improves the performance of the network by realizing the pixel-level spatial modeling ability through two-dimensional spatial conditions, resulting in a 1.1% improvement on model accuracy; initializing the model weights with transfer learning improves the training starting point of the model, resulting in a 2.6% improvement in model recognition accuracy. Finally, by introducing the mixed attention mechanism PMAM module, FReLU activation function and transfer learning, GCNet achieves an accuracy of 97.9% in the household garbage dataset. The model achieves a 4.7% accuracy improvement compared to the baseline network with only a small increase in cost. GCNet achieves improved prediction accuracy while ensuring light weight.

4.2. Algorithm comparison and analysis

4.2.1. Performance on the self-built dataset

There exist a variety of image classification models based on convolutional neural networks, and in order to illustrate the performance of GCNet, experimental comparisons are conducted using a variety of current mainstream models. The experiment selects ResNet, VGG, MobileNet and other series of models in recent years, and trains them in the self-built dataset of this research. Due to the fact that most models have pre-trained weights, for fairness, all models are initialized with the weights using transfer learning. In the training process, after each training cycle is completed, the training loss of the model and the accuracy rate in the test set are recorded, so as to observe the training situation of the model and ensure that each model completes the training of the model under the condition of convergence. The training curves for each model on the self-built dataset are shown in Fig. 10, and numerical comparison of the final training results is shown in Table 3. where A is the test accuracy, E is the cycle of training, and L is the training loss value.

From the training curve, it can be seen that ShuffleNet v2 has higher accuracy than ResNet50, VGG16, and MobileNet v2 in the garbage classification task, and is close to ResNet101, achieving excellent results in the garbage classification task, proving the performance advantage of ShuffleNet v2. GCNet is improved based on ShuffleNet v2. With only a small increase in cost, the accuracy rate is 4.5% and 4.7% higher than ShuffleNet v2 and ResNet-101 respectively, and the parameter amount of the model is only about 1/30 of that of the large-scale network ResNet101. The specific reasons are as follows:

1) ShuffleNet v2 itself is an excellent lightweight network that uses channel shuffling to enhance the feature extraction capability of the model and designs an efficient network based on four criteria;

2) This study introduces a parallel mixed attention mechanism PMAM module to enhance features in the dimension of channels and space and suppress irrelevant features;

3) Using FReLU instead of ReLU activation function to achieve pixel-level spatial information modeling, which helps to improve model accuracy;

4) The transfer learning approach can better initialize the model weight parameters and improve the accuracy in similar tasks.

4.2.2. Performance on the public dataset

In order to further analyze the performance of GCNet in other image

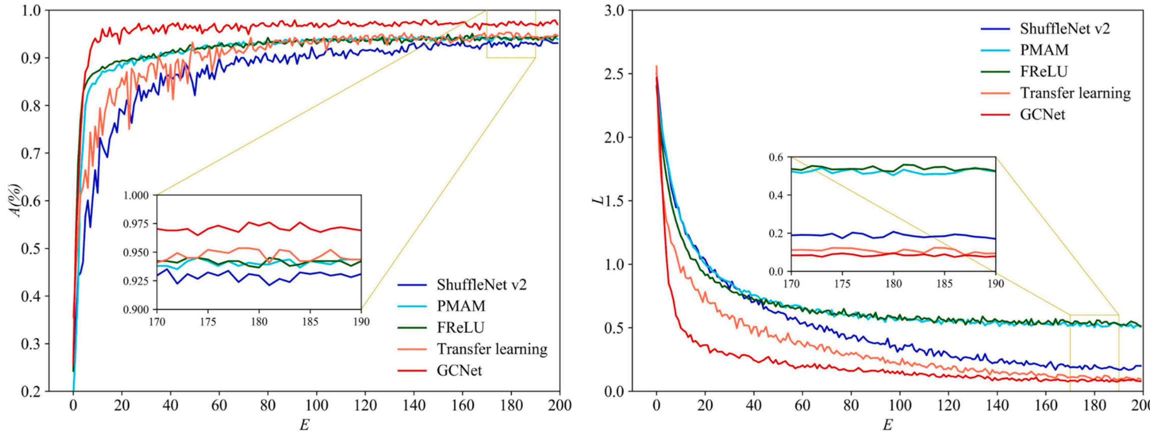

Fig. 9. The training curve of ablation experiment on the self-built dataset.

Table 2
Results of ablation experiment

Model	factor			A/%	P/M	FLOPs/M
	PMAM	FReLU	Transfer learning			
ShuffleNet v2	-	-	-	93.2	1.267	149.59
	✓	-	-	94.1	1.268	149.60
	-	✓	-	94.3	1.335	165.96
	-	-	✓	95.8	1.267	149.59
GCNet	✓	✓	✓	97.9	1.336	165.97

classification tasks, each model is trained using large public datasets CIFAR-100 and tiny-ImageNet. CIFAR-100 and tiny-ImageNet are large datasets, which are very challenging in image classification tasks, and the accuracy of the models on these datasets is generally not high. In order to prove the difficulty of the dataset, we recorded the changes in the accuracy of GCNet with and without transfer learning on tiny-ImageNet dataset. The results are shown in Fig. 11. It can be seen that the accuracy of GCNet using transfer learning is 69.8%, which is significantly higher than the model without transfer learning. At the same time, the experiment also proved that there is no negative transfer learning in the experiment.

The accuracy comparison of each model on the test set is shown in the Table 4. In the table, A_{IM} is the test accuracy rate of the model on tiny-ImageNet, and A_{CF} is the test accuracy rate of the model on CIFAR-100. It can be seen that the accuracy of GCNet on the tiny-ImageNet and CIFAR-100 datasets is 69.8% and 76.2% respectively, both surpassing the benchmark network ShuffleNet v2 and second only to the ResNet101 model. The reason why the accuracy of GCNet is lower than ResNet101 is that ResNet101's network structure is complex, the network layer is too deep, and the number of parameters is large. It can often achieve better learning results in large datasets. However, the ResNet101 model is large in size and takes a long time for inference, so it will not be deployed and used in actual application scenarios. Therefore, GCNet is equally suitable for classification tasks in other scenarios and is an efficient image classification model.

4.3. System test

4.3.1. Recognition speed test

The GCNet, MobileNet v2, and ResNet50 models are deployed on the edge device Raspberry Pi 4B, and the program is used to continuously call the model 100 times for garbage image recognition, and the recognition speed curve obtained is shown in Fig. 12. In the figure, N is the number of identifications, and T is the identification elapsed time. It

Table 3
Comparison of training results of each model on the self-built dataset .

Model	A/%	P/M	FLOPs/G
GCNet	97.9	1.336	0.16
ShuffleNet v2	93.4	1.268	0.15
MobileNet v2	92.6	2.245	0.32
ResNet-50	92.3	23.53	4.12
ResNet-101	93.2	42.53	7.81
VGG16	91.2	134.31	15.50

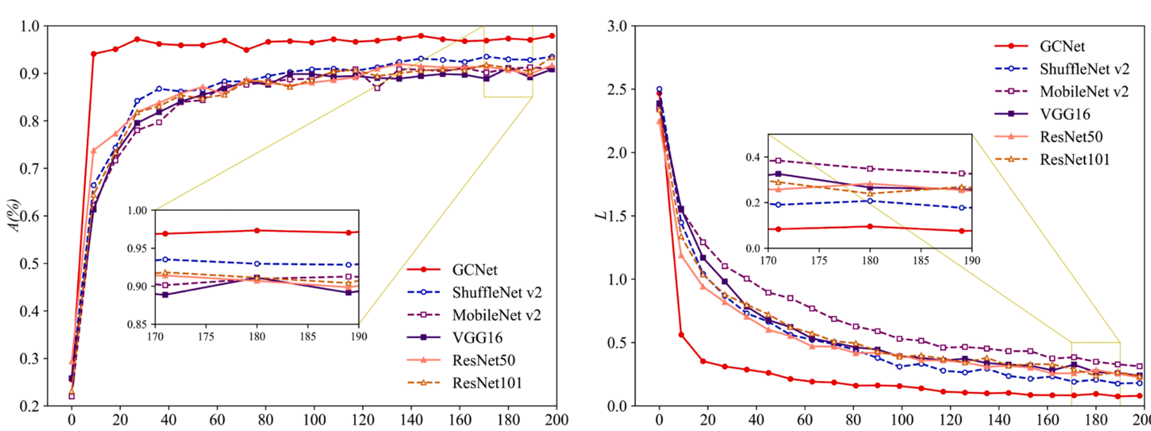

Fig. 10. The training curve of each model on the self-built dataset.

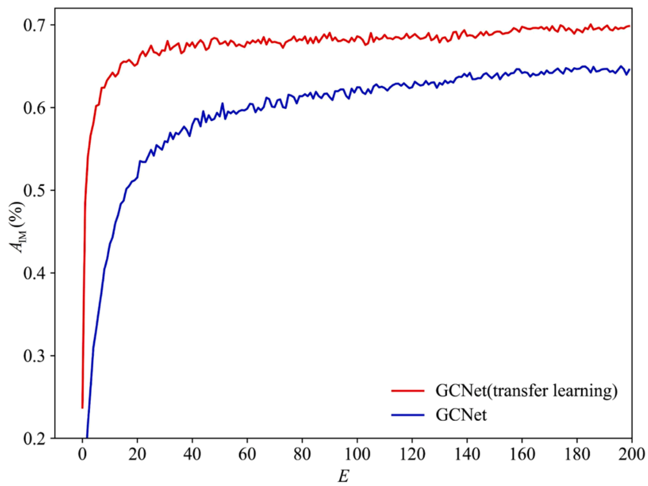

Fig. 11. Accuracy of GCNet with and without transfer learning.

Table 4
Test accuracy of each model on the public dataset .

Model	A_{IM}/%	A_{CF}/	FLOPs/G
GCNet	69.8	76.2	0.16
ShuffleNet v2	66.9	74.8	0.15
MobileNet v2	66.2	74.0	0.32
ResNet50	69.4	74.6	4.12
ResNet101	72.3	76.5	7.81
VGG16	64.5	70.5	15.50

can be seen from the test results that GCNet has achieved good real-time results on the Raspberry Pi. The single recognition time is about 105ms, which is slightly faster than MobileNet v2, which is also a lightweight model, and is far better than other large models,ensuring real-time classification of edge devices.

4.3.2. Object classification test

Perform a visual test on the recognition results of GCNet, select a set of common household garbage pictures, and pass these pictures into the deployed GCNet model. The recognized garbage categories and corresponding classification confidences are marked below the pictures. The result is shown in Fig. 13. It can be seen that the generalization ability of the model is strong, and the categorie of garbage can be correctly identified in the case of a variety of single objects, multiple similar objects, blurred images, and complex backgrounds. Therefore, considering the recognition speed and accuracy of the model, GCNet is a very effective garbage image classification model.

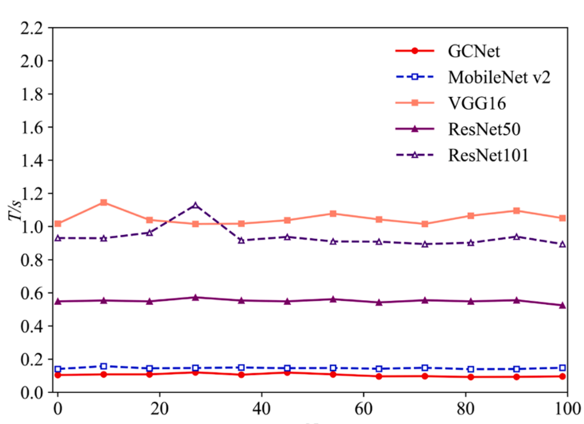

Fig. 12. The recognition velocity curve of the model.

4.3.3. System comprehensive test

The system is equipped with host computer software to facilitate the visualization of the identification process. The running interface is shown in Fig. 14. Put the bottle into the conveyor belt, and after entering the recognition area, it is recognized as a recyclable item immediately, and the classification icon and related statistical information are displayed. It can be seen that the software has achieved the effect of real-time monitoring and quantitative statistics on the status of garbage identification, meeting the visual monitoring requirements of the system workflow.

The system is put through a complete functional test and ten categories of domestic garbage are randomly thrown into the system in turn, and the corresponding test results are shown in Table 5. It can be seen that the system is stable in operation and can accurately identify garbage categories and make correct decisions; the mechanical drive performance is satisfied and can carry out garbage transportation and collection as we want. Therefore, the classification system has successfully applied GCNet, which can complete the task of automatic garbage classification and collection.

5. Conclusion

With the increasing focus on environmental protection and sustainable utilization of resources, garbage classification is an urgent problem for mankind. The current classification methods rely too much on manual participation, which is easily affected by personal quality, attention, sense of responsibility, and so on. Efficient and reliable automatic classification technology is extremely important and will be the inevitable trend of social development. Applying artificial intelligence technology to garbage classification can improve classification efficiency and further reduce labor costs. Therefore, a lightweight and efficient garbage classification model GCNet and a classification system using GCNet are designed. GCNet is an improved model based on ShuffleNet v2. The parallel hybrid attention mechanism in the original model is optimized. A new activation function is used, and the model is trained via transfer learning. The contributions of this paper are as follows:

1) A lightweight garbage image classification model GCNet is proposed, which not only ensures real-time performance, but also improves the recognition accuracy. GCNet simplifies the hardware requirements, reduces the computing cost, and meets the needs of practical applications;

2) We apply GCNet to the garbage classification system and will realize the efficient classification of all kinds of domestic garbage with the improvement of the engineering level;

3) GCNet can be applied to intelligent garbage sorting equipment, such as intelligent trash cans, garbage sorters, intelligent garbage stations. It also has utility in mobile applications and public propaganda.

The experimental results show that embedding the PMAM module can enhance the important features extracted by the model and improve the accuracy; compared with the traditional ReLU function, the FReLU activation function contributes more to the convolution network model; the training method, transfer learning, can better initialize the weight parameters of the network. The system has the advantages of simple structure, low cost, and stable operation. The processing time of single garbage is about 0.88 seconds. The method proposed in this paper has good academic significance and practical application value for the application of artificial intelligence technology in the field of garbage classification.

In this paper, there are still some limitations, which can be improved in future work. First of all, in our real life, garbage has different forms and more categories. How to build an effective data collection platform or expand datasets by using adversarial learning is a direction worthy of exploration. Secondly, the garbage classification system in this study

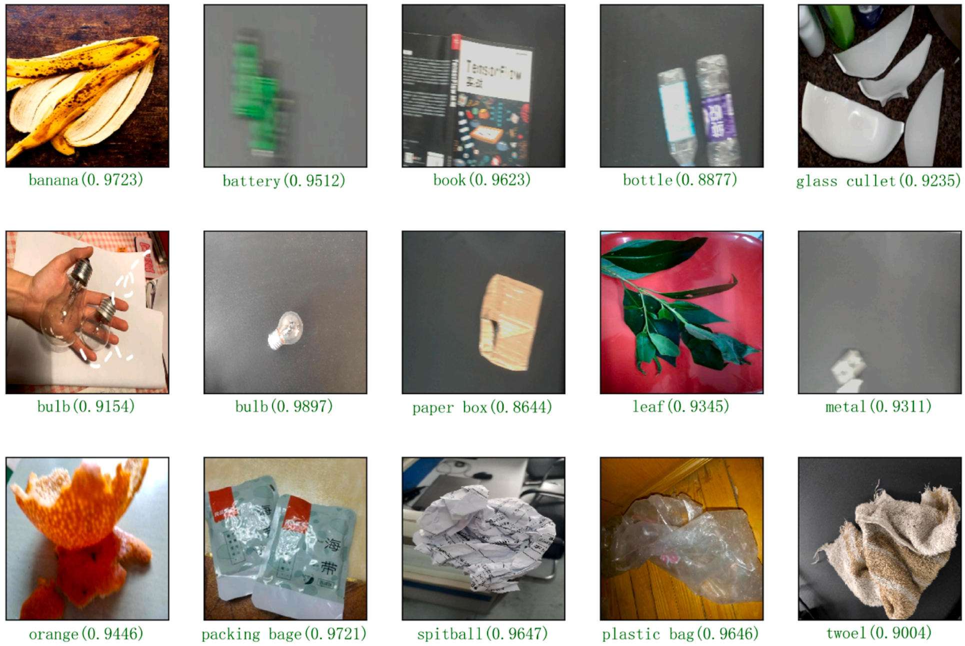

Fig. 13. Recognition results of garbage pictures.

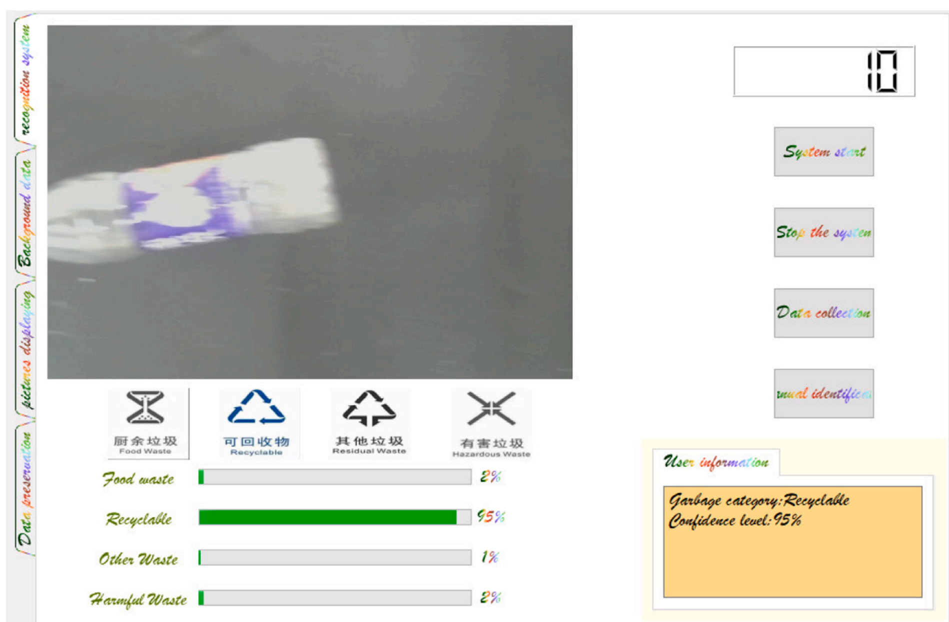

Fig. 14. Software monitoring interface.

Table 5
System test results .

Garbage name	Result	The motor rotation	Collection box
battery	battery	0	harmful garbage
plastic bag	plastic bag	+90	other garbage
book	book	+180	recyclable garbage
bottle	bottle	0	recyclable garbage
banana	banana	–90	kitchen garbage
blub	blub	+180	harmful garbage
carton	carton	–90	harmful garbage
leaf	leaf	–90	kitchen garbage
metal	metal	+90	recyclable garbage
orange	orange	–90	kitchen garbage

only supports a single object. In future work, a simple and effective garbage loading platform can be designed from the perspective of engineering to ensure that only a single object appears at a time. In addition, image segmentation technology can also be used to return multi-target garbage for judgment. Our ultimate goal is to apply GCNet to more waste sorting facilities to effectively promote resource recycling and sustainable social development.

CRediT authorship contribution statement

Zhichao Chen: Conceptualization, Methodology, Software, Validation, Resources, Writing – review & editing. **Jie Yang:** Conceptualization, Supervision, Project administration, Writing – original draft, Funding acquisition. **Lifang Chen:** Investigation, Data curation. **Haining Jiao:** Supervision, Project administration, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors gratefully appreciate the financial support provided by the National Natural Science Foundation of China (61763016) and the Jiangxi Province 03 special and 5G funded projects (20204ABC03A15).

References

China is implementing "garbage classification" action. Environ. Pollut. 259, 2020, 113707. https://doi.org/10.1016/j.envpol.2019.113707.

Azhaguramyaa, V.R., Janet, J., Narayanan, V., Sabari, R.S., Santhosh, K.K., 2021. An intelligent system for waste materials segregation using IoT and deep learning. J. Phys. Conf. Ser. 1916 (1), 012028. https://doi.org/10.1088/1742-6596/1916/1/012028, 8pp.

Chen, A., Chen, J.R., Chui, J., Fan, C., Han, W., 2019. Research on risks and countermeasures of "cities besieged by waste" in Chinaan empirical analysis based on diis. Bulletin of Chinese Academy of Sciences 34 (07), 797–806. https://doi.org/10.16418/j.issn.1000-3045.2019.07.009.

Chheda, D., Darde, D., Chitalia, S., 2014. Smart projectors using remote controlled raspberry pi. Int J Comput Appl 82 (16), 6–11. https://doi.org/10.5120/14245-2250.

Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800–1807, 10.1109/CVPR.2017.195.

Díaz-Romero, D., Sterkens, W., Van den Eynde, S., Goedemé, T., Dewulf, W., Peeters, J., 2021. Deep learning computer vision for the separation of cast- and wrought-aluminum scrap. Resour. Conserv. Recycl. 172, 105685. https://doi.org/10.1016/j.resconrec.2021.105685.

Fu, G., Sun, P., Zhu, W., Yang, J., Cao, Y., Yang, M.Y., Cao, Y., 2019. A deep-learning-based approach for fast and robust steel surface defects classification. Opt Lasers Eng 121, 397–405. https://doi.org/10.1016/j.optlaseng.2019.05.005.

Gao, M., Chen, Y., Zhang, Z., Feng, Y., Fan, W., 2021. Classification algorithm of garbage images based on novel spatial attention mechanism and transfer learning. Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice 41 (2), 498–512. https://doi.org/10.12011/SETP2020-1645.

Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, vol. 15 of Proceedings of Machine Learning Research, PMLR, Fort Lauderdale, FL, USA, pp. 315–323 https://proceedings.mlr.press/v15/glorot11a.html.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. 770–778, 10.1109/CVPR.2016.90.

Jia, G., Lam, H.-K., Xu, Y., 2021. Classification of covid-19 chest x-ray and ct images using a type of dynamic CNN modification method. Comput. Biol. Med. 134, 104425. https://doi.org/10.1016/j.compbiomed.2021.104425.

Kang, Z., Yang, J., Guo, H., 2020. Automatic garbage classification system based on machine vision. Journal of Zhejiang University (Engineering Science) 7, 1272–1280. https://doi.org/10.3785/j.issn.1008-973X.2020.07.004.

Khan, M.A., Paul, P., Rashid, M., Hossain, M., Ahad, M.A.R., 2020. An ai-based visual aid with integrated reading assistant for the completely blind. IEEE Trans Hum Mach Syst 50 (6), 507–517. https://doi.org/10.1109/THMS.2020.3027534.

Krizhevsky, A., Hinton, G., et al., 2009. Learning multiple layers of features from tiny images. Citeseer.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25, 1097–1105. https://doi.org/10.1145/3065386.

Loshchilov, I., Hutter, F., 2017. SGDR: stochastic gradient descent with warm restarts. 5th International Conference on Learning Representations, ICLR 2017,Toulon, France, april 24–26, 2017, Conference Track Proceedings, Openreview.Net https://openreview.net/forum?id=Skq89Scxx.

Ma, N., Zhang, X., Sun, J., 2020. Funnel activation for visual recognition. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (Eds.), Computer vision – ECCV 2020. Springer International Publishing, Cham, pp. 351–368, 10.1007/978-3-030-58621-8_21.

Ma, N., Zhang, X., Zheng, H.-T., Sun, J., 2018. Shufflenet v2: practical guidelines for efficient CNN architecture design. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018. Springer International Publishing, Cham, pp. 122–138, 10.1007/978-3-030-01264-9_8.

Mao, W.-L., Chen, W.-C., Wang, C.-T., Lin, Y.H., 2021. Recycling waste classification using optimized convolutional neural network. Resour. Conserv. Recycl. 164, 105132. https://doi.org/10.1016/j.resconrec.2020.105132.

Meng, X., Tan, X., Wang, Y., Wen, Z., Tao, Y., Qian, Y., 2019. Investigation on decision-making mechanism of residents' household solid waste classification and recycling behaviors. Resour. Conserv. Recycl. 140, 224–234. https://doi.org/10.1016/j.resconrec.2018.09.021.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Front Plant Sci 7, 1419. https://doi.org/10.3389/fpls.2016.01419.

Nikolaidis, S., Refanidis, I., 2021. Using distributed ledger technology to democratize neural network training. Applied Intelligence 1–17. https://doi.org/10.1007/s10489-021-02340-3.

Relekar, H., Shanmugam, P., 2021. Transfer learning based ship classification in sentinel-1 images incorporating scale variant features. Adv. Space Res. https://doi.org/10.1016/j.asr.2021.08.042.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. 4510–4520,10.1109/CVPR.2018.00474.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 10.1109/CVPR.2015.7298594.

Talo, M., 2019. Automated classification of histopathology images using transfer learning. Artif Intell Med 101, 101743. https://doi.org/10.1016/j.artmed.2019.101743.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2018. Springer International Publishing, Cham, pp. 270–279, 10.1007/978-3-030-01424-7_27.

Tang, Z., Yang, J., Li, Z., Qi, F., 2020. Grape disease image classification based on lightweight convolution neural networks and channelwise attention. Comput. Electron. Agric. 178, 105735. https://doi.org/10.1016/j.compag.2020.105735.

Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D., 2016. Matching Networks for One Shot Learning. Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16. Curran Associates Inc., Red Hook, NY, USA, pp. 3637–3645, 10.5555/3157382.3157504.

Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer vision – ECCV 2018. Springer International Publishing, Cham, pp. 3–19, 10.1007/978-3-030-01234-2_1.

Xu, S., Li, G., Ou, H., Wu, W., 2015. Research and demonstration of dynamic intelligent logistics system of the collection and transportation process of giant municipal garbage. Proceedings of china modern logistics engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 39–50, 10.1007/978-3-662-44674-4_4.

Yang, S., Wei, J., Cheng, P., 2021. Spillover of different regulatory policies for waste sorting: potential influence on energy-saving policy acceptability. Waste Manage. (Oxford) 125, 112–121. https://doi.org/10.1016/j.wasman.2021.02.008.

Yang, Z., Li, D., 2020. Wasnet: a neural network-based garbage collection management system. IEEE Access 8, 103984–103993. https://doi.org/10.1109/ACCESS.2020.2999678.

Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., Liu, X., 2021a. Recyclable waste image recognition based on deep learning. Resour. Conserv. Recycl. 171, 105636. https://doi.org/10.1016/j.resconrec.2021.105636.

Zhang, S., Chen, Y., Yang, Z., Gong, H., 2021b. Computer vision based two-stage waste recognition-retrieval algorithm for waste classification, resources. Conservation and Recycling 169, 105543. https://doi.org/10.1016/j.resconrec.2021.105543.

Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 IEEE/CVF conference on computer vision and pattern recognition, pp. 6848–6856, 10.1109/CVPR.2018.00716.

Zhang, Z., Li, J., Shao, W., Peng, Z., Zhang, R., Wang, X., Luo, P., 2019. Differentiable Learning-to-group Channels via Groupable Convolutional Neural Networks. IEEE/CVF international conference on computer vision (ICCV), 2019, pp. 3541–3550, 10.1109/ICCV.2019.00364.