ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20220420124810
Original scientific paper

An Improved ResNet-50 for Garbage Image Classification

Xiaoxuan MA, Zhiwen LI, Lei ZHANG*

Abstract: In order to solve the classification model's shortcomings, this study suggests a new trash classification model that is generated by altering the structure of the ResNet-50 network. The improvement is divided into two sections. The first section is to change the residual block. To filter the input features, the attention module is inserted into the residual block. Simultaneously, the downsampling process in the residual block is changed to decrease information loss. The second section is multi-scale feature fusion. To optimize feature usage, horizontal and vertical multi-scale feature fusion is integrated to the primary network structure. Because of the filtering and reuse of image features, the enhanced model can achieve higher classification performance than existing models for small data sets with few samples. The experimental results show that the modified model outperforms the original ResNet-50 model on the TrashNet dataset by 7.62% and is more robust. In the meanwhile, our model is more accurate than other advanced methods.

Keywords: attention module; garbage classification; multi-scale feature fusion; ResNet

1 INTRODUCTION

The problem of environmental pollution has gotten more serious as a result of fast economic growth and the increase of people's living conditions [1]. The pollution of the environment produced by the fast growth in waste production has harmed the planet and all of its species [2]. China began requiring individuals to sort and discard their trash in 2019. Front-end collection is strongly dependent on people's sense of responsibility in this scenario. Even yet, it might be difficult for persons who lack experience to appropriately categorize all sorts of household garbage. The waste recycling situation has increasingly deteriorated with the fast growth in medical waste generation, particularly in the context of the present COVID-19 epidemic. Finding an effective intelligent automated trash sorting system is therefore of tremendous academic and practical importance.

The amount of data has grown exponentially in the last decade as the mobile internet has developed, and the accumulation of big data has given the major basis for deep learning. Meanwhile, deep learning is undergoing rapid progress as a result of advances in computer technology in data collecting, storage, and processing [3]. Deep learning has now pervaded all elements of computer vision, yielding impressive results in image classification, target detection, and semantic segmentation [4]. Deep learning-based systems have several benefits over conventional machine learning methods, including extensive modeling capabilities [5] and end-to-end learning methodology. Deep learning relies heavily on data [6], but because trash categorization is such a new field, there's no standardized dataset available at the time for neural network learning. As a result, in 2016, Mindy Yang and Gary Thung created the TrashNet dataset [7] for rubbish picture categorization. Following that, the amount of work being done on using deep learning to categorize trash photos has steadily risen. Despite the fact that the TrashNet dataset has been widely utilized, for its tiny quantity of photos and lack of feature information, various efforts based on it have not yielded satisfactory results.

Deeper neural networks are increasingly being used for deep learning-based garbage image classification. Deepening the network structure can assist achieve stronger nonlinear representation capabilities [8] and better fitting of complicated feature information by simplifying the task at each layer. TrashNet, on the other hand, is a small dataset with a single context that has less feature information, fewer data samples, and more inter-class similarity. In this case, it is difficult to enhance classification performance by increasing network depth in this situation. Using a supervised approach, this work trains and evaluates various deep classification models, then improves on the best-performing model. It can extract improved image features for classification by modifying the network structure and introducing multi-scale feature fusion without deepening the network depth. On small-scale datasets, the approach suggested in this research can achieve improved classification accuracy.

The remainder of the paper is laid out as follows. Section 2 examines some of the relevant garbage classification research. The proposed enhanced version of the classification model is described in Section 3. Experiments in Section 4 are used to analyse the impact of the improvement. The conclusion is given in Section 5.

2 RELATED WORKS

Image classification is continuously diversifying due to the rapid growth of artificial intelligence, with the prominent models being AlexNet [9], VGG [10], Inception [11], and ResNet [12]. The researchers ran multiple trials with these models and attempted to make improvements in order to acquire better outcomes. Garbage classification, being a relatively young subject, lacked a consistent dataset for neural network learning in its early stages. In 2016, Yang and Thung collected and structured the TrashNet dataset [7], and identified it using SIFT and CNN-based SVM, with accuracy of 63% and 22%, respectively. Many academics have since used this dataset in their rubbish picture categorization research.

Mandar Satvilkar, for example, utilized the RF algorithm to classify waste photos from the TrashNet data set with 62.61% [23]. Bernardo S. Costa et al. classified six types of waste photos in the TrashNet data set using the KNN method, with an accuracy of 88% [24]. In another research, on the TrashNet dataset, Kennedy Tom's suggested OscarNet network [13] (fine-tuned by vgg19)

obtained 88.42% classification accuracy. Bernardo S. Costa et al. suggested a fine-tuned AlexNet network with 91% accuracy in October 2018 [14], as well as a fine-tuned VGG16 network with 93% accuracy. Chu et al. [28] proposed a Multi-layer Hybrid deep-learning System (MHS) to sort waste automatically. A high-resolution camera with sensors was used to extract image features and other features. The MHS used a CNN and a multi-layer perceptron (MLP) to combine the image with features and classify wastes as recyclable or not. They achieved an accuracy higher than 90% under two different testing scenarios. RahmiArda Aral et al. tested multiple classic networks on the TrashNet data set. They achieved 89% accuracy using Inception-Resnet V2 and 89% accuracy using DenseNet121 [15]. Adedeji and Wang [16] proposed an intelligent waste classification system that uses ResNet-50 model to extract the features and SVM to classify them into different categories such as glass, metal, paper, etc. The TrashNet dataset was used and an accuracy of 87% was achieved. The authors Vo et al. [17] collected 5904 images belonging to three classes- Organic, Inorganic, and Medical waste to create a new dataset called VN-trash dataset. They developed a deep neural network model for trash classification named DNN-TC that outperformed the existing ResNext model and achieved an accuracy of 94% and 98% for TrashNet and VN-trash dataset, respectively.

3 RESEARCH METHOD

3.1 Basic Model Selection

In order to select a suitable image classification model, the cifar-10 dataset is used to compare the advantages and disadvantages of AlexNet, VGG, Inception, and ResNet models. The training in 200 iterative steps under the same training strategy is performed and the training results are shown in Tab. 1. From the accuracy and error of the validation data in the table, it can be seen that the models in the ResNet series perform best, with ResNet-50 and ResNet-152 excelling in both accuracy and error loss, respectively. Because the deeper the model, the larger the number of parameters, the higher the hardware requirement, and the slower the computation speed, ResNet-50 was chosen as the base model in this garbage classification algorithm after a trade-off between classification accuracy and computation speed. Fig. 1 depicts the structure of one of the ResNet-50 models.

Table 1 Comparison of accuracy and loss of different networks on the cifar-10 dataset

Network	Train		Val		Parameters
	Accuracy	Loss	Accuracy	Loss	
AlexNet	0.8644	0.2673	0.7550	1.5667	9,639,178
VGG-16	0.9777	0.1764	0.8831	0.9599	33,638,218
VGG-19	0.9817	0.1752	0.8892	0.8149	38,947,914
InceptionV2	0.9699	0.1823	0.9114	0.4823	11,264,111
InceptionV3	0.9754	0.2540	0.9157	0.4763	23,851,784
ResNet-34	0.9841	0.1755	0.9221	0.4510	21,315,338
ResNet-50	0.9874	0.1709	0.9226	0.4111	38,213,514
ResNet-101	0.9852	0.1716	0.9204	0.4205	75,196,810
ResNet-152	0.9868	0.1650	0.9279	0.3923	105,666,442

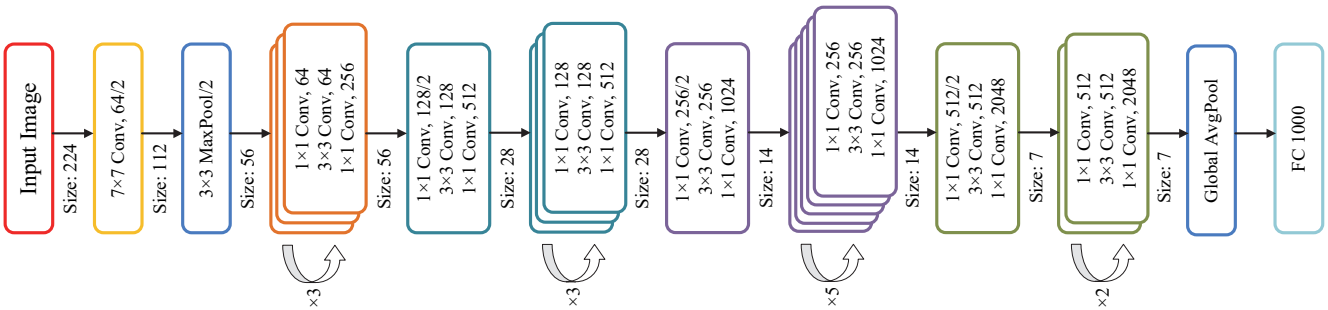

Figure 1 The structure of ResNet-50 network

3.2 Model Modification

3.2.1 Residual Block Modification

Woo et al. introduced CBAM [18], a novel convolutional attention module that adaptively directs the neural network model where to focus information. Much study has been done on network depth and width in order to increase the performance of deep neural networks. As illustrated in Fig. 2, CBAM is a typical attention mechanism module that has two sequential submodules, the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), that execute adaptive filtering of input features in the channel and spatial dimensions, respectively.

The channel attention module first receives the input characteristics. To efficiently compute channel attention, the tensor all-channel feature matrix is first calculated using global max pooling and global average pooling to obtain two matrices. For learning weight optimization, these two weight matrices are supplied into the same multilayer perceptron. The two output components are then added together to form a channel weighting module. A sigmoid activation function compresses the data to between 0 and 1, multiplying it with the input features. As illustrated in Eq. (1) and Eq. (2):

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

$$F' = M_C(F) \times F \quad (2)$$

where F denotes the input features; *AvgPool* and *MaxPool* represent the global average pooling and the global max pooling, respectively; *MLP* represents a multilayer perceptron containing two fully connected layers with ReLU as the first layer activation function; σ denotes the sigmoid activation function; M_C is the channel attention module, and F' represents the features obtained after passing the channel attention module.

The feature maps enter the spatial attention module after passing through the channel attention module. The average pooling and max pooling are used to process the input in the spatial attention module, however compressed sampling is done in the channel dimension. Following the pooling procedure, two two-dimensional spatial matrices are formed, which are then concatenated in the channel dimension. After that, a convolutional layer is utilized, using a single output channel and a sigmoid activation function. Eqs. (3) and (4) demonstrate that:

$$M_S(F') = \sigma(f([AvgPool(F'); MaxPool(F')])) \tag{3}$$

$$F'' = M_S(F') \times F' \tag{4}$$

where M_S is the spatial attention module; f denotes the operation of the convolutional layer, and F'' represents the features obtained after passing through the spatial attention module.

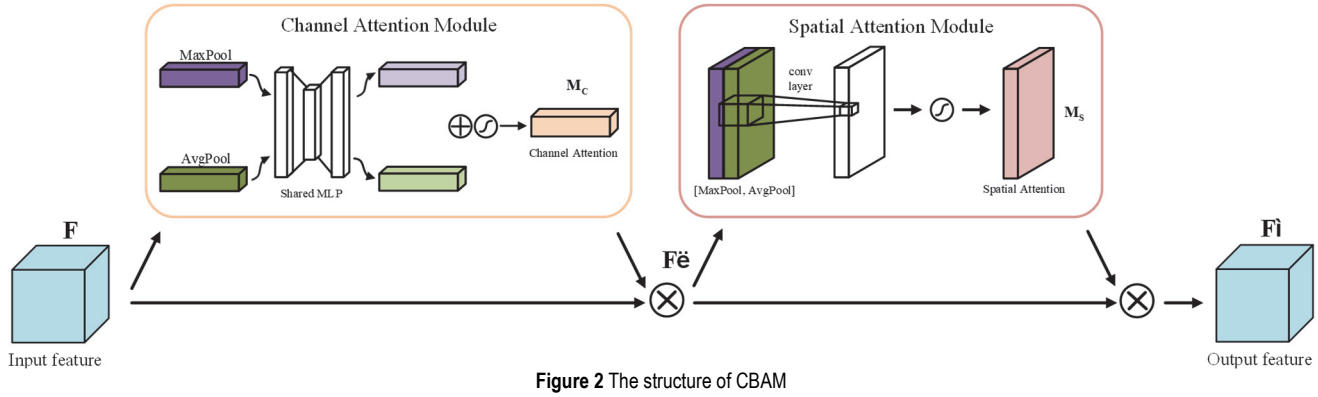

Figure 2 The structure of CBAM

The CBAM module is integrated in each of ResNet-50's underlying residual blocks in this work. When downsampling is required, the classic ResNet model employs an 1×1 convolutional kernel with a step size of 2 to execute the convolutional operation, which invariably results in information loss. As a result, convolutional operations with a convolutional kernel of 1 × 1 and a step size of 2 will be avoided in this article. Fig. 3 depicts the new base block structure, with (a) the original structure and (b) the modified structure.

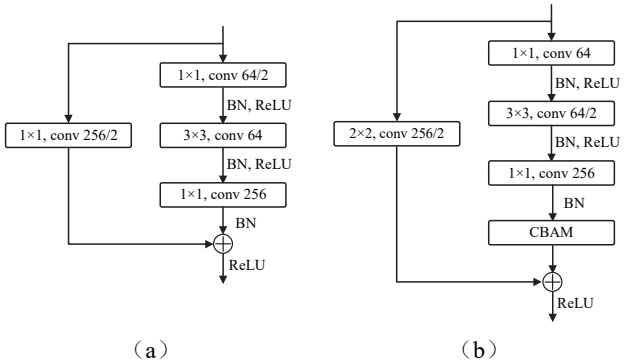

Figure 3 Structure diagram of the base residual block for the original ResNet-50 and the modified

The 1 × 1 convolution for downsampling in the residual component is also changed to be implemented in the 3 × 3 convolution layer, in addition to introducing the CBAM module.When downsampling is required, the direct mapping component additionally adjusts the convolution kernel size from 1 × 1 to 2 × 2.

3.2.2 Residual Block Modification

The shallow component of the ResNet-50 network is optimized in this section to extract features from multiple dimensions of the image for fusion using parallel operations. The max pooling process at the start of the network was adjusted and replaced with three simultaneous feature extraction operations. The original pooling process is preserved in the first route, i.e., a maximum pooling of size 3 × 3 and a step size of 2. The second route performs one convolution operation with a convolution kernel size of 3 × 3 and a step size of 2. The third way executes two convolution operations with 3 × 3 convolution kernels, however the second convolution has a step size of 2. Finally, the outputs of the three parallel routes are combined to complete feature fusion in multiple dimensions and extract more image features, hence boosting the model's performance.

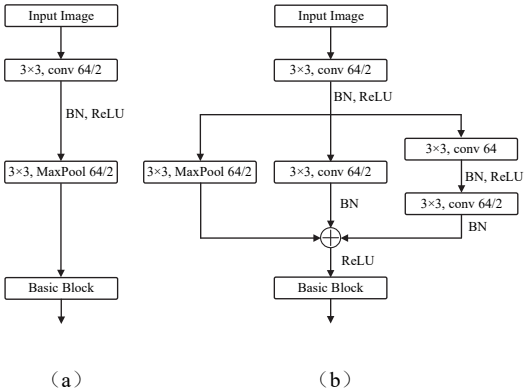

Figure 4 Structure diagram of shallow network for the original ResNet-50 and the modified

Fig. 4 shows a comparison of the model's shallow structure before and after the modification.

Vertical information fusion and horizontal information fusion for multi-scale feature fusion are proposed. The residual block, which is also the smallest base unit in the ResNet network, is the network's essential structure as shown in Fig. 3a. The entire network structure is made up of four huge residual blocks, each of which is made up of numerous base residual blocks. The four major residual blocks, like the ResNet-50 structure, include 3, 4, 6, and 3 base residual blocks, respectively.

After a downsampling procedure, the output features of the first big residual block are fused with the input of the third large residual block, and the size changes from 75 × 75 × 64 to 38 × 38 × 512 (with a network input size of 300 × 300 as an example). Similarly, downsampling processes fuse the output features of the first big residual block with the input and output of the fourth large residual block, resulting in dimensions of 19 × 19 × 1024 and 10 × 10 × 2048, respectively.

The output of the second residual block is passed to the input and output of the fourth large residual block after a downsampling operation, similar to the method used in the first residual block, and the output of the third residual block is passed to the output of the fourth large residual block after a downsampling operation where the downsampling operation consists of two convolution operations, starting with a convolution layer that adjusts the length and width, with a convolution kernel of 2 × 2 and a step size of 2. The second layer of convolution uses a 1 × 1 convolution kernel to adjust the number of channels. Multiple features are fused by the summation operation. Fig. 5 depicts the specific operation flow.

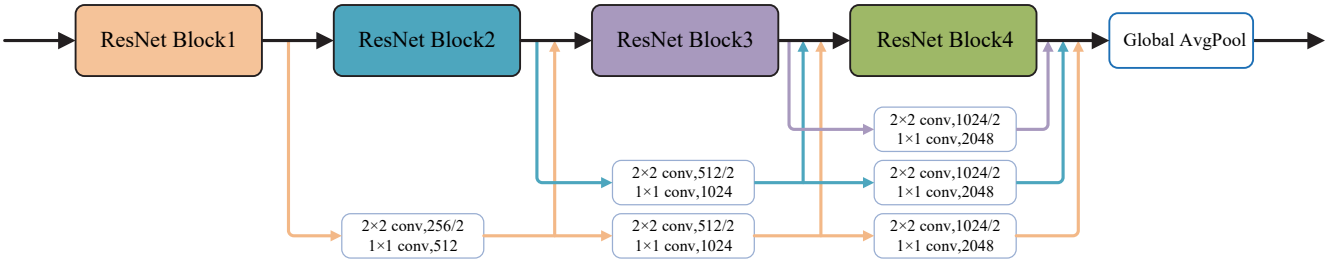

Figure 5 Schematic diagram of vertical fusion of multi-scale features.

4 RESULTS AND DISCUSSION

4.1 Data Set

The TrashNet dataset [7] was utilized in this study, and it comprises RGB photographs of six different types of garbage, with only one type of garbage appearing in each image. Glass, cardboard, metal, paper, plastic, and trash are the six types of waste. There are presently 2527 photos in the dataset, with a distribution of 501 for glass, 594 for paper, 403 for cardboard, 482 for plastic, 410 for metal, and 137 for trash for each category. These pictures were created by mounting the object on a white poster board and photographing it using natural and/or artificial light. All of the images have been reduced in size to 512 × 384 pixels. Fig. 6 depicts representative photos from the TrashNet dataset's six categories.

Figure 6 Trash samples in the TrashNet dataset: (a) glass; (b) cardboard; (c) metal; (d) paper; (e) plastic; (f) trash

The data set is separated into training and validation sets in an 8:2 ratio in this paper. Before being fed into the network, the photos are cropped to 300 × 300 pixels and their pixel values are normalized between 0 and 1.

4.2 Evaluation Index

In order to verify the classification performance of the improved ResNet network, evaluation metrics need to be set. The commonly accepted evaluation metrics in academia are accuracy, loss, and recall. In this paper, accuracy and loss are used as evaluation metrics and time complexity and model robustness are discussed.

When the model is trained to a later stage, the parameters will stabilize, and the accuracy and loss will fluctuate in a small range. Therefore, the average accuracy and loss of the last ten epochs are taken as the absolute accuracy and loss. The expressions are as follows.

$$Accuracy_{\text{avg}} = \frac{\sum_{e=E-9}^{E} Accuracy_e}{10} \tag{8}$$

$$Loss_{\text{avg}} = \frac{\sum_{e=E-9}^{E} Loss_e}{10} \tag{9}$$

where $Accuracy_e$ denotes the validation classification accuracy obtained by the validation set after the e-th epoch of training, and $Accuracy_{\text{avg}}$ denotes the average classification accuracy obtained from the last 10 epochs of validation; $Loss_e$ represents the loss obtained by the validation set after the e-th epoch of training, and $Loss_{\text{avg}}$ is the average loss obtained from the last 10 epochs of validation; E indicates the total number of epochs needed to train a model.

4.3 The Parameter Settings

AdamW [19] optimizer is used to train the model to train 250 epochs with an initial learning rate of $1e^{-3}$, weight decay factor of $1e^{-9}$, and batch size of 10. The learning rate decay strategy is exponential decay, and for every 1.3 epochs, the learning rate will be 0.96 times the previous one. In addition, due to the small number of images in this dataset, data enhancement techniques are used to expand the training samples, including rotate, crop, scale flip, and other transformations. In order to ensure that the experimental results reflect only the performance of the network structure and thus verify the effectiveness of the proposed method, data augmentation with identical relevant settings is used in the training process. The hyperparameter settings are shown in Tab. 2.

Table 2 Data enhancement parameter settings

Project	Settings
shear_range	0.1
zoom_range	0.1
width_shift_range	0.1
height_shift_range	0.1
horizontal_flip	True
vertical_flip	True
rotation_range	20
fill_mode	nearest

The computer configuration of the experimental platform is a GPU deep learning server with two 24-core 2.2 GHz processors, 64 G of RAM, and 16G of RTX5000 professional graphics card. The operating system is Windows 10, the python version is 3.7, and the deep learning framework and version is Tensorflow 2.3.0.

4.4 Experimental Results Analysis

Firstly, we compared the classification performance of three network models: the original ResNet-50 network, the ResNet-50-A model obtained by adding the improvements proposed in subsection 3.2.1 of this paper, and the ResNet-50-B model obtained by adding the modifications proposed in subsection 3.2.2 based on the ResNet-50-A model as shown in Tab. 3. Finally, the model is compared with other state-of-the-art models.

Table 3 Experimental comparison models

Model	Modifications
ResNet-50-A	Embedded CBAM module + modified downsampling method
ResNet-50-B	ResNet-50-A + horizontal and vertical multi-scale fusion

4.4.1 Comparison of Loss, Accuracy and Time Complexity

Tab. 4 shows the average validation loss, validation accuracy, time taken for single-image (S-I) validation, and the number of parameters calculated for the last ten epochs of the training process for the three network models with the same dataset and training strategy. Fig. 7 shows the changes in the validation loss and validation accuracy of the three networks trained for 250 epochs. The three networks tend to stabilize after about 125 epochs of training. Fig. 8 shows the confusion matrix of the actual and predicted labels for the three network models by heat map patterns. It is evident from both error and accuracy that every improvement proposed in this paper can positively impact the classification task. The ResNet-50-A model obtained by adding the CBAM module and modifying the downsampling method is more accurate than the original ResNet-50 model by 3.94%, while the validation loss is reduced by nearly double. And the ResNet-50-B model obtained by adding all the improvements is higher than the ResNet-50-A model with 3.68% accuracy and higher than the original ResNet-50 model with 7.62% accuracy, and this model obtained the least validation loss.

The number of parameters increases due to the increasing complexity of the proposed model structure from the original ResNet-50 model to the ResNet-50-A model to the ResNet-50-B model. The ResNet-50-A model increases the time complexity by 32.6% with a 3.94% accuracy improvement, but the ResNet-50-B network achieves a 7.62% accuracy improvement with a 40.5% increase in time complexity compared to the original ResNet-50. This shows that the improved model can significantly improve network classification performance at the cost of a relatively small increase in time complexity.

Table 4 Average validation loss, accuracy, number of parameters and time spent on single image validation for the three networks

Model	Valloss	Valaccuracy	Parameters	S-I Val time / ms
ResNet-50	0.8039	84.46%	38,205,318	45.56
ResNet-50-A	0.4472	88.40%	48,965,222	60.43
ResNet-50-B	0.3371	92.08%	71,555,366	64.01

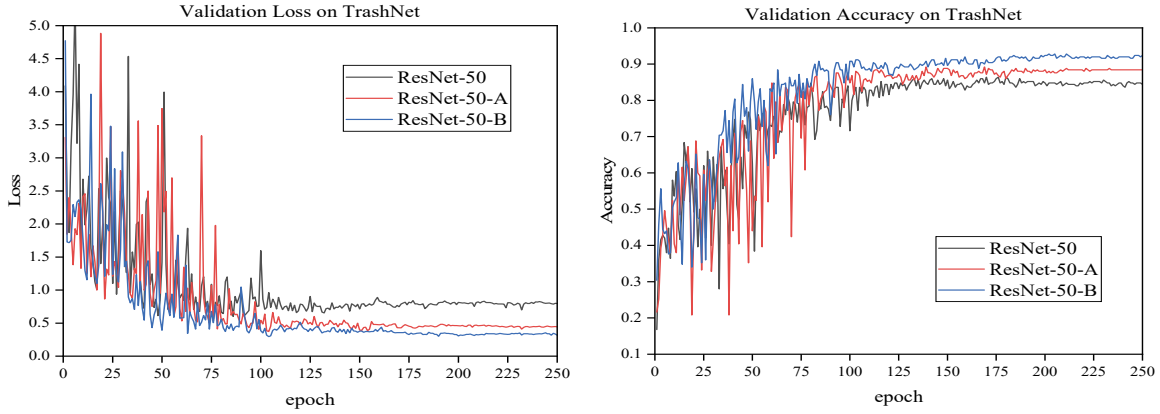

Figure 7 Variation of validation loss (left) and validation accuracy (right) during training of the three networks

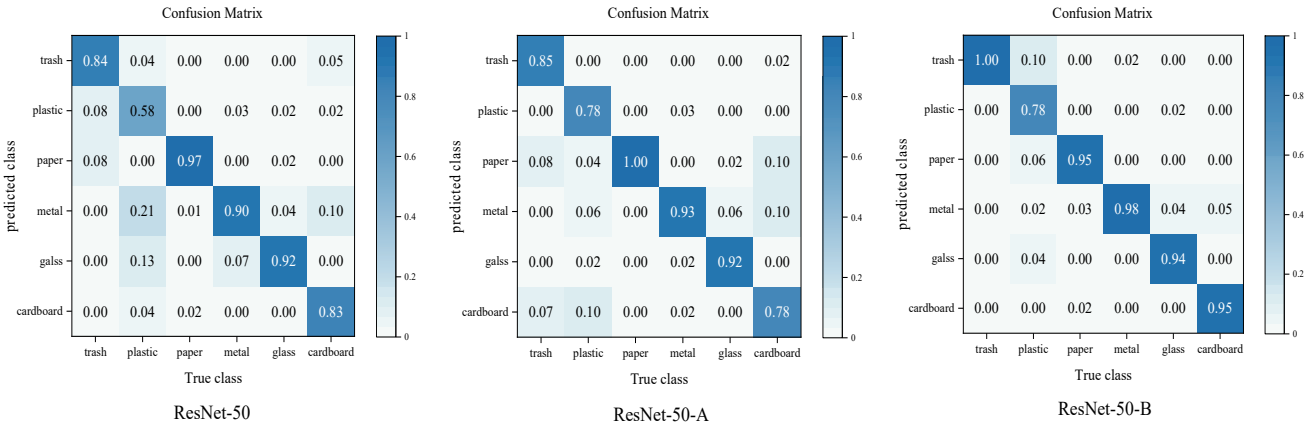

Figure 8 Confusion matrix of actual labels and predicted labels for three network models

4.4.2 Comparison of Robustness

In addition, several studies have shown that neural networks are susceptible to external disturbances. Network models can make incorrect predictions when the ambient light around an object changes or when the object is occluded [20, 21]. For garbage classification, the most common situation in real life is incomplete or ambiguous garbage samples. Therefore, in this paper, some experiments are conducted to test the robustness of ResNet-50, ResNet-50-A and ResNet-50-B for the occlusion case.

As shown in Fig. 9, we refer to [22] to divide nine regions in the image, i.e., nine occlusion cases are considered. Nine new test sets were created based on the different occlusion positions. The position of number 1 of all images of the new test set 1 is masked by the gray rectangle with *RGB* = (192, 192, 192), the position of number 2 of all images of the new test set 2 is masked, and so on.

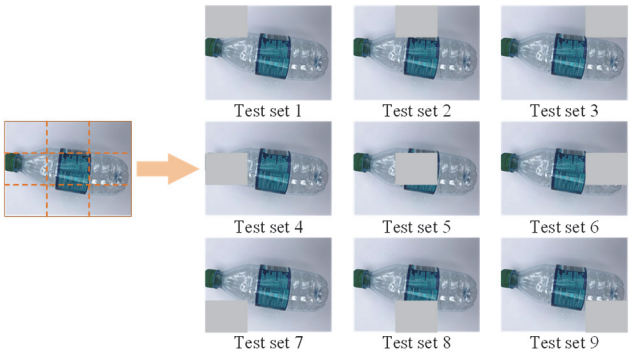

Figure 9 Nine new test sets were created by masking the images in the original validation set, depending on the masking position

These nine new test sets are then used to test our three previously trained network models. Fig. 10 shows the overall accuracy of the three models on the nine test sets. It can be seen that the original ResNet-50 performs the worst, which offers poor robustness. And the ResNet-50-B model shows the best performance on the original image set, which indicates that the ResNet-50-B model has good robustness and further proves its effectiveness. In practical applications, the performance of garbage image classification may be more affected by the robustness of the model. Hence, the model proposed in this paper has more practical significance in improving classification stability.

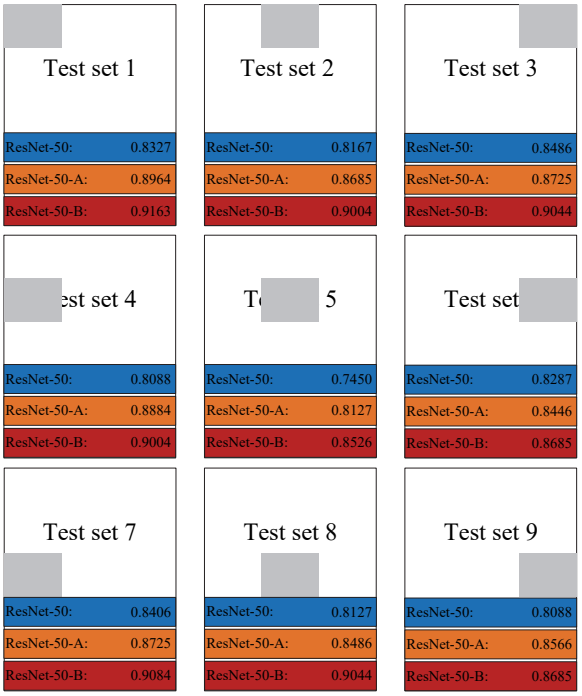

Figure 10 Test accuracy of three network models on 9 new test sets

4.4.3 Comparison with some other Advanced Models

For the experiments compared with other methods, we selected four machine learning methods, SVM, XGB, RF, KNN, and ten deep learning models, and the comparison results are shown in Fig. 11.

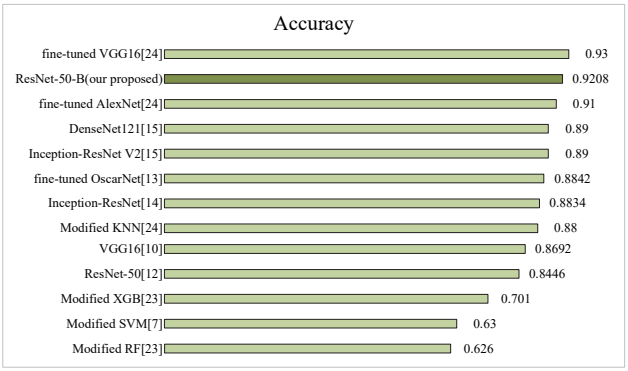

Figure 11 Accuracy comparison of ResNet-50-B with some advanced models

Among them, the fine-tuned VGG16 proposed by B. S. Costa et al. has slightly higher accuracy than the model in this paper, but the result was obtained by transfer learning. Instead of initializing the weights to train from scratch, it is a weight initialization using the pre-trained weights obtained by training on the ImageNet dataset. In contrast, our model is trained on the TrashNet dataset by randomly initializing the weights. In [15], the authors used the VGG16 network with random initialization weights to start training on the Trashnet dataset and finally obtained a classification accuracy of 76.94%. Meanwhile, we trained the VGG16 network with the BN layer added using the training strategy of this paper. We got a final classification accuracy of 86.92%, which is still 5.16% less than the accuracy of the proposed network in this paper. Therefore, our model outperforms most of the current state-of-the-art network models at the network level.

5 CONCLUSION

For the garbage classification task, an improved model based on ResNet-50 is proposed in this paper. The model is enhanced in feature screening, downsampling, and feature fusion. Several currently popular classification models were selected by standard training datasets, and ResNet-50 was chosen as the fundamental network model in the trade-off between accuracy and time complexity. It was then improved by adding an attention mechanism, modifying the downsampling method, and adding horizontal and vertical multi-scale feature fusion.

The original ResNet-50 model was tested with an accuracy of 84.46% by experimenting on the classical garbage dataset TrashNet. The accuracy of the ResNet-50-A model with the addition of the attention module and the modified downsampling method was improved to 88.4%, while the loss was reduced by almost half. Finally, the accuracy of the ResNet-50-B model with the addition of multiscale feature fusion to the ResNet-50-A model was improved to 92.08%, and the loss was further reduced. In testing the model robustness experiments, better robustness is shown in our model. Compared with some related excellent methods, the proposed model can provide a higher classification accuracy.

However, improving the accuracy will inevitably increase the network size and thus the computational effort. So the idea of the subsequent research is to improve the classification accuracy without increasing the computation.

Acknowledgments

This work is supported in part by the National Key Research and Development Program of China under Grant of No. 2020YFB2103604 and No. 2020YFF0305504.

6 REFERENCES

[1] Rajamanickam, R. & Nagan, S. (2018). Assessment of Comprehensive Environmental Pollution Index of Kurichi Industrial Cluster, Coimbatore District, Tamil Nadu, India - a Case Study. *Journal of Ecological Engineering*, *19*(1). https://doi.org/10.12911/22998993/78747

[2] Deng, Z., Weng, D., Chen, J., Liu, R., Wang, Z., Bao, J., Wu, Y. et al. (2019). AirVis: Visual analytics of air pollution propagation. *IEEE transactions on visualization and computer graphics*, *26*(1), 800-810. https://doi.org/10.1109/TVCG.2019.2934670

[3] Khellal, A., Ma, H., & Fei, Q. (2018). Convolutional neural network based on extreme learning machine for maritime ships recognition in infrared images. *Sensors*, *18*(5), 1490. https://doi.org/10.3390/s18051490

[4] De Oliveira, D. C. & Wehrmeister, M. A. (2018). Using deep learning and low-cost RGB and thermal cameras to detect pedestrians in aerial images captured by multirotor UAV. *Sensors*, *18*(7), 2244. https://doi.org/10.3390/s18072244

[5] Enguehard, J., O'Halloran, P., & Gholipour, A. (2019). Semi-supervised learning with deep embedded clustering for image classification and segmentation. *IEEE Access*, *7*, 11093-11104. https://doi.org/10.1109/ACCESS.2019.2891970

[6] Yan, M. (2019). Adaptive learning knowledge networks for few-shot learning. *IEEE Access*, *7*, 119041-119051. https://doi.org/10.1109/ACCESS.2019.2934694

[7] Yang, M. & Thung, G. (2016). Classification of trash for recyclability status. *CS229 project report*, *2016*, 3.

[8] Peng, J., Kang, S., Ning, Z., Deng, H., Shen, J., Xu, Y., Liu, L. et al. (2020). Residual convolutional neural network for predicting response of trans arterial chemoembolization in hepatocellular carcinoma from CT imaging. *European radiology*, *30*(1), 413-424. https://doi.org/10.1007/s00330-019-06318-1

[9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[10] Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. et al. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

[13] Kennedy, T. (2018). *OscarNet: Using transfer learning to classify disposable waste. CS230 Report: Deep Learning*. Stanford University, CA, Winter.

[14] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. *Thirty-first AAAI conference on artificial intelligence.* https://doi.org/10.1609/aaai.v31i1.11231

[15] Aral, R. A., Keskin, Ş. R., Kaya, M., & Hacıömeroğlu, M. (2018, December). Classification of trashnet dataset based on deep learning models. *IEEE International Conference on Big Data (Big Data)*, 2058-2062. https://doi.org/10.1109/BigData.2018.8622212

[16] Adedeji, O. & Wang, Z. (2019). Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, *35*, 607-612. https://doi.org/10.1016/j.promfg.2019.05.086

[17] Vo, A. H., Vo, M. T., & Le, T. (2019). A novel framework for trash classification using deep transfer learning. *IEEE Access*, *7*, 178631-178639. https://doi.org/10.1109/ACCESS.2019.2959033

[18] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision (ECCV)*, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[19] Loshchilov, I. & Hutter, F. (2017). Decoupled weight decay regularization.

[20] Cen, F. & Wang, G. (2019). Dictionary representation of deep features for occlusion-robust face recognition. *IEEE Access*, *7*, 26595-26605. https://doi.org/10.1109/ACCESS.2019.2901376
[21] Shin, D. K., Ahmed, M. U., & Rhee, P. K. (2018). Incremental deep learning for robust object detection in unknown cluttered environments. *IEEE Access*, *6*, 61748-61760. https://doi.org/10.1109/ACCESS.2018.2875720
[22] Shi, C., Xia, R., & Wang, L. (2020). A novel multi-branch channel expansion network for garbage image classification. *IEEE Access*, *8*, 154436-154452. https://doi.org/10.1109/ACCESS.2020.3016116
[23] Satvilkar, M. (2018). *Image based trash classification using machine learning algorithms for recyclability status*. Doctoral dissertation, Dublin, National College of Ireland.
[24] Costa, B. S., Bernardes, A. C., Pereira, J. V., Zampa, V. H., Pereira, V. A., Matos, G. F., Silva, A. F. et al. (2018). Artificial intelligence in automated sorting in trash recycling. *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional*, 198-205. https://doi.org/10.5753/eniac.2018.4416
[25] Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., & Xiong, X. (2018). Multilayer hybrid deep-learning method for waste classification and recycling. *Computational Intelligence and Neuroscience*, *2018.* https://doi.org/10.1155/2018/5060857

Contact information:

Xiaoxuan MA, PhD, Associate Professor
Beijing University of Civil Engineering and Architecture,
School of Electrical and Information Engineering,
No. 15, Yongyuan Road, Huangcun, Daxing District, Beijing, China
E-mail: maxiaoxuan@bucea.edu.cn

Zhiwen LI, MS
Beijing University of Civil Engineering and Architecture,
School of Electrical and Information Engineering,
No. 15, Yongyuan Road, Huangcun, Daxing District, Beijing, China
E-mail: 2108550020021@stu.bucea.edu.cn

Lei ZHANG, PhD, Associate Professor
(Corresponding author)
Beijing University of Civil Engineering and Architecture,
School of Electrical and Information Engineering,
No. 15, Yongyuan Road, Huangcun, Daxing District, Beijing, China
E-mail: lei.zhang@bucea.edu.cn