

A LIGHTWEIGHT MODEL FOR PAVEMENT GARBAGE CLASSIFICATION BASED ON DEEP LEARNING

Guoqiang Chen¹, Wenhui Tong¹, Yanan Cheng¹, Jun Dai¹

¹*School of Mechanical and Power Engineering, Henan Polytechnic University, Jiaozuo 454003, China*

Email: chengq@hpu.edu.cn

Abstract: In response to the requirement of urban environmental health to classify and recycle citizens' garbage, a lightweight model for pavement garbage classification based on deep learning is proposed considering the powerful performance of computer convolutional neural networks in image classification. In order to reduce the scale of the network model to run better on embedded devices, on the basis of the original convolutional neural network framework, the ghost module is reasonably embedded into the lightweight model for pavement garbage classification and the ordinary convolution layers in the path aggregation network (PANet) are replaced by the depthwise separable convolution layers. In order to improve accuracy and real-time performance, the attention mechanism squeeze and excitation layer (SELayer) is embedded to fuse more spatial features. Then, a self-made dataset is built through internet search and manual photography. The self-made dataset contains seven types of common pavement garbage images. Through experimental demonstration, the proposed model can well locate the garbage and identify the types of garbage. And the proposed model has the advantage of small size. The proposed model also achieves 15.58 frames per second (FPS) real-time performance, and the size of the model is only 5.18 mbyte (MB). The FPS increases by 27.5%. The model size is reduced by about two thirds.

Keywords: Garbage Classification; Deep Learning; Lightweight Model; Ghost Module; Attention Mechanism; Depthwise Separable Convolution.

1. Introduction

With the acceleration of urbanization and the boom of the urban population, the dilemma of "urban garbage siege" has become one of the environmental pollution problems in cities. According to the latest report of International Lianhe Zaobao, the global garbage volume will increase by 70% by 2050 [1]. In this context, how to achieve the reduction, recycling and harmlessness of urban garbage has become an important issue for urban management. Experience from developed countries shows that wastes reuse is an important measure to solve this problem [2]. The realization of garbage classification plays a key role in the realization of wastes reuse. Among the garbage, hazardous wastes such as waste fluorescent tubes, waste thermometers, waste medicines and their packaging, and waste photographic papers need to be landfilled. In the meanwhile, there are some useable wastes in the garbage, such as paper, plastic bags, and beverage bottles. If we can separate useful wastes from garbage, not only can we protect the environment, but it can also promote social and economic development. For example, one ton of waste paper can produce 850 kilograms of good paper, which can save 300 kilograms of wood; one ton of waste plastic bottles can be made into 600 kilograms of unleaded gasoline or diesel.

Every autumn and winter, many pavements are covered with a lot of fallen leaves. Fallen leaves are neither pollution nor domestic waste. But our current treatment method is to classify them as municipal solid waste, and then they will be burned or buried. Burning fallen leaves will pollute the atmosphere, and burying fallen leaves will cause a waste of resources. Fallen leaves can be used to make fertilizer. If we sort them out of garbage and process them, they will become organic fertilizers that can provide nutrients to other plants.

Garbage classification is beneficial to society, economic and ecology [3]. It helps with reducing the consumption of land resources and promoting the garbage recycling. It can also reduce environmental pollution. Therefore, it is of great significance to classify garbage. However, the currently classifying garbage mainly relies on manual identification. Manual sorting is not only inefficient, but also easy to damage the health of workers [4-5]. To completely solve this problem, we should use artificial intelligence technology to classify garbage [6]. The intelligent classification and recycling will inevitably become the future garbage treatment mode.

In recent years, deep learning has developed rapidly and has penetrated into many fields. Deep learning-based technology is being significantly used to solve many real-world problems in order to make

the system more accurate and efficient. These techniques based on deep learning can be implemented in almost every domain like medical, health, security, automation, and many more [7]. How to use deep learning to classify garbage efficiently has become a research hotspot.

Automatically identifying garbage scattered area can improve the cleaning efficiency and reduce secondary pollution. Zeng et al. [8] proposed a method for monitoring garbage distribution in large areas with airborne hyper spectral data. They completed the detection of the location and size of garbage regions by using unsupervised region proposal generation algorithm selective search and none maximum suppression. You et al. [9] proposed an automatic garbage scattered area detection model based on remote sensing images. Experiments show that their proposed model can obtain the location information of all garbage scattered areas accurately and rapidly. Therefore, it is effective and convenient to use deep learning technology to automatically identify the garbage scattered area.

Automatically identifying garbage scattered area is only the first step of garbage classification. After locating the garbage scattered area, the type of garbage needs to be identified. Only by distinguishing the types of various wastes in the garbage can useful wastes such as fallen leaves and plastic bottles be separated from the garbage. Yuan et al. [10] designed a 23-layer convolutional neural network model to accurately classify the garbage in real time. Zhihong et al. [11] applied the region proposal generation and the visual geometry group -16 (VGG-16) model to identify plastic bottles in garbage under complex background. Mittal et al. [12] presented a novel smartphone app, called SpotGarbage, which detects and coarsely segments garbage regions in a user-clicked geotagged image. Ramalingam et al. [13] proposed a method for detecting and locating pavement defects and garbage using deep convolutional neural network. Based on ResNet-34 algorithm, Kang et al. [14] proposed a garbage classification algorithm. They also further optimized the network structure of the algorithm from the multi feature fusion of input images, the feature reuse of the residual unit, and the design of a new activation function. Their garbage classification algorithm can classify the garbage quickly and accurately. Zhang et al. [15] proposed a novel two-stage waste recognition-retrieval algorithm, which can automatically classify garbage into four categories.

The methods of garbage classification based on deep learning can be roughly divided into: the method based on residual neural network (ResNet), the method based on dense convolutional network (DenseNet), the method based on combination of convolutional neural network and transfer learning, and the method based on one-stage object detection

algorithm. The one-stage object detection network has a simple structure. At the same time, it has the advantage of high detection accuracy. Therefore, it has extensive research and application value in the field of real-time target detection [16]. The you only look once (YOLO) is one of the most classic one-stage object detection algorithms [17-18]. In YOLO algorithms, the convolution neural network is used to extract features without region proposal network. Faster detection speed is achieved by inputting the features into the regression network to directly return the class probability and object bounding box. And the whole algorithm adopts the end-to-end network to evaluate the whole image by regression, so that the real-time performance can be guaranteed reliably.

In summary, although many scholars have designed algorithms that can automatically locate garbage scattered areas and identify the type of garbage, they take up a lot of memory. So, they cannot run on embedded devices. In this regard, this study proposes a lightweight model for pavement garbage classification that can run on embedded devices. The proposed model in the study can automatically identify seven types of common pavement garbage, including beverage bottle, pericarp, dry cell, disposable fast-food box, plastic bag, cigarette butt and fallen leaf.

2. A Lightweight Model for Pavement Garbage Classification

The model proposed in the study is based on the you only look once v5 (YOLOv5), which has good performance in fast object identification.

The study optimizes the structure of the model to make it suitable for garbage classification. In order to reduce the model size, the ghost module is added to the backbone network. Moreover, by replacing the ordinary convolution layers with the depthwise separable convolution layers, the number of parameters in the PANet network is further reduced, thereby reducing the model size. In order to ensure detection accuracy while reducing the size of the model, the attention mechanism SELayer is added to the backbone network. And the overall structure of the model is shown in Figure 1.

Firstly, the image features are extracted from the backbone network, and then they are input into the PANet network. Finally, the input feature maps are processed by up-sampling and down-sampling to predict the category and bounding box.

This study replaces the BottleneckCSP with a ghost module of 1×1 . And the ghost module of 2×2 is used to replace the convolution layer for down-sampling and feature extraction. The model uses the attention mechanism SELayer to obtain the importance of each feature channel. And the ordinary convolution is replaced with the depthwise

separable convolution. $\times N$ in Figure 1 means the module is stacked N times.

For example, the GhostBottleneck $\times 3$ means that GhostBottleneck module is stacked 3 times.

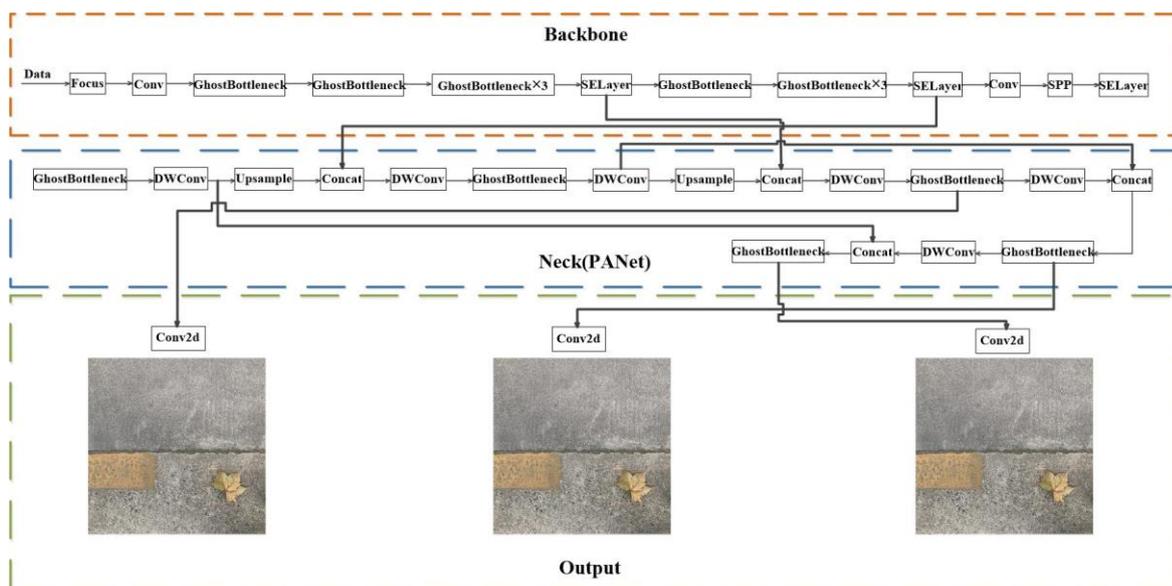


Figure 1: The Overall Structure of the Model

2.1 Ghost Module Embedding

The ghost module is the core module of GhostNet, a new lightweight neural network proposed by Huawei Noah's Ark Laboratory [19]. Integrating the ghost module into the existing convolution neural network can reduce the computation cost and the number of parameters of the model. The reason is that the ghost module can generate more feature maps through extremely small cost operations. In order to solve the memory and speed problems of the garbage classification model, we replace the

original backbone network BottleneckCSP module of YOLOv5 with the ghost module with the step size 1 and the step size 2 respectively.

The ghost module is shown in Figure 2. In the ghost module, the feature maps are convoluted to get the $H' \times W' \times n$ feature maps, and then the n channel feature maps are mapped linearly. The GhostBottleneck is formed by stacking two ghost modules. The first ghost module is used to expand the number of channels of the feature maps. The second ghost module reduces the number of channels to match the shortcut path.

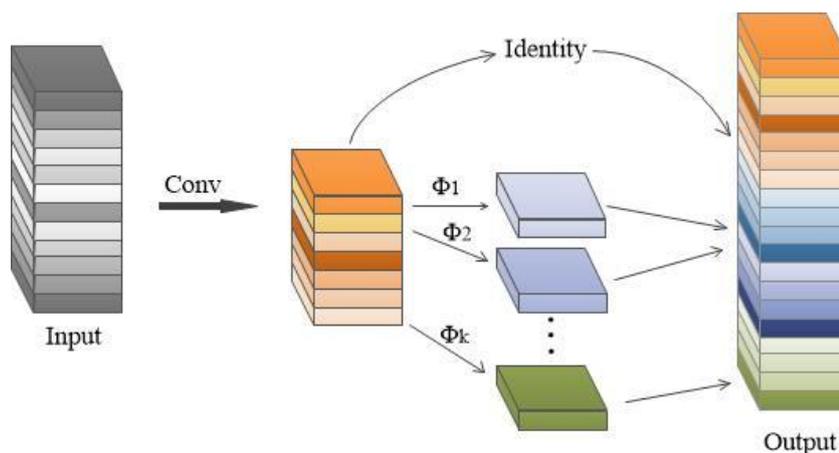


Figure 2: The Ghost Module

2.2 Convolution Layers Changing

The ordinary convolution layers in the model proposed in the study are replaced by the depthwise separable convolution layers. The parameter number in the PANet network is further reduced.

Therefore, the model proposed in the study is further lightweight.

The depthwise separable convolution is shown in Figure 3. The depthwise separable convolution is divided into two steps: the depthwise convolution and the pointwise convolution.

Unlike conventional convolution, one convolution kernel of depthwise convolution is responsible for one channel, and one channel is convolved by only one convolution kernel. And the depthwise convolution is performed entirely in a two-dimensional plane. Each channel of the input layer is convolved independently.

Therefore, the feature information of different channels is not effectively utilized in the same spatial location. However, the Pointwise convolution can combine the previous feature maps in the depth direction. And the kernel size of the Pointwise convolution is $1 \times 1 \times N$ (N represents the depth of the upper output).

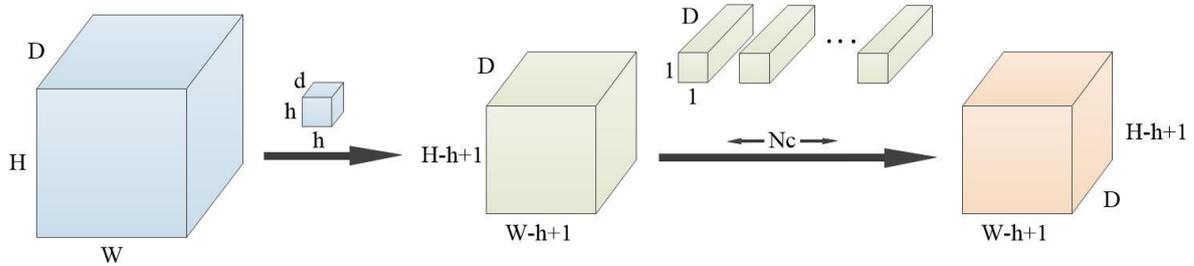


Figure 3: The Depthwise Separable Convolution

$$P_{conv} = (H - h + 1) \times (W - h + 1) \times N_c \times h \times h \times D \tag{1}$$

$$P_{DWconv} = (H - h + 1) \times (W - h + 1) \times N_c \times h \times h \times D + N_c \times 1 \times 1 \times D \times (H - h + 1) \times (W - h + 1) \tag{2}$$

$$\frac{P_{DWconv}}{P_{conv}} = \frac{1}{N_c} + \frac{1}{h^2} \tag{3}$$

For an ordinary 2D convolution, the input size is $H \times W \times D$, N_c represents the convolution kernel number, the kernel size is $h \times h \times D$, and h is an even number. For a random image with three channels, the ordinary convolution parameters are P_{conv} and the depthwise separable convolution parameters are P_{DWconv} .

Therefore, for ordinary 2D convolution, the training time of ordinary convolution is h^2 times of the training time of depthwise separable convolution.

They are equivalent to equation 1 and equation 2 respectively. When the depthwise separable convolution is performed, there are many output channels. In other words, the N_c number in equation 3 is much greater than h .

2.3 Selayer Embedding

Embedding ghost module and replacing the ordinary convolution layers with the depthwise separable convolution layers will reduce the detection accuracy of the model. In order to maintain the model detection accuracy, the attention mechanism SELayer is added to the backbone network.

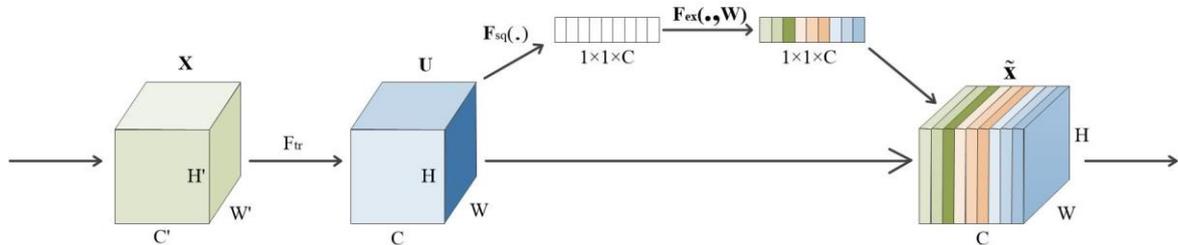


Figure 4: The Selayer Module

The SELayer module is shown in Figure 4. Firstly, the spatial dimension is compressed by squeeze operation to get the $1 \times 1 \times C$ feature maps in the SELayer module. Secondly, the importance of each channel is predicted by using the full connection layer.

Then, the importance of different channels is obtained through the excitation operation. Finally, in order to promote important features and suppress unimportant features, a weighting operation is performed to multiply the activation values of each learned channel.

3. Experimental Analysis

3.1 Experimental Environment

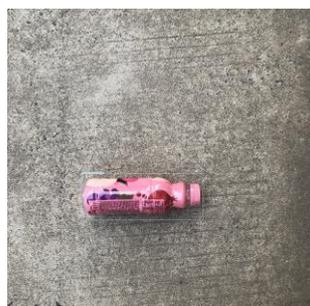
In order to train the neural network model, the study builds a training platform that uses Ubuntu 18.04 system and GeForce GTX 2060Ti.

Other configurations include CUDA 10.1, deep learning acceleration library CUDNN v7.6 and Python interface environment.

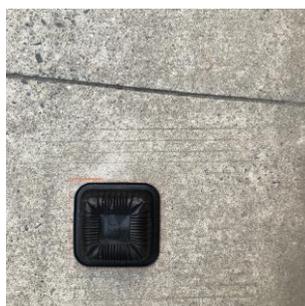
And the model uses the Pytorch deep learning framework and the Python programming language.

Table 1. The Number of Garbage Images in Each Category

| | Trainset | Testset | Dataset |
|--------------------------|----------|---------|---------|
| beverage bottle | 179 | 45 | 224 |
| pericarp | 600 | 161 | 761 |
| dry cell | 257 | 65 | 322 |
| disposable fast-food box | 161 | 41 | 202 |
| plastic bag | 289 | 73 | 362 |
| cigarette butt | 221 | 56 | 277 |
| fallen leaf | 235 | 59 | 294 |



(a) beverage bottle



(b) disposable fast-food box



(c) fallen leaf



(d) pericarp



(e) plastic bag



(f) cigarette butt



(g) dry cell



(h) cigarette butt & dry cell



(i) fallen leaf & cigarette butt

Figure 5: Performance of Our Model

3.2 Experimental Dataset

At present, there are few open-source datasets about garbage. Researchers generally use self-photographing and internet search to establish datasets. In 2020, Huawei held the “garbage classification Challenge Cup” competition and released a dataset on household garbage. The dataset has 4 categories and 44 subcategories, with a total of 14964 images. Our study is mainly aimed at the common pavement garbage. Therefore, when making our dataset, this study only uses common pavement garbage images in the dataset released by Huawei. And considering that fallen leaves are very common on the pavement, 294 images of fallen leaves are taken manually when making our dataset. Finally, these images form a dataset about common pavement garbage. The self-made dataset contains 2442 images. As shown in Table 1, the self-made dataset is divided into seven categories: beverage bottle, pericarp, dry cell, disposable fast-food box, plastic bag, cigarette butt and fallen leaf.

Among them, the trainset contains 1942 images and the testset contains 500 images.

3.3 Experimental Results

In order to clearly show the impact of different modules, this study compares the results through ablation experiments. As shown in Table 2, this study compares different modules from four aspects: mAP (%)0.50, mAP (%)0.50:0.95 FPS and model size. The ghost module reduces the model size without decreasing the detection accuracy. By adding the attention mechanism SELayer and depthwise separable convolution, the model size is decreased from 9.60MB to 5.18MB; and mAP (%) 0.50 and mAP (%) 0.50:0.95 are reduced by 1.52% and 2.25% respectively. The experimental results show that the size of the model is reduced by optimizing the structure. The detection accuracy of the model has not been greatly affected. And the results of target recognition of the lightweight model for garbage classification are shown in Figure 5.

Table 2. Performance Results of Different Modules on the Self-made Dataset

| | YOLOv5s | The lightweight model | | |
|---------------------------------|---------|-----------------------|---------|--------|
| Ghost module | × | ✓ | ✓ | ✓ |
| SELayer | × | × | ✓ | ✓ |
| Depthwise separable convolution | × | × | × | ✓ |
| mAP (%)0.50 | 0.6696 | 0.6829 | 0.6832 | 0.6677 |
| mAP (%)0.50:0.95 | 0.4542 | 0.4704 | 0.4693 | 0.4479 |
| FPS | 65.77 | 75.38 | 69.5 | 34.87 |
| Model Size | 14.1 MB | 9.60 MB | 9.69 MB | 5.18 B |

4. Discussions

The increasing amount of household garbage not only brings resistance to the development of the city, but also seriously damages the ecological environment. Changing the promiscuity of garbage by classification is an important prerequisite for harmless treatment. This study presents a lightweight model for pavement garbage classification based on YOLOv5s, which realizes the automatic location and classification of seven types of common pavement garbage.

Based on YOLOv5, the backbone network of the proposed model reasonably embeds the ghost module and the attention mechanism SELayer. In the proposed model, the ordinary convolution layers of the PANet network were replaced by the depthwise separable convolution layers. The model proposed in the study can identify the types of common pavement garbage. And it is more lightweight than other models.

The model can also identify and locate fallen leaves on the pavement well. The fallen leaf is a common recyclable waste. Therefore, the model proposed in the study lays a theoretical foundation for the intelligent collection of fallen leaves and better achieves the recycling of fallen leaves.

There are still many problems to be solved in further research due to the limited capacity and time in this study although some achievements have been made in this research:

(1) There are other kinds of garbage on the pavement. In the future, we can further add other kinds of garbage images into the dataset.

(2) In order to adapt to the high complexity of pavement garbage, the following work can continue to optimize the network model to achieve the identification and classification of multi-scale and small targets.

5. Conclusions

The study presents a lightweight model for pavement garbage classification based on YOLOv5s. Firstly, in order to reduce the model size, the ghost module is introduced to replace the original module to extract features. Secondly, the depthwise separable convolution is embedded to further reduce the model size. Thirdly, the attention mechanism SELayer is embedded reasonably to fuse more features in space and extract multi-scale spatial information to improve the detection accuracy. Finally, the different module is verified by ablation experiments on the self-made dataset. Through training on the self-made dataset which contains fallen leaves images, the proposed model shows a good performance in classifying pavement garbage.

Acknowledgements

This work is supported by Fundamental Research Funds for the Universities of Henan Province (No.NSFRF200401), the Key Technology R&D Program of Henan Province of China (No. 212102210045, No. 182102310706) and National Natural Science Foundation of China (No. U1304525).

References

- [1] Kang Z, Yang J, Li G, et al. An automatic garbage classification system based on deep learning [J]. IEEE Access, 2020, 8: 140019-140029.
- [2] Gamba R J, Oskamp S. Factors influencing community residents' participation in commingled curbside recycling programs [J]. Environment and Behavior, 1994, 26(5): 587-612.
- [3] Tong Y, Liu J, Liu S. China is implementing "Garbage Classification" action [J]. Environmental Pollution, 2020, 259: 113707.
- [4] Zhao G, The status quo and countermeasures of municipal solid waste sorting and processing [J], Environmental and Development, 2020,32(10):32-35. (in Chinese)
- [5] Ma J, Research on the classification of municipal solid waste [J]. Environmental and Development, 2020,32(09):31-34. (in Chinese)
- [6] Qi X, Gong X, Li J, et al. Deep learning based garbage image processing and recognition [J]. Computer Knowledge and Technology, 2021,17(09):20-24. (in Chinese)
- [7] Panwar H, Gupta P K, Siddiqui M K, et al. AquaVision: automating the detection of waste in water bodies using deep transfer learning [J]. Case Studies in Chemical and Environmental Engineering, 2020, 2: 100026.
- [8] Zeng D, Zhang S, Chen F, et al. Multi-scale CNN based garbage detection of airborne hyperspectral data [J]. IEEE Access, 2019, 7: 104514-104527.
- [9] You T, Chen W, Wang H, et al. Automatic garbage scattered area detection with data augmentation and transfer learning in SUAV low-altitude remote sensing images [J]. Mathematical Problems in Engineering, 2020, 2020.
- [10] Yuan J, Nan X, Li C, et al. Research on real-time multiple single garbage classification based on convolutional neural network [J]. Mathematical Problems in Engineering, 2020, 2020.
- [11] Zhihong C, Hebin Z, Yanbo W, et al. A vision-based robotic grasping system using deep learning for garbage sorting [C]//2017 36th Chinese Control Conference (CCC). IEEE, 2017: 11223-11226.
- [12] Mittal G, Yagnik K B, Garg M, et al. Spotgarbage: smartphone app to detect garbage using deep learning [C]//Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016: 940-945.
- [13] Ramalingam B, Hayat A A, Elara M R, et al. Deep learning based pavement inspection using self-reconfigurable robot [J]. Sensors, 2021, 21(8): 2595.
- [14] Kang Z, Yang J, Li G, et al. An automatic garbage classification system based on deep learning [J]. IEEE Access, 2020, 8: 140019-140029.
- [15] Zhang S, Chen Y, Yang Z, et al. Computer vision based two-stage waste recognition-retrieval algorithm for waste classification [J]. Resources, Conservation and Recycling, 2021, 169: 105543.
- [16] Liu J, Meng W. Review on single-stage object detection algorithm based on deep learning [J]. Aero Weaponry, 2020, 27(3):44-53. (in Chinese)
- [17] Joseph Redmon, Santosh Divvala, Ross Girshick, et al. You only look once: 319unified, real-time object detection. In IEEE Conference on Computer Vision and Pattern 320Recognition, pages 779-788, 2016. 321
- [18] Joseph Redmon, Ali Farhadi. YOLO9000: better, faster, stronger, In IEEE Conference on 322Computer Vision and Pattern Recognition. pages 7263-7271, 2017. 323
- [19] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.