Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4793555, 12 pages
https://doi.org/10.1155/2022/4793555

Research Article

Image Recognition for Garbage Classification Based on Transfer Learning and Model Fusion

Wei Liu, Hengjie Ouyang, Qu Liu, Sihan Cai, Chun Wang, Junjie Xie, and Wei Hu

School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, Hunan, China

Correspondence should be addressed to Wei Liu; weiliu@hnucm.edu.cn

Received 22 June 2022; Accepted 13 July 2022; Published 2 August 2022

Academic Editor: Zaoli Yang

Garbage is an underutilized resource, and garbage classification is one of the effective ways to make full use of these resources. In order to realize the automation of garbage classification, some deep learning models are used for garbage images recognition. A novel garbage image recognition model Garbage Classification Net (GCNet) based on transfer learning and model fusion is proposed in this paper. After extracting garbage image features, EfficientNetv2, Vision Transformer, and DenseNet, respectively, are combined to construct the neural network model of GCNet. Data augmentation is used to expand the dataset and 41,650 garbage images are contained in the new dataset. Compared with other models through experiments, the results show that the proposed model has good convergence, high recall rate and accuracy, and short recognition time.

1. Introduction

With the continual and rapid development of the economy, environmental pollution is becoming more serious, endangering the lives of billions of people, reducing life expectancy, and harming the growth and development of children [1].

Garbage is the primary source of pollution. “Garbage in the city” and “garbage in the countryside” are becoming more and more of an issue for towns and villages. Garbage classification is a symbol of social and environmental development.

To promote the work of garbage classification and delivery, many cities, represented by Shanghai, have issued mandatory garbage classification laws. However, there are still significant issues with garbage classification. For instance, residents’ awareness of garbage classification is still relatively low, and many people do not understand garbage classification and have unclear standards for garbage classification.

Garbage classification methods that are automated can assist in resolving these challenges. China launched the first garbage classification system in Shanghai [2]. However, there has a problem that the automated garbage classification system cannot classify the garbage images accurately [3]. Therefore, this paper proposes an image recognition algorithm for garbage classification based on transfer learning and model fusion.

As the basis of the above algorithms, in recent years, deep learning has developed rapidly due to the improvement of computational power and theoretical systems. Compared with traditional image feature extraction methods, deep learning does not require the preextraction of features [4]. In the era of big data, models are allowed to learn from large-scale data. Therefore, deep learning has greater learning ability, better adaptability, and a higher upper limit.

Deep learning is highly dependent on data, but different countries have different standards of domestic garbage classification, and there is no suitable dataset in China or even internationally in terms of dataset selection. Therefore, this paper uses transfer learning to make up for the lack of datasets. Transfer learning is a machine learning method that takes knowledge from one domain and transfers it to another domain, enabling better learning results in the target domain.

Convolutional Neural Network (CNN), the most fundamental network for deep learning, was first proposed by

LeCun and others. After continuous development, Krizhevsky [5] and others used it for the first time for classification tasks and achieved excellent results. CNN is also the most widely used deep learning algorithm in the field of computer vision (CV).

However, the types of household garbage are complex, and the distinctions between them are not clear. Ordinary CNN has difficulty learning the differences between categories and cannot complete the classification effectively. As a result, model fusion is used to improve the model's underlying feature extraction ability, which improves the model's learning ability.

2. Related Work

Research on garbage classification, Yang and his colleagues from Stanford University created the public TrashNet Dataset. There are 2,527 images separated into six categories: 403 cardboard, 501 glass, 410 metal, 594 paper, 482 plastic, and 137 other waste materials. Yang and Thung used a method called support vector machine (SVM) to perform early trials on this dataset with a 63% accuracy [6].

Satvilkar used an algorithm called random forest (RF) to classify these images and achieved an accuracy of 62.61%. The RF algorithm is a classifier that trains and predicts samples through multiple decision trees, each of which plays a role in the final decision of the predicting outcomes [7, 8]. In the era of big data and samples, RF training can highly parallelize data, which improve training speed. Later, Satvilkar did experiments using another algorithm called XGBoost with an accuracy of 70.1% [3]. The XGBoost algorithm is an improvement on the Gradient Boosted Decision Tree (GBDT) algorithm. It is based on two integrated tree-based learning classifiers named RF and GBDT [9], which have the advantage of being less prone to overfitting.

Costa et al. and others used the K-nearest neighbor (KNN) algorithm to classify images and achieved an accuracy of 88.0% [10]. The KNN classification algorithm is easy to implement, has remarkable classification performance [11], and is frequently used in image classification. In the KNN algorithm, the determination of an image category is based on the class of the nearest one or more images [12].

Traditional machine learning methods described above have been around for a long time and have achieved good results in the field of image processing. However, these methods usually consist of several independent processes, so they require a great amount of storage space for intermediate results, causing cumbersome and unintelligent implementation procedures.

Now, many scholars have been using deep learning methods to solve problems in the field of image processing. Rabano et al. applied the MobileNet model to the TrashNet Dataset with an accuracy of 87.2%. This model application was successfully installed on Samsung Galaxy S6 Edge+ mobile phone [13]. Ruiz et al. used a combined Inception-ResNet model on the TrashNet and achieved an average accuracy of 88.6% [14]. Adedeji et al. used the 50-layer residual net pretrain (ResNet-50) CNN model as the extractor and replaced the full-connected layer with an SVM in the later classification stage. An accuracy of 87% was achieved on this dataset [15]. Aral et al. did experiments with two fine-tuned model (95% for the DenseNet121 and 94% for the InceptionResNetV2) [16]. Ozkaya et al. compared a variety of combinations of network and classifier for extracting classification features, and then found the best combination of the GoogleNet and the SVM classifier with an accuracy of 97.86%, which is the best result on the TrashNet Dataset by far [17].

In addition to some nonpublic datasets, Mittal et al. created the GINI dataset of 2561 images, used the GarbNet model, and obtained an average accuracy of 87.69% [18].

Yang et al. proposed a GarbageNet model. It uses the garbage classification dataset from the Huawei Cloud Garbage Classification Challenge, employs transfer learning, and learns noise-resistant features through a feature synthesis module. In addition, they designed a memory pool and a metric-based classifier to improve the model without retraining it. The best performance was achieved with an average accuracy of 96.96% [19].

Guo et al. investigated an algorithmic model for garbage classification based on EfficientNet. The dataset from Huawei Artificial Intelligence Competition was used. To prevent some irrelevant information in the images from affecting the training of the model, an attention mechanism was added after the EfficientNet output to emphasize or select the important information of the target-processing object and suppress some irrelevant details, enabling the model to focus on key features and better recognize the images. The final average accuracy rate reached 93.47% [20].

Fu et al. proposed a new migration learning-based GNet model for rubbish classification and an improved MobileNetV3 model, with an average accuracy of 92.62% [21].

In conclusion, the powerful feature characterization capability of convolutional neural networks not only completely liberates the process of manual extraction of image features in traditional image classification but also makes good use of the huge amount of current image data, which have extensive research significance.

However, there are still great challenges in applying convolutional neural networks to garbage classification:

(1) The accuracy of convolutional neural networks relies heavily on the quality of the training dataset. But there are a few publicly domestic and international datasets. This increases the resistance to the application of convolutional neural networks.

(2) The image background is single, and the algorithm's generalization cannot be proved.

(3) The types of household garbage are complex, and the differences between them are not obvious. Therefore, it is difficult for ordinary convolutional neural networks to learn the differences between different categories and complete classification effectively.

In recent years, self-attention-based architectures, particularly Transformers, have become the preferred model for natural language processing (NLP) [22]. Inspired by this, Dosovitskiy et al. attempted to apply Transformers to the field of the image. After they performed pretraining on

Google's JFT dataset, VIT approaches or beats state of the art on multiple image recognition benchmarks. The best model on ImageNet achieved an accuracy of 88.55% [23].

VIT does not need to rely on CNN architecture to achieve good results on image classification tasks. This paper compares this model with others.

In summary, the main contributions of this paper are as follows:

(1) In addition to collecting lots of datasets of existing garbage images on the Internet, this paper photographs and labels more than 2,000 garbage images for processing. Finally, this paper uses more than 40,000 garbage images and uses data augmentation techniques to enrich the dataset.

(2) Using pretrained models on ImageNet through transfer learning, this paper greatly improves the identification results of classification tasks with insufficient samples.

(3) Considering the problem that ordinary convolutional neural networks do not have strong generalization ability, this paper designs a network model based on model fusion, combining various pretrained models, to effectively learn the differences between garbage categories. Finally, this paper produces predicted results and completes garbage classification.

3. Proposed Methodology

3.1. Parameter Debugging Based on Adam's Adaptive Method. Tuning parameters is a major difficulty in deep learning. By iteratively updating each sample, Stochastic Gradient Descent (SGD) improves the overall optimization efficiency with the loss of a small fraction of precision while increasing the number of iterations by a certain amount. The number of extra iterations is significantly less than the number of samples. In the training process, a fixed learning rate is usually used for training, using gradient descent for the parameters θ. g_t is the gradient and η is the learning rate.

$$\begin{aligned} g_t &= \nabla_{\theta_{t-1}} f(\theta_{t-1}), \\ \theta_t &= \theta_{t-1} - \eta * g_t. \end{aligned} \tag{1}$$

However, SGD has several obvious drawbacks:

(1) SGD is parameter-sensitive and must pay close attention to parameter initialization.

(2) It is simple to fall into local minima.

(3) As more data become available, the training process will take longer.

(4) All of the data from the training set are used for each iteration step.

In SGD, each parameter is updated with the same learning rate. However, in practical application, each parameter has different importance, so different learning rates should be dynamically adapted for different parameters, to achieve faster convergence objective function.

To make a dynamic update of learning rate by Adagrad adaptive method, square the gradient of each iteration of each parameter, then take the square root after accumulation, and divide the basic learning rate. The learning rate of each parameter is thus tied to its gradient, resulting in a separate learning rate for each parameter, which is referred to as the adaptive learning rate.

Based on gradient descent, a gradient accumulation variable S_t is added:

$$S_t = S_{t-1} + g_t \odot g_t. \tag{2}$$

$\odot$ denotes the dot product between elements, and the learning rate is adjusted by gradient:

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{S_t + \varepsilon}} \odot g_t. \tag{3}$$

ε is the small parameter introduced to maintain numerical stability. It can be seen that the learning rate has changed from a fixed adaptive learning rate to an adaptive learning rate controlled by a gradient accumulation variable.

It is easy to see that as the algorithm continues to iterate, S_t will get bigger and the overall learning rate will get smaller. So in general, Adagrad adaptive method starts as an incentive convergence, and then it becomes a penalty convergence, slower and slower.

The learning rate of each element of the Adagrad adaptive method has been decreasing (or unchanged) in the iteration process, and it is difficult to find a useful solution in the late iteration due to the low learning rate. Given the above problems, RMSProp uses an exponentially weighted average for the gradient and cumulative variables:

$$S_t = \gamma S_{t-1} + (1 - \gamma) g_t \odot g_t. \tag{4}$$

RMSProp and Adagrad algorithms use the same adaptive learning rate method.

Adam is essentially RMSProp with momentum terms, combining the strengths of the Adagrad and RMSProp algorithms. It dynamically adjusts the learning rate of each parameter using first-order moment estimates and second-order moment estimates of the gradient.

Adam's advantage lies mainly in the fact that, after bias correction, there is a defined range of learning rates for each iteration, making the parameters relatively smooth. Formulas are as follows:

$$\begin{aligned} m_t &= \mu * m_{t-1} + (1 - \mu) * g_t, \\ n_t &= v * n_{t-1} + (1 - v) * g_t^2, \\ \widehat{m_t} &= \frac{m_t}{1 - \mu^t}, \\ \widehat{n_t} &= \frac{n_t}{1 - v^t}, \\ \theta_t &= \theta_{t-1} - \frac{\widehat{m_t}}{\sqrt{\widehat{m_t} + \varepsilon}} * \eta. \end{aligned} \tag{5}$$

m_t and n_t is the first and the second moments estimator of the gradient, respectively; $\widehat{m_t}$ and $\widehat{n_t}$ are corrections to m_t and n_t.

3.2. Neural Network Activation Based on ReLU. When Sigmod is used as the activation function in deep neural network training, gradient dispersion phenomenon occurs, network parameters cannot be updated for a long time, and developing deeper network models becomes hard, etc. Therefore, ReLU is used as the activation function of the neural network.

The definition of ReLU is as follows:

$$\mathrm{ReLU}(x) := \max(0, x). \tag{6}$$

Figure 1 shows the ReLU function diagram:

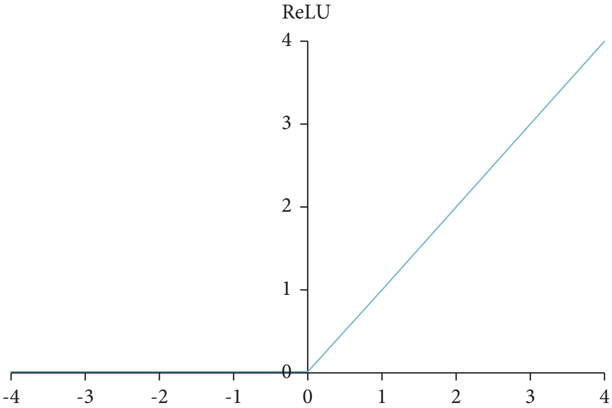


Figure 1: ReLU function diagram.

The derivative of the negative half ReLU function is 0. Once the neuron activation value enters the negative half, the gradient is 0, and the positive value remains unchanged. This is known as unilateral inhibition, and it is more similar to the biological activation model.

In addition, the derivative of the ReLU function is much faster to calculate. The program implementation is an if-else statement, whereas the sigmoid function has to perform a floating-point four operations. So the ReLU function is considerably less computational than the Sigmoid function.

Moreover, when the input signal is strong, the difference between signals can still be preserved, so that the garbage image data can be processed centrally to obtain the image dataset.

3.3. Evaluating Model Performance Based on Cross-Entropy Functions. Garbage recognition is a multiclassification problem. This paper chooses the multiclassification cross-entropy function, which is most commonly used in classification problems, as the loss function, and the training aim is to minimize this loss function. The function is defined as follows: N denotes the number of categories:

$$\text{loss} = -\sum_{i=1}^{N} y_i \log \widehat{y}_i. \tag{7}$$

3.4. Transfer Learning. In the field of deep learning computer vision, without a sufficiently wide range of training samples, the generalization ability of models will be poor. Transfer learning uses pretrained models, with minor changes to the architecture.

If a deep neural network is trained with a vast quantity of data and gains knowledge in the form of “weights” in the neural network, these weights can be extracted and transmitted to other deep neural networks so that other deep neural networks are not trained from scratch [24].

While the garbage types are numerous, and there are a few publicly domestic and international dataset. The employment of appropriate transfer learning can yield positive results in this setting.

This paper conducts transfer learning experiments and training from scratch experiments for five models, ResNet, DenseNet, EfficientNetV2, Vision Transformer, and VGGNet, respectively. These models achieve state-of-the-art performance on ImageNet for object recognition and detection [25]. The purpose is to compare the accuracy and convergence rate of transfer learning and training from scratch on garbage datasets, as well as acquire experimental findings to demonstrate transfer learning’s involvement in garbage classification.

3.5. Model Fusion. Model fusion is a type of model integration that is frequently utilized in Kaggle competitions. The final performance of the model can be enhanced by fusing numerous models, and variations in characteristics across categories in the classification task can be effectively learned.

There are various methods of model fusion, the more common ones being voting, averaging, and stacking.

Voting fusion is suitable for classification tasks, voting on the predicted results of multiple learners, that is, the minority rules the majority. Theoretically, the larger the structural differences across models, the better the voting fusion results for models that are independent of each other.

The averaging method is appropriate for regression and classification problems, in which numerous models’ predictions are averaged. Averaging has the advantage of smoothing the findings and so reducing overfitting.

The Stacking method is based on the original data, training several basic models, then combining the predictions of these basic models into a new training set to train a new classification model.

In this paper, model fusion experiments are performed mainly based on averaging, using pretrained models of DenseNet, EfficientNetV2, and Vision Transformer on ImageNet.

3.6. Garbage Classification Net. The three base models for model fusion are DenseNet, Vision Transformer, and EfficientNetV2. First, each of the three models is run through a global average pooling layer, which is then regularized to

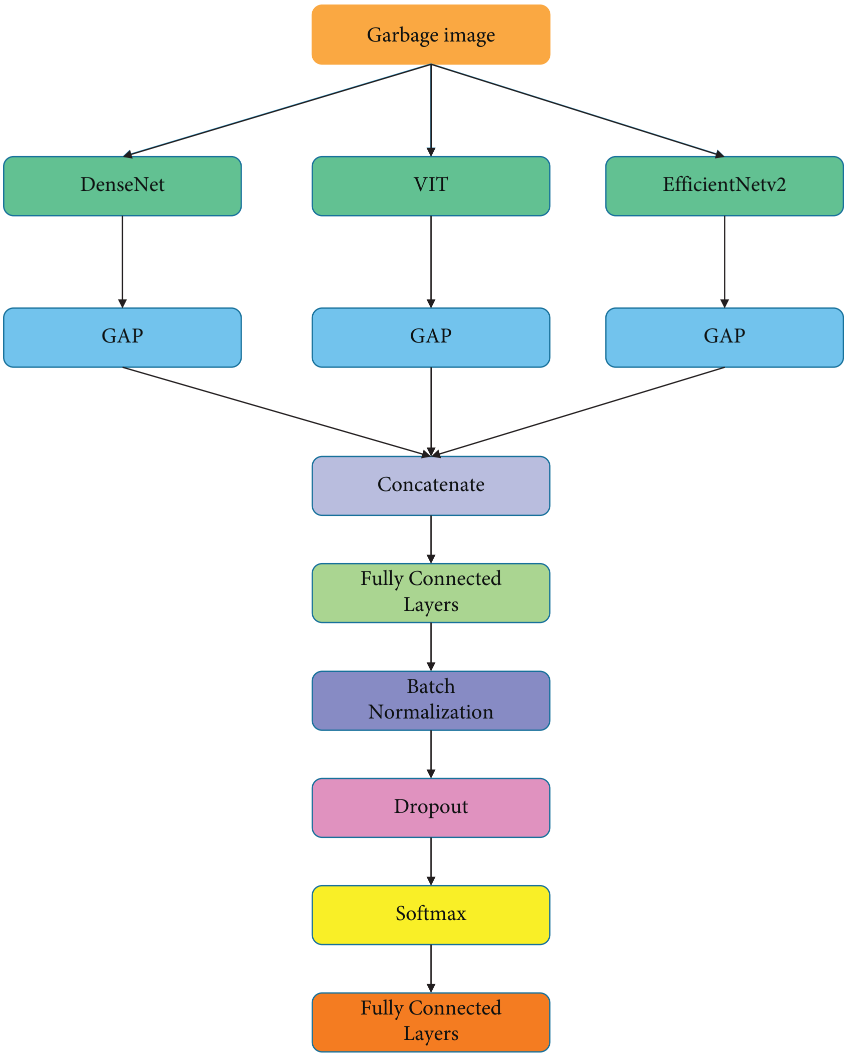


Figure 2: The model structure of GCNet.

prevent overfitting and the feature vectors are obtained. Second, feature fusion is performed using the concatenate layer. Third, the full-connected layer is used to ensure that there are enough features. Fourth, Dropout is added to prevent overfitting. Finally, Softmax is used for classification, and this new network structure is called Garbage Classification Net (GCNet) in this paper. The model structure of GCNet is shown in Figure 2.

4. Experimental Results

4.1. Experimental Configuration. In this paper's experiments, the operating system is Windows 10, 11th Gen Intel(R) Core(TM) i7-11700K @ 3.60 GHz 3.60 GHz with the memory of 32G, and the graphics card model of NVIDIA GeForce RTX 3090 with the memory of 24G.

4.2. Introduction to Datasets and Data Augmentation Techniques. Nowadays, there are few standard publicly available garbage datasets for researchers to use. Therefore, when training the model, this paper divides the garbage images into four categories, namely recyclable garbage, hazardous garbage, kitchen waste, and other garbage. Each category contains multiple objects, including 29 types of recyclable garbage: bags, newspapers, glass, glass bottles and jars, plugs and wires, rechargeable batteries, flyers, spice bottles, pots and pans, metal food cans, wine bottles, old books, old clothes, courier paper bags, plush toys, Tetra Pak packaging such as milk box, Styrofoam, leather shoes,

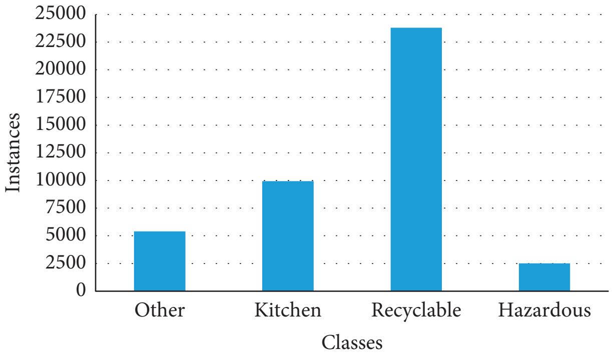
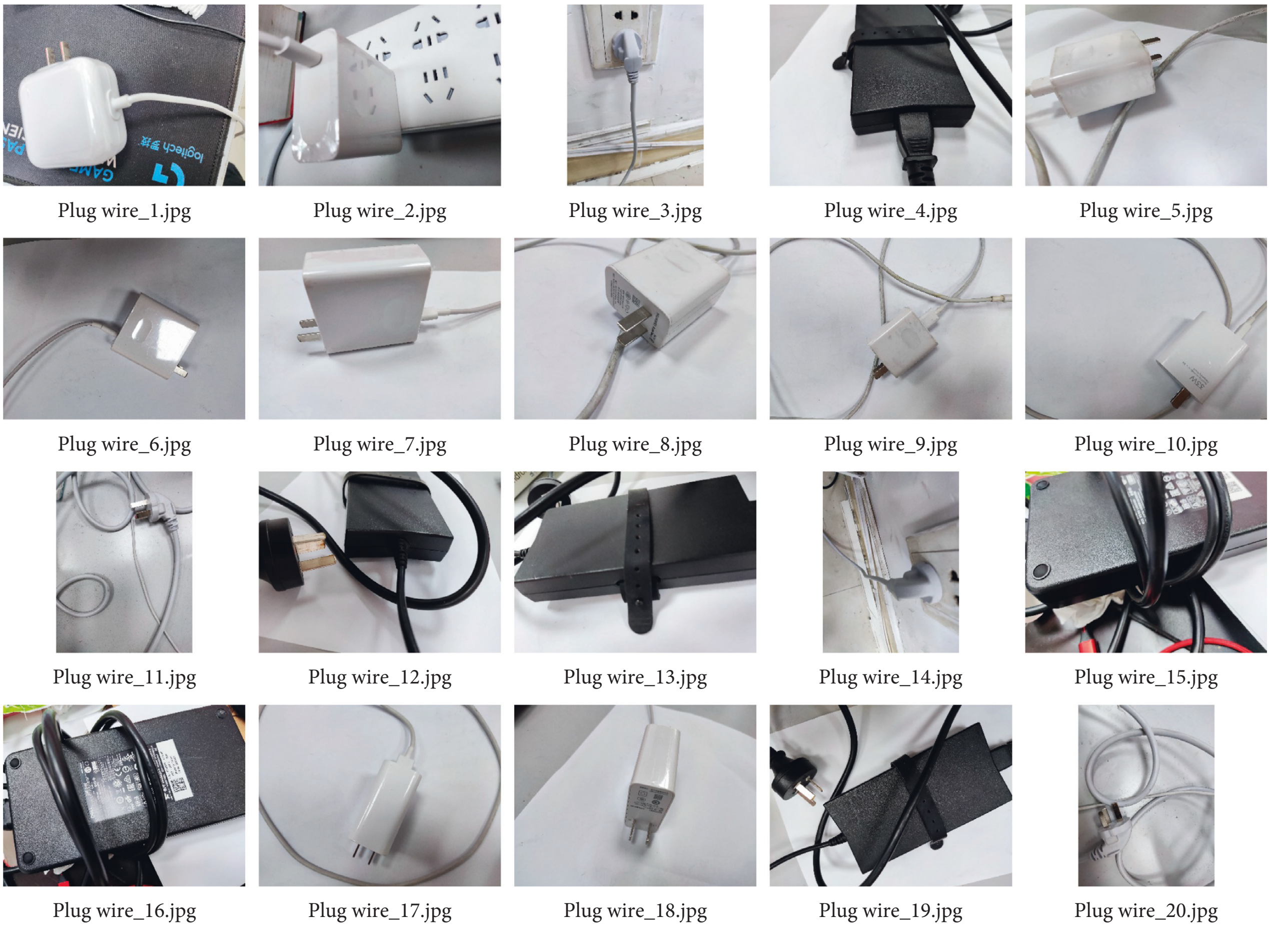


Figure 3: Dataset visualization.



Plug wire_1.jpg Plug wire_2.jpg Plug wire_3.jpg Plug wire_4.jpg Plug wire_5.jpg

Plug wire_6.jpg Plug wire_7.jpg Plug wire_8.jpg Plug wire_9.jpg Plug wire_10.jpg

Plug wire_11.jpg Plug wire_12.jpg Plug wire_13.jpg Plug wire_14.jpg Plug wire_15.jpg

Plug wire_16.jpg Plug wire_17.jpg Plug wire_18.jpg Plug wire_19.jpg Plug wire_20.jpg

FIGURE 4: Hand-taken garbage images.

cooking oil drums, plastic toys, plastic bowls and pots, plastic hangers, shampoo bottles, cans, drink bottles, magazines, chopping boards, pillows, cardboard boxes; eight types of hazardous garbage: herbicide containers, batteries, used lamp bulbs, discarded mercury thermometers, dry batteries, expired medicines, ointments, pesticide containers; 12 types of kitchen waste: vegetable stalks, leaves, tea leaves, big bones, eggshell, fruit peel and rind, fallen leaves, leftovers, fruit pulp, western pastry, fish bones; 11 types of other garbage: shells, cosmetic bottles, diapers, broken flower pots and dishes, toilet paper, soiled plastic, toothpicks, cigarette butts, disposable tableware, paper cups, bamboo chopsticks.

The dataset used in this paper has a total of 41,650 images, with each category of garbage images divided 4 : 1 to give a training set of 33,320 images and a testing set of 8,330 images. Figure 3 shows a visual display of the number of all datasets.

The dataset is divided into three main sections:

FIGURE 5: Original image.

(1) A large number of images are obtained through online crawlers. First, the principle of crawling technology is accessing web resources recursively through keywords. Second, the quality of the data collected is too poor because of inaccurate keyword matching and other reasons. Over 30,000 images are filtered out by manually eliminating images with blurred images, serious watermarks, and the presence of multiple objects.

(2) Domestic garbage dataset is opened by Huawei.

(3) 2,136 images of everyday household garbage are taken by hand. The images are taken from above, to the left, and in front of the object in a well-lit scene to extract features better during training. Figure 4 shows a collection of garbage images.

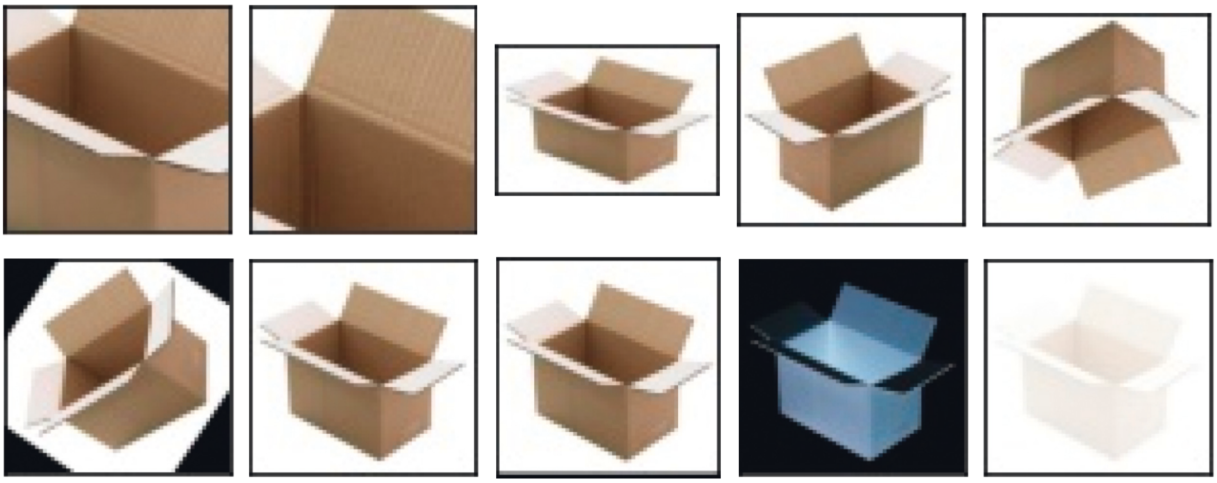


Figure 6: Images after data augmentation.

Due to the complexity of the garbage categories, there is the problem of insignificant differences between different categories. In addition, achieving learning-by-learning data representations is the core of deep learning, and if this paper wants to enhance the robustness of the learning model, this paper must have a large amount of data for training. Therefore, this paper expands the dataset by data augmentation.

As shown in Figure 5, the original images from the garbage classification training set were transformed by certain techniques, such as spatial and chromatic transformations, to obtain 10 enhanced images as shown in Figure 6. The first layer from left to right in order is the center crop, random crop, resize horizontal flip, and vertical flip; the second layer in order is random flip, followed by grayscale, turning the images into squares, color inversion, and γ transform.

4.3. Transfer Learning Experiments. In this experiment, the network structures ResNet, DenseNet, EfficientNetV2, Vision Transformer, and VGGNet are trained in two ways, training from scratch and transfer learning, respectively, and garbage datasets are divided into four types. The aim is to analyze and compare the accuracy and convergence rate of transfer learning and training from scratch on the four types of garbage datasets to derive experimental results.

The PyTorch framework randomly initializes the network parameters by default, and training from scratch is simply a matter of training all the trainable layer parameters on the dataset once the network structure has been designed.

While transfer learning training needs to be divided into two steps: feature extraction and fine-tuning. Feature extraction is to load pretrained weights first and transfer the parameters pretrained in the source domain to the network in this paper, which not only speeds up the model convergence but also enables to gain the generalization ability for better training of the model.

In contrast, fine-tuning is the process of unfreezing some or all of the trainable layers of the original network, using a lower learning rate trained through the garbage dataset of this experiment to fine-tune its original parameters to make it more suitable for the garbage classification task of this experiment.

Table 1: Transfer learning experiment parameter settings.

Training methods	Training parameters	
	Training from scratch	Fine-tuning training
Learning rate	0.001	0.0001
Optimizers	Adam	Adam
Number of iterations	30	30
Lot size	32	32
Datasets	Raw dataset	Raw dataset
Number of training levels	All layers	All layers
Loss function	Cross-entropy function	Cross-entropy function

Table 2: Model fine-tuning details.

Network structure	Training layer
ResNet	Unfreeze all layers
DenseNet	Unfreeze all layers
EfficientNetV2	Unfreeze all layers
Vision Transformer	Unfreeze all layers
VGGNet	Unfreeze all layers

4.3.1. Experimental Parameter Settings. The experiment is divided into two phases: training from scratch and transfer learning. Table 1 shows the basic training parameter settings.

4.3.2. Feature Extraction. In this experiment, the pretrained weights of the five networks ResNet, DenseNet, EfficientNetV2, Vision Transformer, and VGGNet on ImageNet are downloaded separately, and the network parameters of the pretrained weights are transferred to the experimental network.

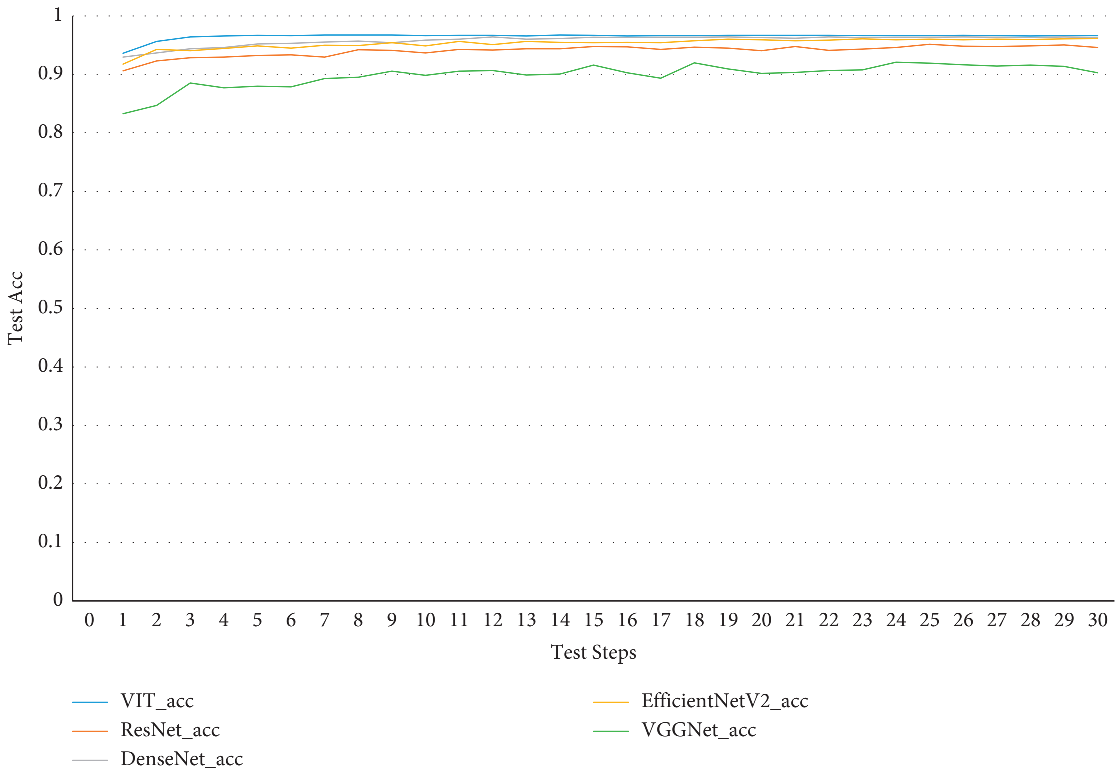


FIGURE 7: Transfer learning test accuracy.

4.3.3. Fine-Tuning. Directly using the pretrained model for classification obviously cannot solve the garbage classification image recognition problem of this experiment. To make the pretrained network weights better adapted to the garbage image data of this experiment, the pretrained model was fine-tuned. First, this paper adjusts all layers of the model, and in the original model as well as the added classification layers. Second, this paper continues training with the garbage image data to better apply it to the garbage image classification problem. Ultimately, this paper improves the recognition accuracy of the model. Table 2 shows the specific details of fine-tuning training for each network model.

4.3.4. Experimental Results. Figures 7 and 8 show the transfer learning's accuracy and loss change curves and Figures 9 and 10 show the training from scratch's accuracy and loss change curves.

As can be seen in Figures 7 and 10, the convergence rate is slower with training from scratch, and it is difficult to train a better model on the garbage dataset in this paper because the garbage image samples are not widely available and the accuracy rate on the test set is not high. In contrast, the model under transfer learning has a faster convergence rate, and on the test set, each model has high accuracy.

The highest accuracies of training from scratch and transfer learning training in the test set are shown in Table 3. The model cannot provide enough underlying features because the garbage image samples are not extensive enough. In contrast, the model after transfer learning, which has powerful underlying features with weights on ImageNet, greatly improves the recognition results for classification tasks with insufficient samples.

4.4. Model Fusion Experiments. In this experiment, multiple models are fused to extract features together. The performance of GCNet is compared to that of individual models to demonstrate its superiority.

After transfer learning experiments, the pretrained models of DenseNet, Vision Transformer, and EfficientNetV2 on ImageNet are obtained and have the highest accuracy for garbage image recognition in this paper. Therefore, the three base models for GCNet are DenseNet, Vision Transformer, and EfficientNetV2.

4.4.1. Experimental Parameter Settings. Table 4 shows the model fusion experimental parameter settings.

4.4.2. Experimental Results. As can be seen from Table 5, the accuracy of the fused model GCNet on the test set (97.54%) is better than the single models DenseNet (96.40%), Vision Transformer (96.75%), and EfficientNetV2 (96.12%). It indicates that model fusion can further improve the generalization ability of the models and effectively learn the differences between garbage categories.

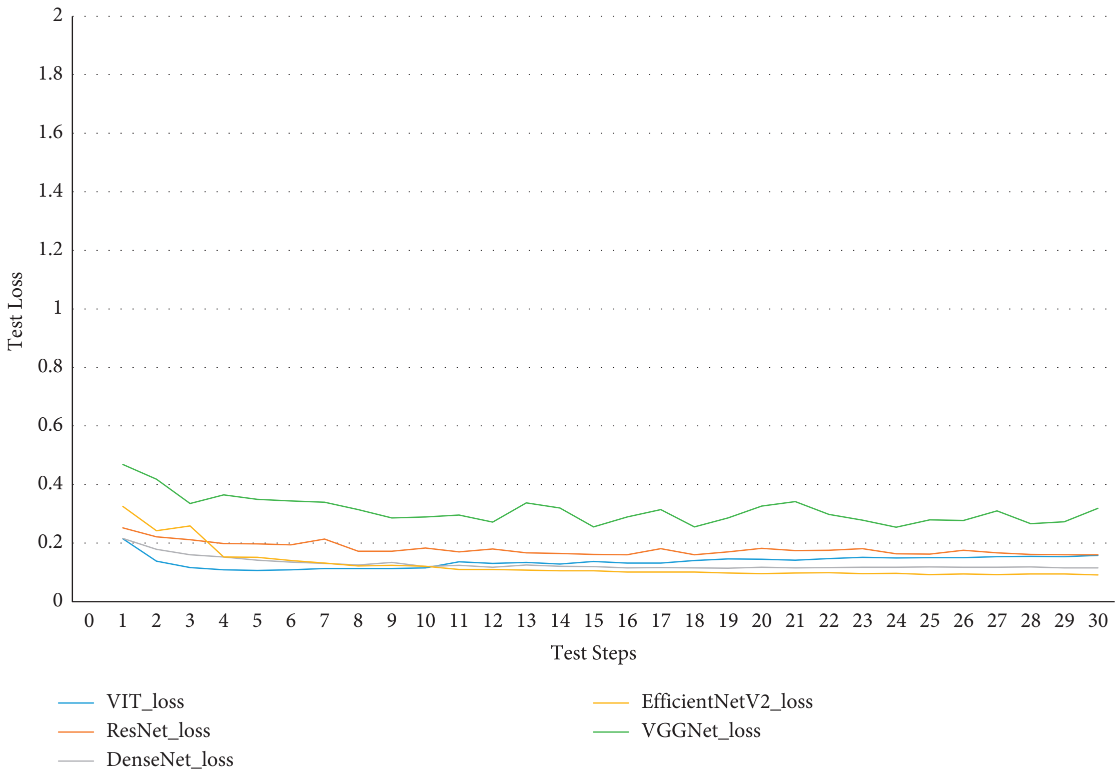


Figure 8: Transfer learning test loss.

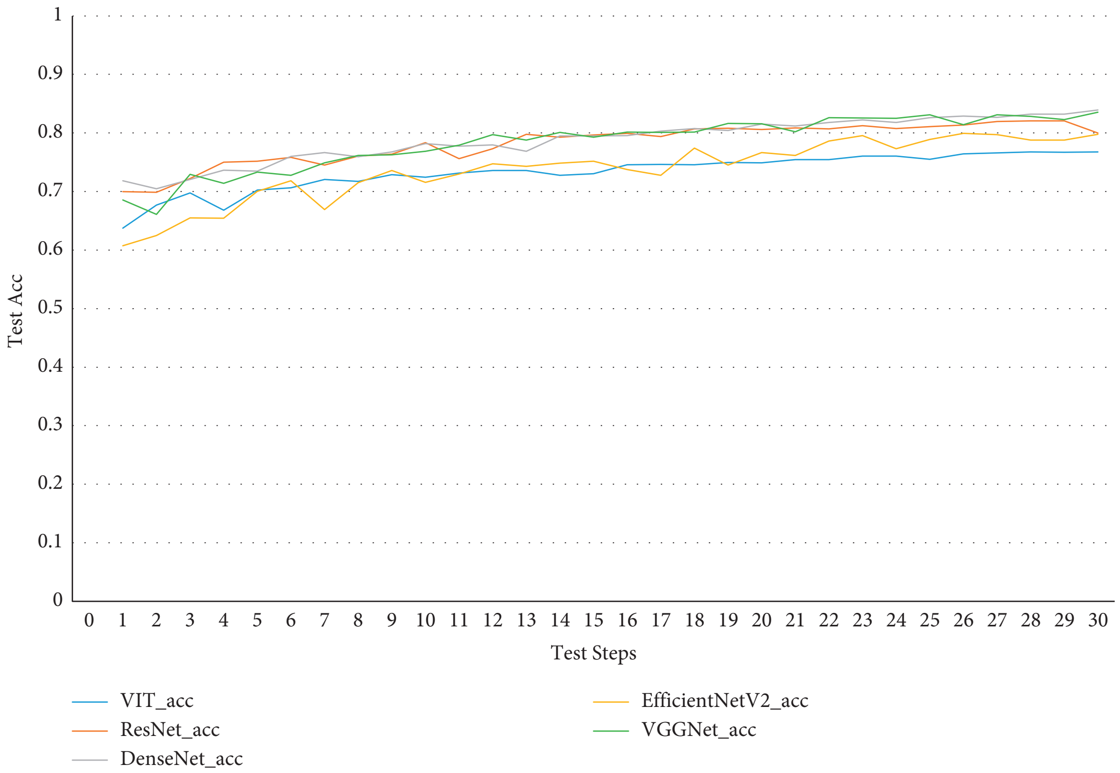


Figure 9: Test accuracy of training from scratch.

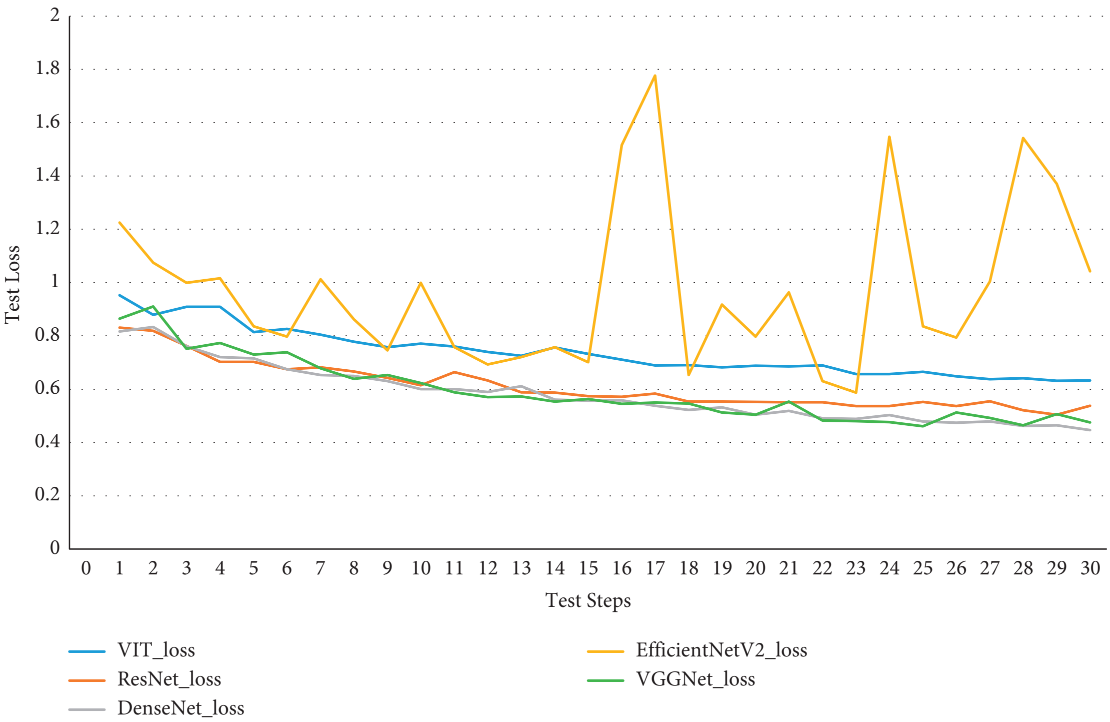


Figure 10: Test loss of training from scratch.

Table 3: Comparison of accuracy rates of transfer learning experiments.

Network structure	Training from scratch (accuracy) (%)	Transfer learning (accuracy) (%)
ResNet	82.07	93.38
DenseNet	83.91	96.40
EfficientNetV2	78.13	96.12
Vision Transformer	76.74	96.75
VGGNet	83.52	93.77

Table 4: Model fusion experiment parameter settings.

Parameters	Learning rate	Lot size	Optimizers	Number of iterations
Value	0.001	30	Adam	30

Table 5: Comparison of accuracy rates of model fusion experiments.

Network structure	Accuracy (%)
GCNet	**97.54**
DenseNet	96.40
EfficientNetV2	96.12
Vision Transformer	96.75

5. Conclusions

A transfer learning and model fusion-based garbage classification image recognition algorithm is proposed for the classification problem.

After transfer learning experiments, it is found that the pretrained models of DenseNet, Vision Transformer, and EfficientNetV2 on ImageNet work best for the garbage image dataset in this paper. At the same time, it also confirms that Transformer is not only suitable for natural language processing related tasks but also for computer vision and the garbage image recognition in this paper.

Therefore, this paper uses DenseNet, Vision Transformer, and EfficientNetV2 as the basic models for model fusion experiments and designs a neural network model named Garbage Classification Net suitable for garbage image recognition. This algorithm achieves the best performance of 97.54% when the inference speed is acceptable, which exceeds most of the mainstream methods.

This paper also improves the generalization ability of the model by filtering and enhancing the dataset obtained from the collection and hand-photographed datasets. However, the following shortcomings still exist in this paper's research:

(1) This paper creates a garbage classification dataset, which provides a database for the training of the

classification model the classification effect is good but the multitarget garbage detection is not achieved and the target detection task needs to be improved.

(2) The model classification still has a certain false detection rate, which needs to be optimized.

(3) Model fusion can also consider more ways of model fusion, as well as choosing other models for fusion.

Based on the above issues, this paper has the following outlook:

(1) Launching research on multiobjective spam detection and expanding the dataset.

(2) The model is further optimized by adding a self-attention mechanism and modifying the model structure to achieve more accurate garbage classification through experiments, which will help to promote the further development of garbage classification.

(3) This paper adopts the fusion of three models, DenseNet, Vision Transformer, EfficientNetV2, and other fusion methods that can be considered in the subsequent research to further improve the classification accuracy.

Data Availability

The dataset used to support the findings of the study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank for the funding support from the Scientific Research Project of the Education Department of Hunan Province (no. 20C1435) and the Natural Science Funding Project of Hunan Province (no. 2022JJ30438).

References

[1] R. Ramasamy, "Assessment of comprehensive environmental pollution index of kurichi industrial cluster, coimbatore district, Tamil nadu, India—a case study," *Journal of Ecological Engineering*, vol. 19, no. 1, pp. 191–199, 2018.

[2] J. Zheng, M. Xu, M. Cai, Z. Wang, and M. Yang, "Modeling group behavior to study innovation diffusion based on cognition and network: an analysis for garbage classification system in Shanghai, China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 18, p. 3349, 2019.

[3] C. Shi, R. Xia, and L. Wang, "A novel multi-branch channel expansion network for garbage image classification," *IEEE Access*, vol. 8, Article ID 154436, 2020.

[4] S. Meng, N. Zhang, and Y. Ren, "X-DenseNet: deep learning for garbage classification based on visual images," *Journal of Physics: Conference Series*, vol. 1575, no. 1, Article ID 012139, 2020.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[6] M. Yang and G. Thung, "Classification of trash for recyclability status," *CS229 Project Report*, vol. 2016, pp. 1–6, 2016.

[7] M. Satvilkar, "Image based trash classification using machine learning algorithms for recyclability status," *Image*, vol. 13, p. 8, 2018.

[8] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, Article ID 180235, 2019.

[9] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters," *IEEE Access*, vol. 7, Article ID 118310, 2019.

[10] B. S. Costa, A. C. S. Bernardes, J. V. A. Pereira et al., "Artificial intelligence in automated sorting in trash recycling," *in Anais do XV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, 2018, São Paulo, Brazil.

[11] S. Zhang, X. Li, M. Zong, and X. Zhu, "Efficient KNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, 2018.

[12] W. Xing and Y. Bei, "Medical health big data classification based on KNN classification algorithm," *IEEE Access*, vol. 8, Article ID 28808, 2019.

[13] S. L. Rabano, M. K. Cabatuan, E. Sybingco, E. P. Dadios, and E. J. Calilung, "Common garbage classification using MobileNet," in *Proceedings of the IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, November 2018.

[14] V. Ruiz, A. Sanchez, J. F. Velez, and B. Raducanu, "Automatic image-based waste classification," in *Proceedings of the International Work-Conference on the Interplay Between Natural and Artificial Computation*, Almería, Spain, June 2019.

[15] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," *Procedia Manufacturing*, vol. 35, pp. 607–612, 2019.

[16] R. A. Aral, S. R. Keskin, M. Kaya, and M. Haciomeroglu, "Classification of TrashNet dataset based on deep learning models," in *Proceedings of the IEEE International Conference on Big Data (Big Data)*, December 2018.

[17] U. Ozkaya and L. Seyfi, "Fine-tuning models comparisons on garbage classification for recyclability," in *Proceedings of the 2nd International Symposium on Innovative Approaches in Scientific Studies*, Samsun, Turkey, December 2019.

[18] G. Mittal, M. Garg, K. B. Yagnik, and N. C. Krishnan, "SpotGarbage: smartphone app to detect garbage using deep learning," in *Proceedings of the Acm International Joint Conference on Pervasive & Ubiquitous Computing ACM*, September 2016.

[19] J. Yang, Z. Zeng, K. Wang, and H. Zou, "GarbageNet: a unified learning framework for robust garbage classification," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 4, 2021.

[20] Q. Guo, Y. Shi, and S. Wang, "Research on deep learning image recognition technology in garbage classification," in

Proceedings of the Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang, China, January 2021.

[21] B. Fu, S. Li, J. Wei, Q. Li, and Q. Wang, "A novel intelligent garbage classification system based on deep learning and an embedded linux system," *IEEE Access*, vol. 9, Article ID 131134, 2021.

[22] A. Vaswani, N. Shazeer, N. Parmar, and J. Uszkoreit, "Attention is all you need," arXiv, 2017, https://arxiv.org/abs/1706.03762?context=cs.

[23] A. Dosovitskiy, B. Lucas, K. Alexander, W. Dirk, and Z. Xiaohua, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, https://arxiv.org/abs/2010.11929.

[24] D. Gupta, S. Jain, F. Shaikh, and G. Singh, *Transfer Learning & the Art of Using Pre-trained Models in Deep Learning*, Analytics Vidhya, 2017.

[25] J. Deng, W. Dong, R. Socher, and K. Li, "ImageNet: a large-scale hierarchical image database," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, June 2009.