Mathematical Biosciences and Engineering

MBE, 20(3): 4741–4759.
DOI: 10.3934/mbe.2023219
Received: 21 September 2022
Revised: 27 November 2022
Accepted: 08 December 2022
Published: 30 December 2022

http://www.aimspress.com/journal/MBE

Research article

Research on deep learning garbage classification system based on fusion of image classification and object detection classification

Zhongxue Yang[1], Yiqin Bao[1,*], Yuan Liu[2], Qiang Zhao[3], Hao Zheng[1] and YuLu Bao[4]

[1] School of information engineering, Nanjing XiaoZhuang University, Nanjing 211171, China
[2] Business School of Jinling University of science and technology, Nanjing 211199, China
[3] Department of Information Systems Schulich School of Business, Toronto 416647, Canada
[4] Nanjing RuiHuaTeng intellectual property Co., Ltd., Nanjing 211175, China

* **Correspondence:** Email: baoyiqin@njxzc.edu.cn; Tel: +8613851549080.

Abstract: With the development of national economy, the output of waste is also increasing. People's living standards are constantly improving, and the problem of garbage pollution is increasingly serious, which has a great impact on the environment. Garbage classification and processing has become the focus of today. This topic studies the garbage classification system based on deep learning convolutional neural network, which integrates the garbage classification and recognition methods of image classification and object detection. First, the data sets and data labels used are made, and then the garbage classification data are trained and tested through ResNet and MobileNetV2 algorithms, Three algorithms of YOLOv5 family are used to train and test garbage object data. Finally, five research results of garbage classification are merged. Through consensus voting algorithm, the recognition rate of image classification is improved to 2%. Practice has proved that the recognition rate of garbage image classification has been increased to about 98%, and it has been transplanted to the raspberry pie microcomputer to achieve ideal results.

Keywords: remote upgrade; load balancing; genetic algorithm; power monitoring terminal; voting algorithm

1. Introduction

In recent years, domestic garbage pollution has become the main source of environmental pollution, seriously damaging the global ecological environment and affecting the normal life of residents. With the promulgation of the global environmental protection and green development goals, the corresponding waste classification policies have been gradually implemented in all countries. Traditional garbage classification methods have some problems such as relying on manual classification, low classification efficiency and poor classification quality. More and more researchers are applying deep learning technology to garbage classification [1].

In the United States, scientific researchers have created a sort machine called Max-AI, which is composed of vision system, AI system and sorting system. Its vision system uses a multi-layer CNN network model, which can accurately obtain video information and classify garbage day and night. Zenrobotics in Finland has developed a large-scale AI waste classification system. It can identify the material and shape of garbage through sensors, judge the type of garbage through these, and then use four mechanical arms to automatically sort and complete garbage classification. However, these sorting machine systems all realize waste classification at the back end, which requires a large investment and is not precise enough. In China, most cities have carried out waste classification pilot projects. In July 2019, Shanghai officially implemented the domestic waste classification system scheme, which was called the strictest waste classification measure in history at that time, and gave very severe punishment to mixed dumping. However, these policies on waste classification need to be implemented by environmental health personnel. If we really want to do a good job of garbage classification for the whole people, it is not enough to rely on people's self-awareness, and a large number of workers are also required to participate. Therefore, it is necessary to design an artificial intelligence waste classification system at the front end to reduce the workload of sanitation personnel and improve work efficiency.

Many relevant scholars have conducted research on garbage classification through deep learning. Deng et al. [2] conducted research on garbage classification and recovery system based on tensorflow, Melinte et al. [3] studied deep convolution neural network object detector for real-time waste identification, Li et al. [4] realized garbage object recognition and classification based on mask score RCNN, and Lin et al. [5] conducted intelligent garbage classification system based on artificial intelligence technology. The recognition rate is 92%. Wu et al. [6] designed a garbage classification model GC-YOLOv6 based on YOLOv5 object detection network, Liu et al. [7] studied image object detection in bad weather through improved YOLO, Feng et al. [8] studied vehicle information detection through improved YOLOv3 algorithm, and Yang et al. [9] adopted the garbage recognition and detection algorithm based on YOLOv5 to realize garbage classification. Guo et al. [10] studied the deep learning image recognition technology in garbage classification, and Kang et al. [11] proposed a garbage classification algorithm based on ResNet-34 algorithm. Although the above studies have achieved certain results, these studies either use image classification algorithm to achieve garbage classification or object detection to classify garbage, and the accurate recognition rate can be further improved.

In this paper, a garbage image classification and recognition method based on image classification and object detection is proposed. Taking pytorch as the main deep learning library, the data sets used were first produced, and then the training and testing of these garbage classification modules were realized through the garbage image classification module of ResNet and MobileNetV2 algorithms and the garbage object detection module of the three algorithms of YOLOv5 family. The training results of garbage classification were merged through the consensus voting algorithm, and the accurate recognition rate was further improved to 98%, Finally, a visual garbage classification system is built to realize the application. The contributions and innovations of this paper are summarized as follows:

- Designed the deep learning garbage classification system architecture.
- ResNet and MobileNetV2 algorithms are implemented to train and test garbage images.
- Three algorithms of YOLOv5 family are implemented to train and test garbage images.
- A visual garbage classification system integrating image classification and object detection is realized by consensus voting algorithm.

The rest of the paper is organized as follows: the second section introduces the architecture of the garbage classification system, the third section introduces the relevant technical research, the fourth section introduces the training and testing of the image classification model and the object detection model, integrates the training results of garbage classification, and builds a visual garbage classification system. The fifth section summarizes the full text and prospects.

2. Garbage classification system architecture

The overall architecture of the waste classification system is shown in Figure 1. Two kinds of deep learning methods are used in system design: image classification algorithm and object detection algorithm. Image classification algorithm mainly classifies images, using ResNet and MobileNetV2. The object detection algorithm is mainly used to locate and detect the object, using three algorithms of the YOLOv5 family The training results of garbage classification are integrated, and the results are determined through the unanimous voting algorithm table. Finally, a visual garbage classification system is built. The visual system mainly has three functions: uploading pictures, real-time recognition by cameras, and image recognition.

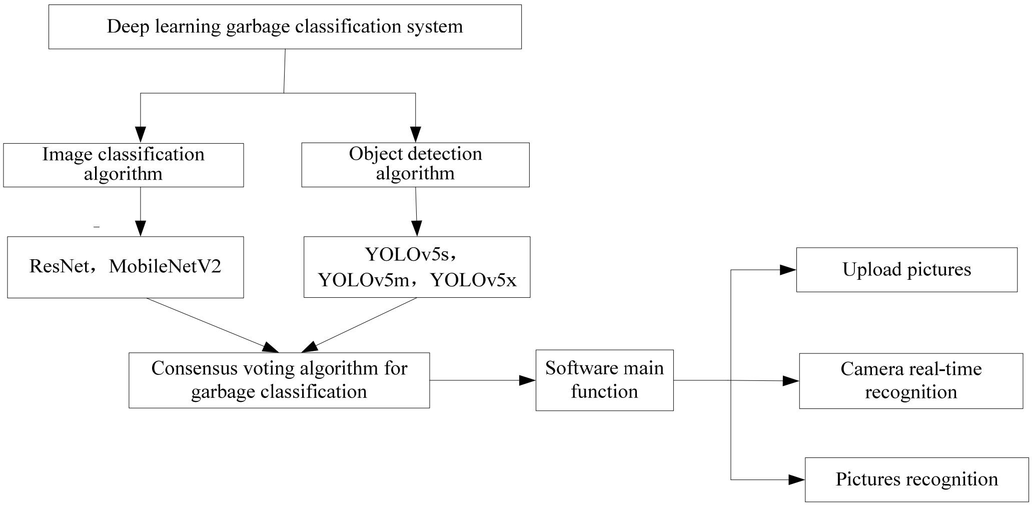

Deep learning garbage classification system
Image classification algorithm
Object detection algorithm
ResNet，MobileNetV2
YOLOv5s，YOLOv5m，YOLOv5x
Consensus voting algorithm for garbage classification
Software main function
Upload pictures
Camera real-time recognition
Pictures recognition

Figure 1. Overall structure diagram of waste classification system.

- Image classification method: select ResNet and MobileNetV2. First, collect the images of the data set, process the images, identify the data, and build the convolution neural network. Then, carry out the training set and verification set for the data set, use the model for training and testing, and adjust the parameters. Finally, get the model training results and save the model training results.

- Object detection method: three algorithms of YOLOv5 family are selected. Similarly, the data set image collection, image processing, data identification, convolution neural network construction, training set and verification set of the data set, training and testing with the model, adjusting the parameters to compare and analyze the results, and saving the model training results.
- Consensus voting algorithm: it combines the training results of image classification method and object detection method, and improves the recognition accuracy through consensus voting algorithm.
- Main interface of garbage classification system: it is built in the environment of pytorch to realize the design of GUI interface. The main buttons are: upload pictures, real-time identification of cameras, and image identification. Image recognition is to recognize the uploaded images by using the results of training of image classification method and object detection method; camera real-time recognition is real-time recognition through the camera. The garbage sorting system is finally transplanted to the raspberry terminal.

3. Related technologies and applications

3.1. Deep learning

Deep learning is a machine learning algorithm based on artificial neural network, which was proposed by Hington et al. [12] in 2006 after several years of development, various deep learning algorithms have strong feature extraction ability, rich data information expression ability and strong generalization. Breakthroughs have been made in computer vision, natural language processing, speech recognition and other fields [13]. Convolution neural network algorithm is a typical deep learning algorithm. It can automatically extract features without losing the structure information of the original image. It uses convolution operation to protect the spatial structure of the original information to a certain extent, and uses the weight value to reduce the parameters to be trained. Therefore, the model has achieved certain results in different fields of image recognition. At present, convolutional neural networks such as ResNet, MobileNetV2 and YOLOv5 have been widely used in our daily life and solved many problems.

3.2. ResNet and its application

There are many types of convolutional neural networks, among which AlexNet and ResNet have deep structures. AlexNet has five convolution layers, plus multiple layer pooling layers and two full connection layers, which can create networks with more than ten layers. ResNet, the deepest type, has 152 layers of networks, which is very deep in the convolutional neural network structure. The full name of ResNet is residual network, also known as residual network, Using ResNet allows achieving good performance and efficiency, even if the network develops in a deeper direction [14,15]. The main component of ResNet is the residual module. Residual network is a deep neural network, which follows the basic idea of using fast connections to skip blocks. It is one of the core of classical computer vision tasks and is widely used in object classification. The presentation of the network solves the degradation and overfitting problems caused by the increase of layers in the network.

ResNet deep residual learning framework is shown in Figure 2. The residual module is composed of two dense layers and one skip connection. The activation functions of the two dense layers are both relu functions. The main idea of ResNet is to add a direct connection channel in the network, called expressway network, which allows the original input information to be directly transferred to the next layer [16]. Classic ResNet models include ResNet50, ResNet18, ResNet101, etc. It has been shown that the performance of ResNet50 in image scene classification is superior to other CNN models in the Imagenet dataset [[17].

We use ResNet50 in the garbage classification system. ResNet50 consists of 50 layers and more than 25.6 million parameters. This set combines convolution, identification block (input = output) and full connectivity layer. The identification x corresponds to the input value of the original block or signal. The output value of the residual block is the sum of the input value and the output value of the block in the inner layer of the block.

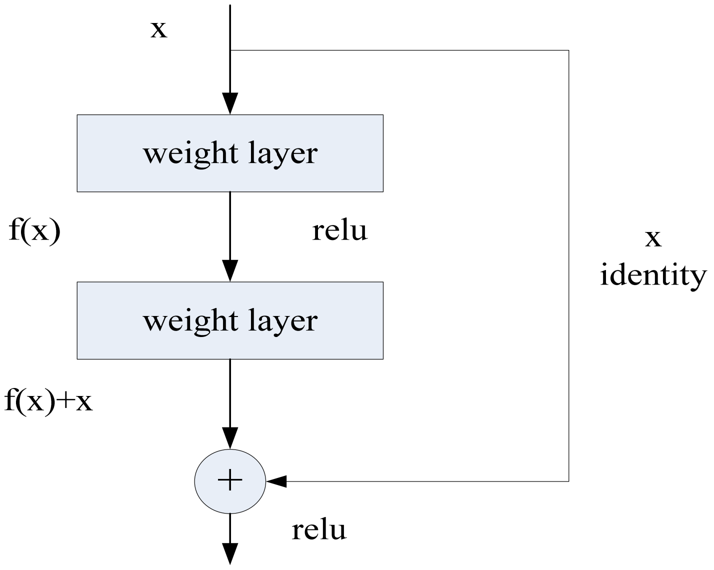


Figure 2. ResNet deep residual learning framework.

3.3. MobileNetV2 and its application

The traditional convolutional neural network has a large memory requirement and a large amount of computation, which makes it unable to run on mobile devices and embedded devices. On the premise that the accuracy rate is slightly reduced, MobileNet greatly reduces the amount of parameters and calculation of the model. MobileNetV2 is an improvement on MobileNetV1 and is a lightweight neural network [16]. MobileNetV2 retains the deep separable convolution of V1 version and adds linear bottleneck and inverted residual [17]. The model structure table of MobileNetV2 is shown in Table 1, where t is the multiple of the internal dimension of the bottleneck layer, c is the dimension of the feature, n is the number of repetitions of the bottleneck layer, and s is the step of the first conv of the bottleneck layer.

Table 1. MobileNetV2 structure table.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$52^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1 × 1	-	1280	1	1
$7^2 \times 1280$	avgpool 7 × 7	-	-	1	-
1 × 1 × 1280	conv2d 1 × 1	-	k	-	-3

When implementing the garbage classification system, we found that the expansion rate between 5 and 10 would lead to almost the same performance curve. The smaller network would have better performance at a lower expansion rate, while the larger network would have better performance at a higher expansion rate. MobileNetV2 mainly applies the expansion factor of 6 to the size of the input tensor. For example, for a bottleneck layer that uses 64 channel input tensor and generates a tensor with 128 channels, the intermediate extension layer is 64 × 6 = 384 channels.

● Linear bottleneck: for the deep separable convolution of MobileNetV1, the M-dimensional space compressed by the width multiplier will pass through a nonlinear transformation ReLU. According to the property of ReLU, if the input feature is negative, the feature of the channel will be cleared. The original feature has been compressed, which will further lose feature information; If the input characteristic is a positive number and the output characteristic through the activation layer is the original input value, it is equivalent to linear transformation. The specific structure of the bottleneck layer is shown in Table 2. Input the conv + ReLU layer through 1 to increase the dimension from k dimension to tk dimension, and then through 3 × 3conv + ReLU can separate convolution to downsample the image (when stripe > 1). At this time, the feature dimension is already tk dimension. Finally, the dimension is reduced by 1 × 1 conv (no ReLU), and the dimension is reduced from tk to k dimension.

Table 2. Bottleneck layer structure table.

Input	Operator	Output
h × w × k	1 × 1 conv2d, ReLU6	H × w × (tk)
h × w × tk	3 × 3 dwise s = s, ReLU6	h/s × w/s × (tk)
h/s × w/s × tk	Linear 1 × conv2d	245

● Backward residual: the residual block has been proved in ResNet, which helps to improve the accuracy and build a deeper network. Therefore, MobileNetV2 also introduces similar blocks. The process of classical residual block is: 1 × 1 (dimension reduction) – > 3 × 3 (convolution) – > 1 × 1 (dimension elevation). However, the feature extraction from the depth convolution layer is limited to the input feature dimension. If the residual block is used, the input feature map will be compressed first through 1 × 1 point wise convolution, and then after the depth convolution, the extracted features will be less. Therefore, MobileNetV2 first expands the channels of the feature map through 1 × 1 point by point convolution operation, enriches the number of features, and further improves the accuracy. This process just reverses the order of the residual block, which is the origin of the backward residual: 1 × 1 (ascending dimension) – > 3 × 3 (dw conv + ReLU) – > 1 × 1 (descending dimension + linear transformation).

3.4. *YOLOv5*

YOLOv5 is a classical algorithm for object detection [20]. The object detection architecture is divided into two stages. The difference is that the two stages have a region proposal process, which is similar to a screening process. The network will generate positions and categories according to candidate regions, while the single stage directly generates positions and categories from images, YOLO is a one-stage method. YOLO is the abbreviation of you only look once, which means that convolutional neural network can output results only by looking at the picture once. YOLO has released a total of 6 versions, of which the first version of V1 has played a pioneering role. The later series are improvements on V1 to improve performance. YOLOv5 is the fifth version. Compared with the previous versions, YOLOv5 absorbs their advantages, such as the high speed of detecting objects and the high accuracy of detecting small objects [21–24]. The overall structure of YOLOv5 is shown in Figure 3. It can be seen that the YOLOv5 network is mainly divided into four parts: input end, backbone, neck and input end.

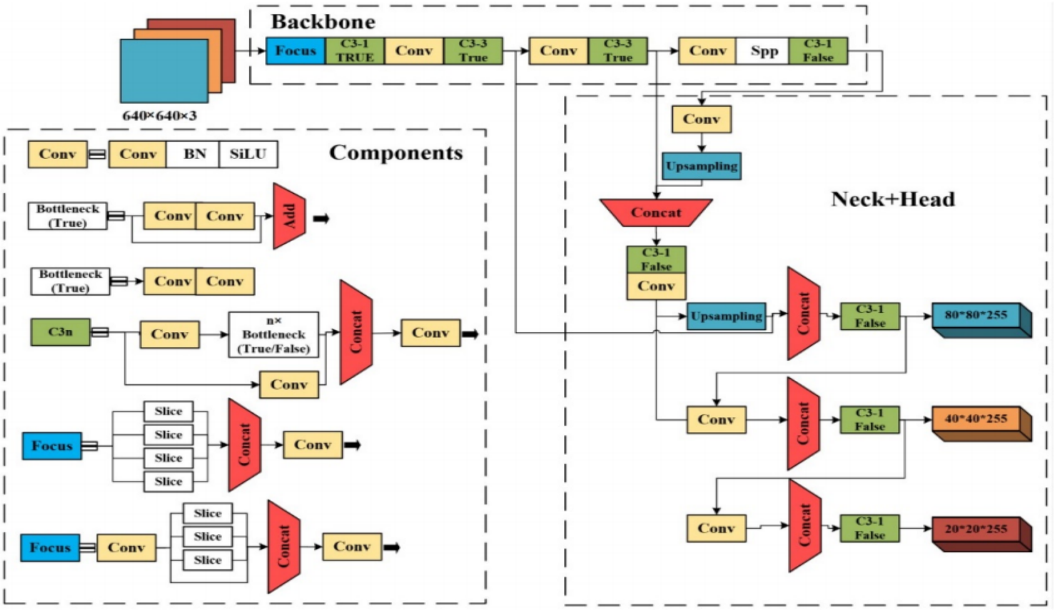


Figure 3. Schematic diagram of YOLOv5 network structure.

Object detection YOLO is a technology used to classify or predict specific object categories in the image. There may be many objects in the input image. Its task is to determine the position and category of the objects. For example, Figure 4(a) contains three objects, three objects is three cans, and one of them is identified to identify its type and position. Figure 4(b) only identifies that the image is a can, as shown in Figure 4, It is the comparison between the object detection picture and the image classification picture.

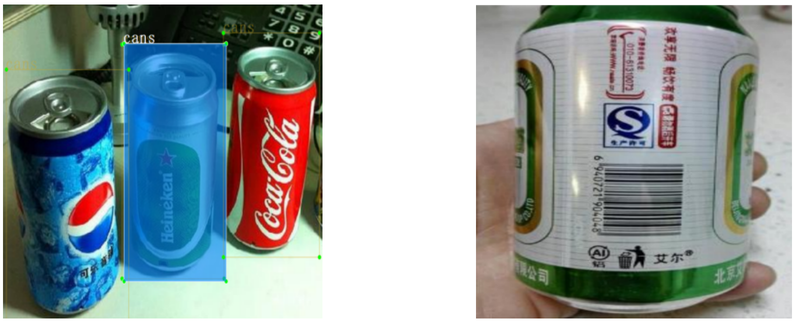


Figure 4(a) Three objects (cans). **Figure 4(b)** A picture (cans).

Figure 4. Comparison between the object detection picture and the image classification picture.

4. Deep learning model of garbage classification

4.1. Image classification model

4.1.1. Data acquisition and data labeling

Deep learning garbage image recognition requires preparing garbage images as data sets. Garbage images come from three aspects: 1) COCO dataset, 2) VOC2007 dataset, and 3) Garbage image data set taken by oneself. Garbage images are divided into four categories: hazardous garbage, recyclable garbage, kitchen waste, and other garbage. The four categories are divided into 245 sub categories. In the study, the storage of classification categories and training image files is realized through the directory and file structure. The directory name: large class name + small class name, such as: harmful garbage_ small class name, as shown in Figure 5.

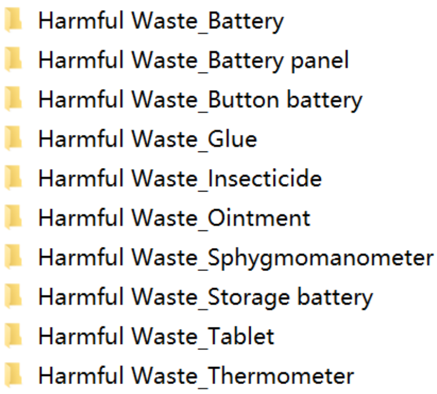

Harmful Waste_Battery
Harmful Waste_Battery panel
Harmful Waste_Button battery
Harmful Waste_Glue
Harmful Waste_Insecticide
Harmful Waste_Ointment
Harmful Waste_Sphygmomanometer
Harmful Waste_Storage battery
Harmful Waste_Tablet
Harmful Waste_Thermometer

Figure 5. Hazardous waste catalogue.

Image file: img + subclass name + number.jpg, for example: the image file name of hazardous waste battery is img_Battery_No.jpg, as shown in Figure 6(a). The shoe image file in recyclable garbage is img_shoes_No.jpg, as shown in Figure 6(b). The image file of tea leaves in kitchen waste is img_ Tea_No.jpg is shown in Figure 6(c). The mask image file in other garbage is img_Mask_ No.jpg is shown in Figure 6(d).

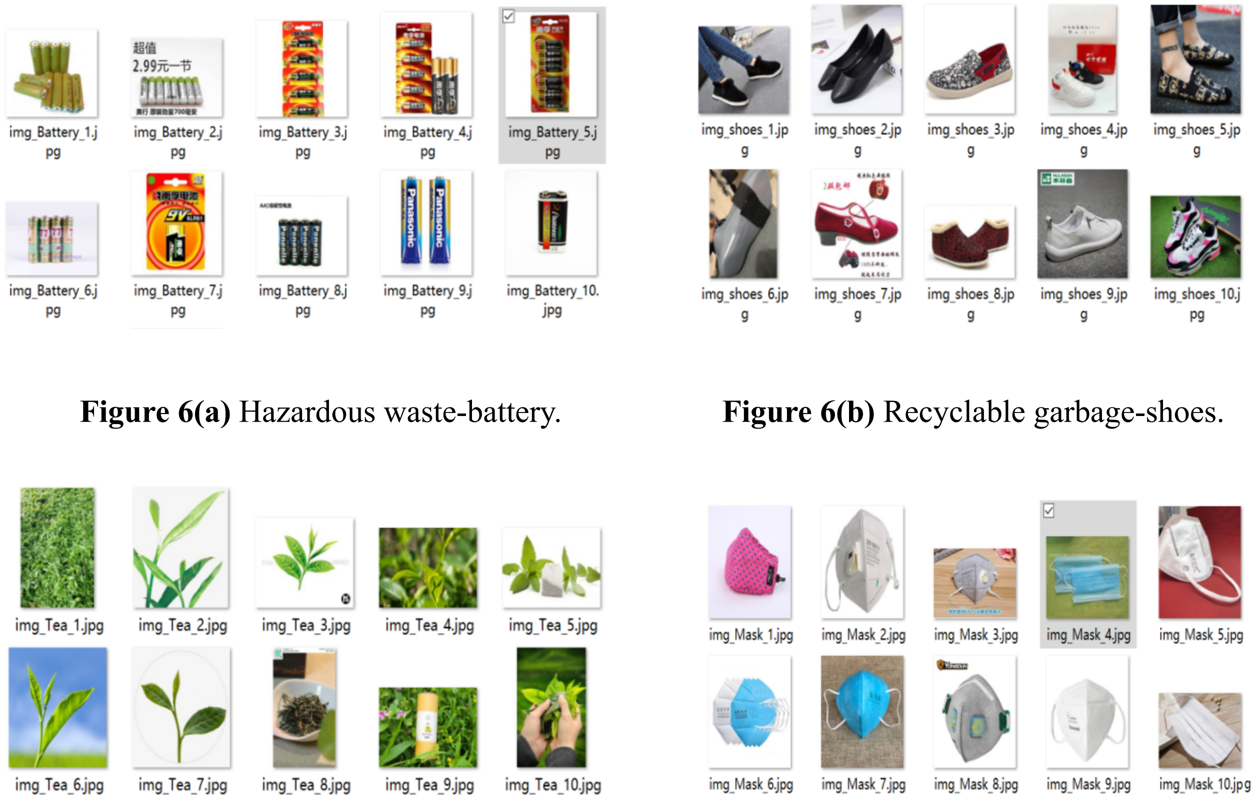


Figure 6(a) Hazardous waste-battery.

Figure 6(b) Recyclable garbage-shoes.

Figure 6(c) Kitchen waste-tea.

Figure 6(d) Other garbage-masks.

Figure 6. Garbage image file.

4.1.2. Training of image classification model training

- Division of data sets: the total number of data sets is 80,012, and the data sets are divided into 80–20% training sets and verification sets, as shown in Table 3.

Table 3. Division of image data set.

Dataset name	Number of pictures	Proportion in total	Number of garbage type
Training set	64,010	80%	245
Validation set	16,002	20%	245

- Model training: ResNet model and MobileNetV2 model are trained respectively. Due to the large number of pictures in the data set, the training round is set to 30 rounds for test training.

The accuracy and loss function of the two models after training are compared. As shown in Figure 7 (a),(b):

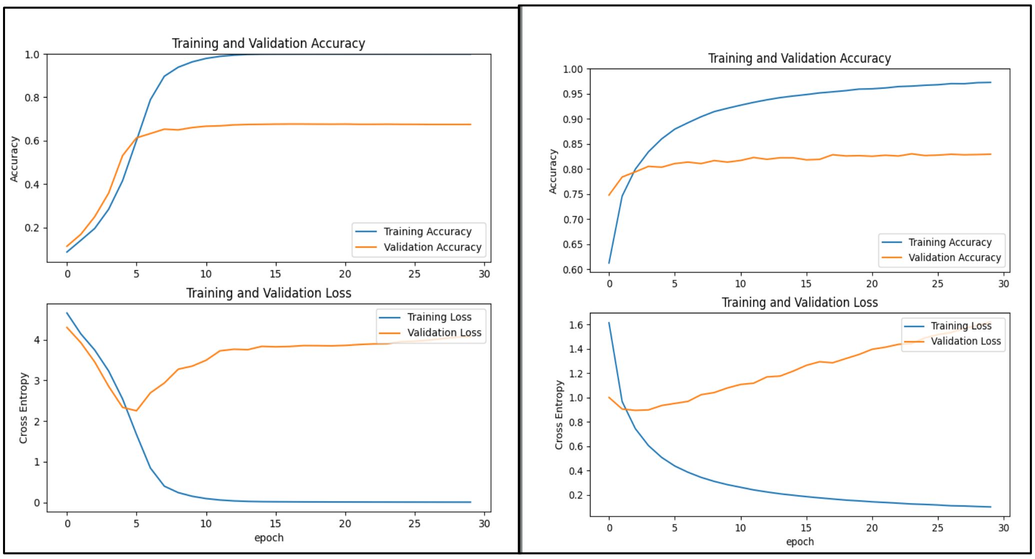


Figure 7(a) ResNet. **Figure 7(b)** MobileNetV2.

Figure 7. Accuracy and loss of two models.

- Training results: the training results of the two models are shown in Table 4:

Table 4. Comparison of model training results.

Model name:	Accuracy	Loss	Training time (hours)
ResNet	99%	0.661	41
MobileNetV2	97%	1.617	50

From Figure 7, we can see that after the first five training sessions of ResNet, the loss function and accuracy rate tend to be gentle, while the accuracy rate of MobileNetV2 reached a high level in the first training session. However, after that, the training loss function and accuracy rate directly entered the low-speed development stage. The training time of the two is not long, and the training rounds set in the experiment are 30 rounds. Although the network depth difference between ResNet and MobileNetV2 is not large, in this experiment, MobileNetV2 achieved a large improvement in accuracy in 30 rounds, and ResNet performed slightly weaker than MobileNetV2.

4.1.3. Testing of image classification model

After the training of the two models is completed, the accuracy of the model needs to be verified. The two models need to be tested separately, mainly on the verification set. Before verification, it is necessary to load the verification data set and the trained model. The test results of the two models are shown in Table 5:

It can be seen from Table 5 that MobileNetV2 occupies a dominant position in the performance of the test set. Through 30 rounds of training and testing, we can basically conclude that MobileNetV2 performs better than ResNet in the data set in this experiment, with better accuracy and reliability.

Table 5. Test results of image classification module.

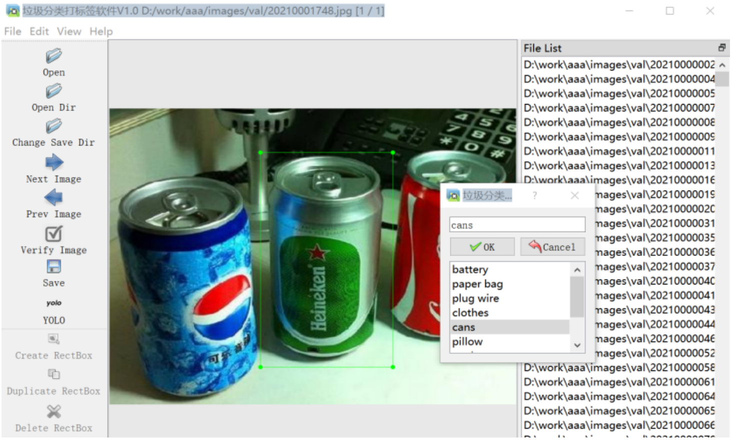
Model name:	Accuracy	Loss
ResNet	81.5	4.0753
MobileNetV2	82.9	1.621

4.2. Object detection model

The object detection model YOLOv5 has four versions, namely YOLOv5s (the smallest model), YOLOv5m, YOLOv5l and YOLOv5x (the largest model). We choose YOLOv5s, YOLOv5m and YOLOv5x for application.

4.2.1. Data acquisition and data labeling

The object detection data set is extracted from the image classification data set. In this study, 54,200 images are extracted from the image classification data set of more than 80,000 images as the object detection module data set, and 254 categories are defined. Since 1) in the object detection, there are many objects in the input image, 2) there may be several objects in a picture, so it is necessary to mark each object through the labeling software, and Pictures are marked with YOLOv5 through the labeling software. As shown in Figure 8, there are multiple cans objects in a picture:



垃圾分类打标签软件V1.0 D:/work/aaa/images/val/20210001748.jpg [1 / 1]

File List

D:\work\aaa\images\val\2021000002
D:\work\aaa\images\val\2021000004
D:\work\aaa\images\val\2021000005
D:\work\aaa\images\val\2021000007
D:\work\aaa\images\val\2021000008
D:\work\aaa\images\val\2021000009
D:\work\aaa\images\val\2021000011
D:\work\aaa\images\val\2021000013
D:\work\aaa\images\val\2021000016
mages\val\2021000019
mages\val\2021000020
mages\val\2021000031
mages\val\2021000035
mages\val\2021000036
mages\val\2021000037
mages\val\2021000040
mages\val\2021000041
mages\val\2021000043
mages\val\2021000044
mages\val\2021000046
mages\val\2021000052
D:\work\aaa\images\val\2021000058
D:\work\aaa\images\val\2021000061
D:\work\aaa\images\val\2021000064
D:\work\aaa\images\val\2021000065
D:\work\aaa\images\val\2021000066

垃圾分类...

cans

Figure 8. Labeling YOLOv5 Format Label.

The labeling process through labelming is as follows:

- Select the label format as YOLO format.
- Select a source picture folder and a destination picture folder.
- Select a picture, such as 2021xxxxxxxx.jpg.
- The generated label file 2021xxxxxxxx.txt is labeled by create rectbox. Since an image may contain multiple objects.

Therefore, the tag file is composed of multiple records, and the format of each record is: type no, pos_x, pos_y, w, h and put them into the object picture folder. An example of a dataset image is shown in Figure 9.

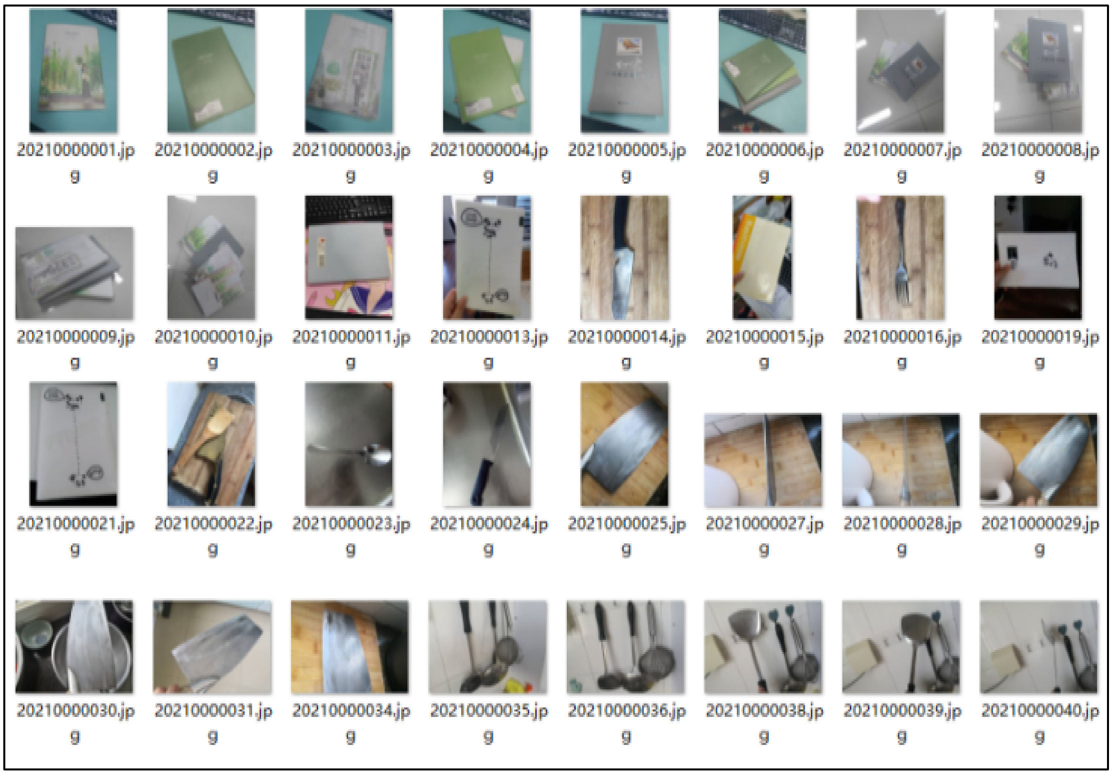


Figure 9. Sample of object detection dataset.

4.2.2. Training of three YOLO5 models

Similar to the image classification model training, before the model training, the data set YOLOv5 tag files and picture files need to be divided into training sets and verification sets, and divided into training sets and verification sets at a ratio of 80–20%, as shown in Table 6:

Table 6. Dataset division of object detection module.

Dataset name	Number of pictures	Proportion in total	Garbage type
Training set	43,360	80%	245
Validation set	10,840	20%	245

During model training, the main parameters need to be adjusted so that the training parameters of the three models of s, m and x are basically similar. The main parameters are as follows:

- epochs: is the training round. In the experiment, the training round is set to 500 rounds, which is larger than the image classification training round.
- batch-size: is the number of pictures that the GPU sees at one time during YOLOv5 training. This value determines the occupied size of the video memory. The default value is 16. If the batch size is increased, the training depth can be greatly increased. However, due to too many pictures in

the training set, the occupied video memory will exceed the actual video memory, resulting in the failure and termination of the experiment.

- img-size: is the fixed size of the picture in this experiment, and the middle note of the experiment is 640 * 640. Too large adjustment of the picture size can increase the number of recognized features, but it will increase the pressure of the GPU, and finally lead to the termination beyond the video memory.
- workers: refers to the number of threads used by the CPU during data loading. The default value is 8. If it is increased, the memory will be exceeded during training, resulting in the program being forced to retreat.

Finally, the training results of the three models are compared, as shown in Table 7:

Table 7. Comparison of training results of object detection module models.

YOLOv5 Name	Training rounds	mAP@0.5 (%)	Accuracy/confidence (%)	Recall/confidence (%)	Training time (hours)
YOLOv5s	500	93.1	94.2	92.5	147
YOLOv5m	500	94.5	91.8	93.4	178
YOLOv5x	500	95.9	92.6	92.2	265

In object detection, the index mAP (mean average precision) can better represent the advantages and disadvantages of the model than the accuracy, mAP@0.5 is the map when IoU is 0.5. It can be seen that after 500 rounds of training, without changing the main parameters, 1) mAP@0.5 of YOLOv5s is the lowest and YOLOv5x is the highest. 2) Training time YOLOv5s is the fastest and YOLOv5x is the slowest. 3) The recall rate of YOLOv5x is lower than that of YOLOv5m, because the batch size is low, which makes the advantages of YOLOv5x not better displayed. 4) The accuracy /confidence of YOLOv5m and YOLOv5x is not as high as that of YOLOv5s, and the map values of the three are not much different. Although the deeper the network is, the better the effect is, the deeper the network is. However, in practice, the depth of the network is not necessarily proportional to the accuracy. It is possible that after a lot of training, the deep model will degrade the network, and the effect is not as good as that of the shallow network model.

4.2.3. Testing of three YOLOv5 models

After the training of the three YOLOv5 models, the accuracy of the model needs to be verified. The main consideration for judging the quality of the model is to test the accuracy on the verification set. Before validation, you need to load the validation data set and the trained model. During model testing, use YOLOv5 to train the best model: best.pt. Data is the location of the test set, and the default location of the training set test set. The parameter file is coco128.yaml. Like the training file, the batch size is the number of pictures that the GPU sees at one time in the test. In order to ensure that the variables are the same as the training, imgsize is also the same as the training file. In order to control the variables, it is still changed to 640 pixels. The test results are shown in Table 8:

Table 8. Comparison of YOLOv5 model test results.

YOLOv5 Name	mAP@0.5 (%)	Accuracy/confidence (%)	Recall/confidence (%)
YOLOv5s	80.4	84.2	90.2
YOLOv5m	85.9	86.8	91.5
YOLOv5x	95.7	92.6	92.2

It can be seen that YOLOv5s mAP@0.5 The lowest and YOLOv5x the highest; The accuracy/confidence of YOLOv5s was the lowest and YOLOv5x was the highest; YOLOv5s had the lowest recall/confidence and YOLOv5x had the highest.

During the test, FPS (frames per second) comparison results of YOLOv5 are shown in Table 9:

Table 9. Comparison of mean FPS results of YOLOv5

YOLOv5 Name	FPS (frames/second)
YOLOv5s	11.2
YOLOv5m	8.4
YOLOv5x	6.8

It can be seen that YOLOv5s is the fastest and YOLOv5sx is the slowest, because the model of YOLOv5s is shallow and the processing speed is fast. The model of YOLOv5x is the deepest and the processing speed is slow. According to the performance of the three models, YOLOv5x has the best comprehensive performance.

4.3. Comparison of object detection and image classification model

For the two classification image classification models and the three object classification models, the training results of the five models are compared, mainly compared with the accuracy, training rounds and consumption time. The specific values are shown in Table 10:

Table 10. Comparison of object detection and image classification models.

Model name	Training rounds	Testset mAP@0.5 (%)	Testset Accuracy (%)	Training time (hours)
ResNet	120	-	81.5	123
MobileNet	120	-	82.9	150
YOLOv5s	500	80.4	-	147
YOLOv5m	500	85.9	-	178
YOLOv5x	500	95.7	-	265

The evaluation parameter used by the image classification module is accuracy, which is different from the meaning expressed by the evaluation parameter precision used by the object detection module. The accuracy refers to the ratio of the number of correctly classified samples to the total number of samples, and the accuracy refers to how many of the predicted positive samples are really positive samples. The accuracy is usually used for image classification, and the accuracy is often used for object detection. object detection model mAP@0.5 Compared with the accuracy of the image classification model, although YOLOv5x is up to 95.7, it can also fuse the image classification and object detection algorithms, and adopt the garbage classification consensus voting algorithm to further improve the garbage classification accuracy.

4.4. Fusion voting of object detection and image classification model

In order to improve the recognition rate of garbage classification, this study combines the research results of image classification model and object detection model, and designs a consensus voting algorithm (GCCV) for garbage classification. Consensus voting algorithm (CV) [25] is a method to improve classification accuracy, which is the improvement of majority voting algorithm, which was was proposed by David. Mc Allister et al in 1990. We define *N*: the total number of functionally equivalent voting modules; Each voting module is *{m1, m2, ..., mn}*; *m* is the approval number, which refers to the number of voting modules that agree to output the result. The unanimous voting algorithm is described as follows [26]:

- If the output result is approved by the majority, that is, m > (n + 1)/2, n > 1, then select this result as the output.
- If there is a unique maximum approval number and the approval number is less than (n + 1)/2, then this result is selected as the output.
- If several outputs have the same maximum approval number, a group of results is randomly selected as the output of the program.
- In other cases, voting is invalid, stop, and transfer to the invalid safety state.

Garbage classification consensus voting algorithm (GCCV) is a consensus voting algorithm using image classification model and object detection model as classifiers. It is defined as N = 5, m1 is ResNet, m2 is MobileNetV2, m3 is YOLOv5s, m4 is YOLOv5m, and m5 is YOLOv5x. Through experiments, the accuracy rate of GCCV for garbage classification and recognition is 97.6%. The comparison table of five garbage recognition models and gccv is shown in Table 11.

Table 11. Comparison of recognition rate.

Models and methods	Accuracy (%)
ResNet	81.5
MobileNet	82.9
YOLOv5s	80.4
YOLOv5m	85.9
YOLOv5x	95.7
GCCV	97.6

The comparison curve drawn by Table 11 is shown in Figure 10. It can be intuitively reflected from Figure 10 that the recognition accuracy of the garbage classification consensus voting algorithm gccv is nearly 2% higher than that of the highest YOLOv5x model, which well illustrates the fusion image classification and object detection algorithm, and achieves a relatively ideal effect through the consensus voting algorithm.

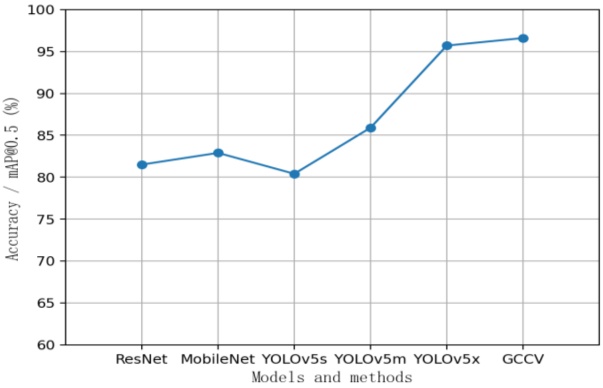
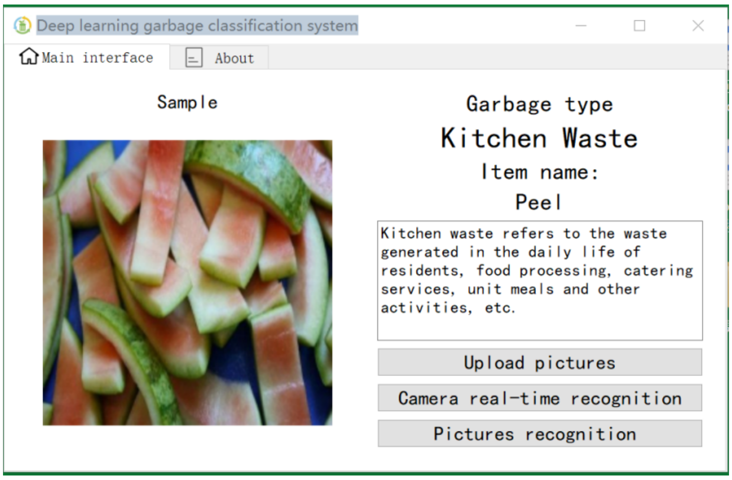


Figure 10. Comparison chart of accuracy rate of identification methods.

4.5. Experimental environment and GUI design

In order to verify the results of the fusion image classification model and target detection model algorithm, after completing the relevant training and testing, we further transplanted to the Raspberry terminal for experiments, and designed a visual interface of the garbage classification system.

● Experimental environment: This experiment uses the Cortex-A72 raspberry 4B + development board with 4 GB memory as the terminal device, and the system is installed as Linux. Raspberry is a basic microcomputer motherboard. Although it is only the size of a card, it has the basic functions of a computer. The specific migration steps include: establishing a virtual environment, configuring the environment required by the system, camera debugging, software downloading, software debugging, etc. the functions are consistent with those of desktop PC.



Kitchen waste refers to the waste generated in the daily life of residents, food processing, catering services, unit meals and other activities, etc.

Figure 11. Main interface of deep learning garbage classification system.

● GUI design：The main interface of the garbage classification system is designed in Python language PyQT5. The main interface is shown in Figure 11. There are three buttons: Upload image button, Camera real-time identification button, and Image identification button. The functions are: 1) Upload image button is used to upload garbage images, and the Image identification button is used to identify garbage images. As shown in Figure 11, the identified image is the peel of kitchen waste. 2) The Camera real-time identification button is used to identify pictures in real time, and the identification results are displayed on the interface.

5. Conclusions and future work

This paper studies a garbage classification system based on deep learning. The deep learning method uses image classification algorithm and object detection algorithm. The image classification algorithm uses ReSNet, MobileNetV2, and the object detection algorithm uses three algorithms of YOLOv5 family. First, the garbage image data is collected and labeled, the model is adjusted, and then the model is trained and tested. Finally, the training results of garbage classification are integrated, and the garbage is identified through the consensus voting algorithm, so that the recognition rate is increased by 2%, and the recognition rate reaches 98%. Finally, the visual garbage classification system is developed to transplant raspberry to achieve the practical level of garbage classification.

In the future, we will continue to improve the deep learning algorithm. There is still room for improvement in the garbage classification and recognition rate. We will continue to enrich the data set to make the garbage classification more accurate, faster and more convenient, reduce the pressure of the sanitation department and workers, and establish a green and beautiful new world.

Acknowledgments

This study was supported by the Natural Science Foundation Project of China (61976118), the Natural Science Foundation Project of Jiangsu Province (BK20180142), and the key topics of the "13th five-year plan" for Education Science in Jiangsu Province (B-b/2020/01/18).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. H. Hu, X. Jiang, X. Liu, R. Ding, S. Ma, B. Wang, Summary of domestic garbage classification and detection based on deep learning, in *2021 7th International Conference on Computer and Communications (ICCC)*, **7** (2021), 858–862. https://doi.org/10.1109/ICCC54389.2021.9674502
2. Y. Deng, Y. Xu, Design of waste classification and recycling system based on tensorflow, *Comput. Knowl. Technol.*, **23** (2021), 50–52.
3. D. O. Melinte, A. M. Travediu, D. N. Dumitriu, Deep convolutional neural networks object detector for real-time waste identification, *Appl. Sci.* **10** (2020), 7301. https://doi.org/10.3390/app10207301
4. S. Li, M. Yan, J. Xu, Garbage object recognition and classification based on mask scoring RCNN, in *2020 International Conference on Culture-oriented Science & Technology (ICCST)*, **6** (2020), 54–58. https://doi.org/10.1109/ICCST50977.2020.00016

5. D. Lin, Z. Chen, M. Wang, J. Zhang, X. Zhou, Design and implementation of intelligent garbage classification system based on artificial intelligence technology, in *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, **13** (2021), 128–134. https://doi.org/10.1109/CICN51697.2021.9574675
6. Z. Wu, D. Zhang, Y. Shao, X. Q. Zhang, X. P. Zhang, Y. Feng, et al., Using YOLOv5 for garbage classification, in *2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, **4** (2021), 35–38. https://doi.org/10.1109/PRAI53619.2021.9550790
7. W. Liu, G. F. Ren, R. S. Yu, S. Guo, J. K. Zhu, L. Zhang, Image-adaptive YOLO for object detection in adverse weather conditions, *arXiv preprint*, (2021), *arXiv:2112.08088v1*.
8. J. M. Feng, M. X. Chu, Y. H. Yang, R. F. Gong, Vehicle information detection based on improved YOLOv3 algorithm, *J. Chongqing Univ.*, **12** (2021), 71–79.
9. G. Yang, Garbage classification system with YOLOV5 based on image recognition, in *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, **6** (2021), 11–18. https://doi.org/10.1109/ICSIP52628.2021.9688725
10. Q. Guo, Y. Shi, S. Wang, Research on deep learning image recognition technology in garbage classification, in *2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, **10** (2021), 92–96. https://doi.org/10.1109/ACCTCS52002.2021.00027
11. Z. Kang, J. Yang, G. Li, Z. Zhang, An automatic garbage classification system based on deep learning, *IEEE Access*, **8** (2020), 140019–140029. https://doi.org/10.1109/ACCESS.2020.3010496
12. G. E. Hinton, S. Osindero, Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.*, **18** (2006), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527
13. D. N. Su, G. T. Cao, Y. N. Wang, H. Wang, H. Ren, Survey of deep learning for radar emitter identification based on small sample, *Comput. Sci.*, **49** (2022), 226–235. https://doi.org/10.11896/jsjkx.210600138
14. A. Krueangsai, S. Supratid, Effects of shortcut-level amount in lightweight ResNet of ResNet on object recognition with distinct number of categories, in *2022 International Electrical Engineering Congress (iEECON)*, (2022), 1–4. https://doi.org/10.1109/iEECON53204.2022.9741665
15. Z. Zhu, W. Zhai, H. Liu, J. Geng, M. Zhou, C. Ji, et al., Juggler-ResNet: A flexible and high-speed ResNet optimization method for intrusion detection system in software-defined industrial networks, *IEEE Trans. Ind. Inf.*, **18** (2022), 4224–4233. https://doi.org/10.1109/TII.2021.3121783
16. M. Hu, Y. Wei, M. Li, H. Yao, W. Deng, M. Tong, et al., Bimodal learning engagement recognition from videos in the classroom, *Sensors*, **22** (2022), 5932–5942. https://doi.org/10.3390/s22165932
17. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778.
18. C. Zhang, T. Yang, J. Yang, Image recognition of wind turbine blade defects using attention-based mobileNetv1-YOLOv4 and transfer learning, *Sensors*, **22** (2022), 6009–6019. https://doi.org/10.3390/s22166009
19. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

20. Q. Luo, J. Wang, M. Gao, Z. He, Y. Yang, H. Zhou, Multiple mechanisms to strengthen the ability of YOLOv5s for real-time identification of vehicle type, *Electronics*, **11** (2022), 2586–2597. https://doi.org/10.3390/electronics11162586
21. L. W. Ye, Z. P. Song, Real time detection method of classroom behavior based on YOLO-v5 improved model, *Changjiang Inf. Commun.*, **7** (2021), 41–45.
22. Z. Li, A. Namiki, S. Suzuki, Q. Wang, T. Zhang, W. Wang, Application of low-altitude UAV remote sensing image object detection based on improved YOLOv5, *Appl. Sci.*, **12** (2022), 8314–8325. https://doi.org/10.3390/app12168314
23. Q. Fu, J. Chen, W. Yang, S. Zheng, Nearshore ship detection on SAR image based on YOLOv5, in *2021 2nd China International SAR Symposium (CISS)*, (2021), 1–4. https://doi.org/10.23919/CISS51089.2021.9652233
24. J. Ieamsaard, S. N. Charoensook, S. Yammen, Deep learning-based face mask detection using YOLOV5, in *2021 9th International Electrical Engineering Congress (iEECON)*, (2021), 428–431. https://doi.org/10.1109/iEECON51072.2021.9440346
25. Z. M. Bao, S. R. Gong, S. Zhong, R. Yan, X. H. Dai, Person re-identification algorithm based on bidirectional KNN ranking optimization, *Comput. Sci.*, **46** (2019), 267–271. https://doi.org/10.11896/jsjkx.181001861
26. X. Liu, Y. Wang, Y. Li, F. Liu, J. Shen, L. Ou, et al., Comparing eight computing algorithms and four consensus methods to analyze relationship between land use pattern and driving forces, *Int. J. Geosci.*, **10** (2019), 12–28. https://doi.org/10.4236/ijg.2019.101002