Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30292

MFF-MobileNetV2: Garbage Image Classification Based on Multi

Mingzhe Dai

Shandong University School of Airspacescience Engineerin, Shandong University (Wihai), Weihai, China
daimingzhe01@outlook.com

Abstract. The problem of garbage classification is one of the main issues in current society. In order to address the low efficiency of manual sorting in garbage classification, the difficulty of traditional techniques in recognizing garbage image features, and the requirement for lightweight and high-precision models, this study proposes a lightweight multi-scale feature fusion technology (MFF-MobileNetV2). The network uses MobileNetV2 as the backbone network and introduces a multi-scale fusion module to capture local details, textures, and global contour features under different receptive fields through parallel dilated convolution. At the same time, it uses channel attention mechanism to enhance key information and suppress noise. In addition, this article also designs a data augmentation strategy targeting the characteristics of garbage images to improve the robustness of the model. Experiments on the TrashNet dataset showed that MFF-MobileNetV2 achieved a classification accuracy of 94.52% on the test set, while maintaining a low inference delay (0.334 ms) and a model size of only 17.12 MB. This study provides a feasible solution for achieving efficient and accurate garbage image classification on resource limited edge devices.

Keywords: Multi-scale fusion, Data augmentation, Lightweight

1. Introduction

The problem of garbage classification is one of the most important issues in garbage problems. In daily life, the lack of garbage classification often leads to the arbitrary burial and incineration of garbage, creating more severe environmental problems [1-3]. However, traditional manual sorting has high error rate and low efficiency, with a per capita processing capacity of less than 1 ton [4-6]. Therefore, facing the challenge of garbage classification, some machine learning based intelligent detection algorithms are necessary [7-9].

This article focuses on the problem of occlusion, contamination, deformation and overlap existing in the garbage image classification. Taking the TrashNet dataset as the experimental basis, a lightweight multi-scale feature fusion network based on MobileNetV2 (MFF-MobileNetV2) is constructed. This paper introduce a parallel dilated convolution module to model the multi-scale feature and combine channel attention mechanism to enhance the expression of key information.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ30292

Meanwhile, an appropriate data enhancement strategy based on the characteristics of garbage images is designed to enhance the robustness of models.

The main contributions of this article are as follows.

(1) Innovation in Lightweight Multi Scale Fusion Network Architecture. This paper proposes a lightweight network structure based on MobileNetV2. The multi-scale fusion module in this paper adopts parallel dilated convolution to capture local detail and texture information as well as global contour feature information at different dilation rates (1, 2, 3). This design can solve the problem of feature loss caused by occlusion and deformation in garbage images. And the whole structure of this network still focuses on compressing the parameters. The size of the model is only 17.12MB, which provides basis for edge deployment.

(2) Intelligent Feature Optimization Mechanism. In order to solve the problem that the key features of garbage images are easily disturbed by noise, this paper introduces a channel attention mechanism (based on SENet structure) to adaptively weight and optimize multi-scale fused features. This mechanism can enhance the expression of key information such as transparent glass area and suppress irrelevant background noise by generating attention weights based on global average pooling and fully connected layers. Experimental results show that this mechanism can significantly improve the robustness of the model to occluded and contaminated images. The accuracy improvement is about 3.2 percentage points.

(3) Data Enhancement Strategy for Garbage Image Characteristics. To enhance the generalization performance of the model in real-world scenarios, this paper designed a data augmentation strategy based on the characteristics of garbage images. The data augmentation strategy designed in this paper includes grid occlusion, random rotation (± 30 °), brightness adjustment (0.8-1.2 coefficients) and random erasing. These methods simulate the problem of garbage images occlusion, uneven lighting and deformation, and enhance the diversity of data sets through online enhancement. Finally, this strategy achieved 94.52% accuracy on the TrashNet data set through early stopping and Dropout regularization in training. These contributions make the accuracy and efficiency of garbage image classification improve from the network design, feature optimization to the data level, and provide feasible solutions for edge computing applications.

2. MFF-MobileNetV2

2.1. Model architecture

This work propose a multi-scale feature fusion lightweight network (MFF-MobileNetV2) as shown in Figure 1.

DOI: 10.54254/2755-2721/2026.TJ30292

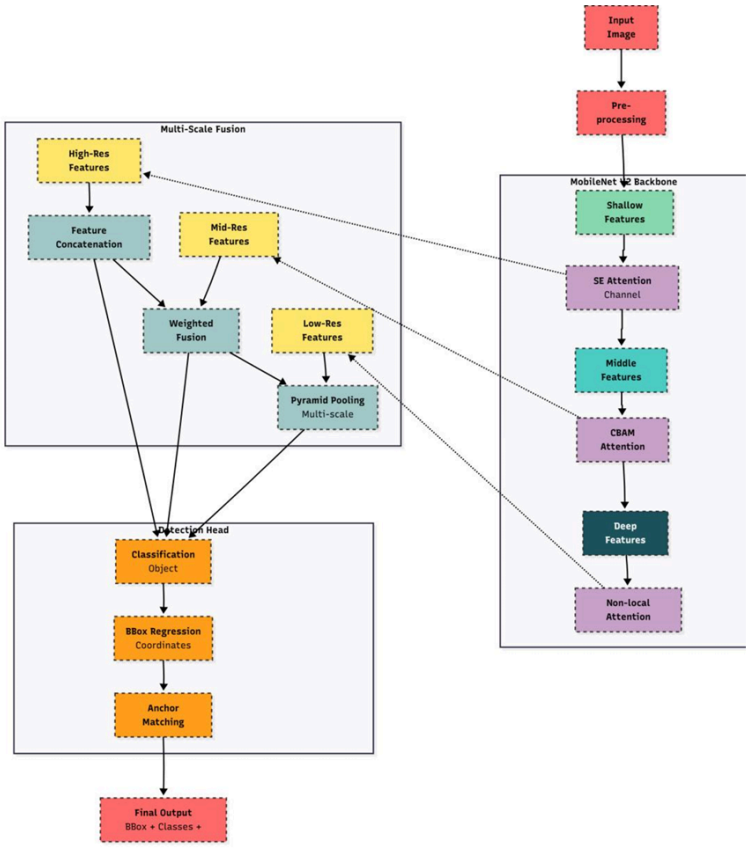


Figure 1. MFF MobileNetV2 (picture credit: original)

The multi-scale feature fusion lightweight network (MFF-MobileNetV2) takes a lightweight MobileNetV2 pre trained network as the backbone network, and its weights are initialized based on the ImageNet data set to achieve the effect of lightness and feature extraction; The core idea of this structure is to design a multi-scale fusion module after the end feature layer (outputting a 1280 channels feature map) of MobileNetV2 (depthwise separable convolution), in order to capture local details, texture and global contour features of different scales through parallel dilated convolution and channel attention, and solve the problem of difficulty in feature recognition caused by occlusion, contamination and deformation of garbage image; The whole network design focuses on lightness and high precision as shown in Figure 1; The whole workflow of the network is that the input layer accepts standardized garbage images, MobileNetV2 backbone network completes feature extraction, and then multi-scale fusion module completes feature compression and multi-scale extension, and then output the six types of garbage probability through global average pooling and fully connected layer; In addition, this paper designs data augmentation strategy based on garbage image characteristics, grid occlusion, random erasing, etc., to enhance the robustness of model in real scene, and finally achieve high classification accuracy (94.52%) and low parameter amount (model volume is only 17.12 MB), providing feasible solution for edge device.

2.2. Multi scale fusion module

Multi scale fusion module is the core part of MFF MobileNetV2, which aims to enhance the ability of model to extract local details, texture and global contour through different receptive fields, and fusion features output by backbone network MobileNetV2 in two steps.

DOI: 10.54254/2755-2721/2026.TJ30292

Feature compression Perform a 1x1 convolution operation on the 1280 channels feature map output by MobileNetV2 backbone network, and compress the number of channels to 256. This step aims to reduce the amount of computation without losing key information. Its expression is as follows.

$$F_{compressed} = Conv_{(1\times1)}(F_{input}) \tag{1}$$

Among them, Finput is the input feature map (1280 channels), Conv1 × 1 is a 1x1 convolutional layer, and Fcompressed is the output 256 channel feature map.

Multi-scale feature capture and fusion. Parallel application of three dilated convolutions on the compressed feature map to replace the traditional large-sized convolution kernel, while achieving the purpose of enlarging receptive field without increasing much parameters. The dilation rate of each dilated convolution is different to capture multi-scale contextual information. The formula for calculating receptive field (RF) is as follows.

$$RF = 1 + \sum_{i=1}^{n}(k_i - 1) \times d_i \tag{2}$$

Among them, ki represents the size of the i-th convolutional kernel, di is the corresponding hole rate, and n is the number of convolutional layers.

Note that the kernel size of three-way dilated convolution is set to 3 × 3 in this study, but the dilation rates are set to 1, 2, and 3 respectively to gradually enlarge the receptive field and cover from local to global feature. After the parallel convolution, can concatenate the three feature maps in channel dimension to get 768 fused feature map. Subsequently, channel attention module (based on SENet structure) is added to weight and optimize the concatenated feature to enhance the key feature (such as transparent area of glass) and suppress the noise. The expression of attention weight is shown as expression (3).

$$\omega = \sigma(W_2 \bullet ReLU(W_1 \bullet GAP(x))) \tag{3}$$

Among them, x is the input feature (768 channels), GAP represents global average pooling operation, Wi and W2are the fully connected layer weight, and σ is Sigmoid activation function. The finally weighted feature will be used in the subsequent classification task, which can greatly improve the robustness of the model in occluded and deformed images. This module realizes the coordination of light and high precision with the complementation of multi-scale information and attention driven, and provides an effective feature fusion method for garbage image classification.

3. Experiment and analysis

3.1. Dataset and evaluation indicators

For model training and evaluation, this paper uses TrashNet dataset. There are totally 2527 garbage images in six categories, including glass, paper, cardboard, plastic, metal and trash. The division of data set takes into account the ratio of training set, validation set and test set are 70%, 15% and 15% respectively, so as to fairly and effectively evaluate the model evaluation. In the data preprocessing stage, all input images are resized to the size of 224 × 224 pixels and pixel value normalized (intensity ranging to [0,1]), so as to meet the input requirements of MobileNetV2 backbone network. In addition, since the occlusion, uneven lighting and deformation of garbage images are commonly

DOI: 10.54254/2755-2721/2026.TJ30292

seen in actual scene, this paper designs special data augmentation strategy (see section 3.2 for details) to improve the generalization ability of model in complex scene.

The evaluation indicators mainly include classification accuracy, weighted average F1 score, model volume (MB) and single graph inference time (ms). Accuracy and F1 score are used to evaluate the classification accuracy, and model volume and inference time are used to evaluate lightweighting and deployment efficiency.

3.2. Experimental setup

This experimental part uses NVIDIA GPU as experimental hard ware, and uses PyTorch as software environment. In the training process, AdamW optimizer is adopted, the initial learning rate of which is 0.001, and the value of weight decay is 1e-4. StepLR strategy is adopted to adjust the learning rate in the process of learning, and the learning rate is reduced to 1/10 of the original in every 10 times of training (cycle/epoch). In the process of regularization, the value of Dropout is set to 0.4 to suppress over-fitting; At the same time, early stopping mechanism is adopted. If the accuracy of validation set does not increase for five consecutive cycles, the training process will be stopped prematurely. The value of batch size is set to 16 by default, and is set to 8 if the memory of GPU is insufficient.The maximum number of training epochs is 30 times.

Data Augmentation strategy(as shown in Table 1)is designed according to the characteristics of garbage images, which is intended to enhance the model's ability to resist occlusion, rotation and lighting: specifically including grid occlusion (partly occluded scene is simulated), random rotation (direction invariance within ± 30 ° is enhanced), brightness adjustment (brightness coefficient is randomly changed within 0.8-1.2 to cope with lighting difference), and random erasing (local areas of image are randomly erased to pollute scene and cope with different lighting conditions). The enhancement operation is adopted online in the training process, and the multidimensional transformation of data set is used to enhance the feature extraction ability of model.

Table 1. Implementation of data enhancement

Enhanced type	Parameter Settings	effect
Grid occlusion	grid_size=56, prob=0.5	Simulate garbage occlusion/pollution scenarios
random rotation	±30°	Enhance angle robustness
Brightness adjustment	±20%	Resist changes in light exposure
Random erase	p=0.2, scale=(0.02,0.1)	Dealing with local occlusion

3.3. Experimental results

MFF-MobileNetV2 obtains 94.52% classification accuracy and 0.9452 weighted average F1 on the test set, and the size of model is only 17.12 MB, and the delay of single image inference is only 0.334 ms. In the training process, when the accuracy of validation set remains at above 94% after 15 cycles, it can be seen that the model converges quickly and has good generalization ability; Early stop mechanism is triggered at the 25th cycle.

DOI: 10.54254/2755-2721/2026.TJ30292

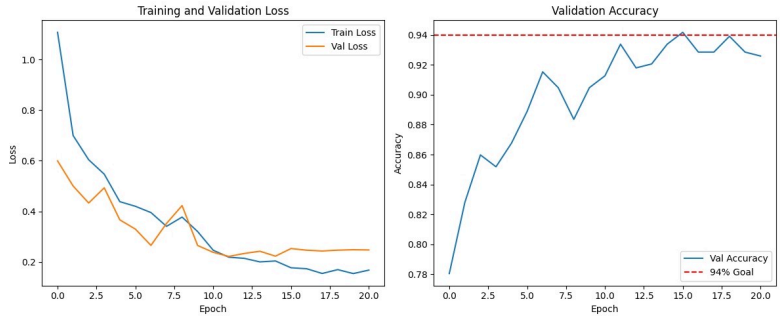


Figure 2. Training curve (MFF MobileNetV2) (picture credit: original)

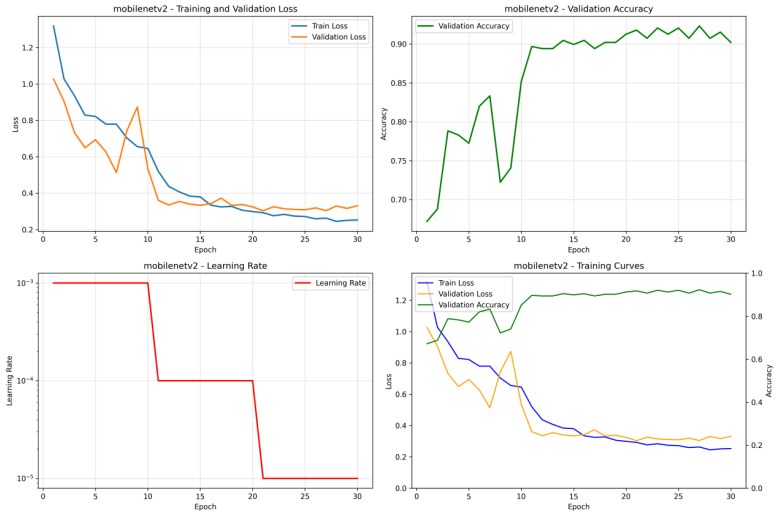


Figure 3. Training curve (MobileNetV2)

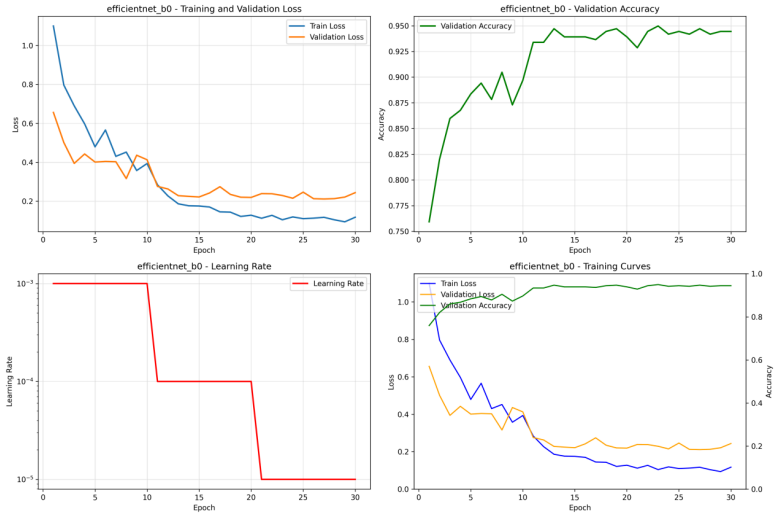


Figure 4. Training curve (EfficientNet-B0) (picture credit: original)

DOI: 10.54254/2755-2721/2026.TJ30292

Table 2. Comparative experiment

model	accuracy	Model volume	Inference latency
MobileNetV2 [10]	92.3%	8.72MB	20ms
EfficientNet-B0 [9]	94.8%	15.6MB	10.48ms
MFF-MobileNetV2	94.52%	17.12 MB	0.334 ms

In addition, this study conducted comparative experiments between MFF-MobileNetV2 and MobileNetV2 with EfficientNet-B0, and the results are shown in Table 2. This study shows through systematic performance evaluation that MFF MobileNetV2 achieves the best balance in accuracy, model size, and latency on the TrashNet dataset. Further analysis of the training curve reveals that MFF-MobileNetV2 (Figure 2) converges rapidly and has significantly better stability than the comparison model within 15 cycles. MobileNetV2 (Figure 3) shows an overfitting trend in the later stage, while EfficientNet-B0 (Figure 4) converges slowly. This verifies the effectiveness of the multi-scale fusion module and attention mechanism in balancing accuracy and efficiency, providing a reliable basis for edge deployment.

4. Conclusion

This article aims at the recognition of occlusion, contamination, deformation and overlap of garbage image classification. This paper takes MobileNetV2 as the backbone network, and takes parallel dilated convolution module to capture local detail and global contour feature under different receptive fields. Channel attention mechanism (SENet structure) is used to improve the key feature expression and suppress irrelevant noise. Data augmentation methods such as grid occlusion and random erasing are designed for garbage image characteristics, which enhance the model robustness in complex scene. The experimental results show that this method obtains the classification accuracy of 94.52% and weighted F1 score of 0.9452 on TrashNet dataset. At the same time, the model size is only 17.12 MB, and the inference delay of single image is only 0.334 ms. Compared with the comparison model, it has advantages of less memory, fast speed and accuracy. It provides a technical solution for the intelligent garbage classification in the environment with limited resources.

Of course, there are still some limitations in this study. Firstly, TrashNet dataset used in this paper is small (only 2527 images, 6 categories), which restricts the generalization ability of model in more garbage categories and real-world scenes. Secondly, the model has not been end-to-end deployed and tested on embedded devices such as Raspberry Pi and Jetson Nano. The stability and efficiency of the model in actual edge environment still need to be further verified.

In the next step, this study will carry out the following two aspects of research: First, validate the model on larger scale and more categories of garbage image datasets (such as Waste Pictures Dataset), to improve the generalization ability of model; The second is to popularize the practical application of model in embedded system. Further compression of model size and reduction of inference latency will be achieved by quantitative techniques (such as INT8 quantification) and hardware collaborative design, to promote the large-scale application of intelligent garbage classification technology in practical scenarios. In addition to providing a new design idea for lightweight waste classification model, this study also provides a preliminary exploration for the application of edge computing in the field of environmental protection.

DOI: 10.54254/2755-2721/2026.TJ30292

References

[1] Tran, H. M., Le, T. M., Pham, H. V., Vu, M. T., & Dao, S. V. T. (2024). An Integrated Learning Approach for Municipal Solid Waste Classification.IEEE Access, 12, 176569-176585.doi: 10.1109/ACCESS.2024.3495982
[2] Walia, J. S., Haridass, K., & Pavithra, L. K. (2025). Deep learning innovations for underwater waste detection: An in-depth analysis. IEEE Access, 13, 88917–88929.
[3] Sayem, F. R., Islam, M. S. B., Naznine, M., et al. (2025). Enhancing waste sorting and recycling efficiency: Robust deep learning-based approach for classification and detection. Neural Computing and Applications, 37, 4567–4583.
[4] Tomas, J. P. Q., Guimong, M. L., Leria, G. A., & Morelos, I. L. (2024). Multiple garbage detection using Mask R-CNN in TensorFlow for household waste classification. In Proceedings of the 2023 6th International Conference on Computational Intelligence and Intelligent Systems(pp. 147–153).
[5] Feng, W., Yan, W., & Xie, J. (2022). Garbage object detection method based on improved Faster R-CNN. In Proceedings of the 5th International Conference on Computer Science and Software Engineering(pp. 415–419).
[6] Liu, T., Guo, X., & Pei, X. (2021). Research on recognition of working area and road garbage for road sweeper based on Mask R-CNN neural network. In Proceedings of the 4th International Conference on Control and Computer Vision(pp. 76–82).
[7] Xie, X. J. (2022). Garbage image classification method based on improved convolution neural network and long short-term memory network. In Proceedings of the 5th International Conference on Computer Science and Software Engineering(pp. 657–663).
[8] Zhang, J., Kang, L., & Xie, X. (2024). Recyclable garbage detection algorithm under occlusion conditions. In Proceedings of the 2023 4th International Conference on Machine Learning and Computer Application(pp. 815–819).
[9] Oza, P., Agrawal, S., Kapadia, M., & Raotole, O. (2025). Solid waste classification using deep neural network: A transfer learning approach.Earth Science Informatics, 18, 255. https: //doi.org/10.1007/s12145-025-01743-x
[10] Yu, L., Mo, F., Lai, Y., & Tian, L. (2023). Edge computing-based intelligent garbage classification and recognition application. In 2023 4th International Conference on Computers and Artificial Intelligence Technology (CAIT)(pp. 136–142).