Computing, Performance and Communication Systems (2025)
Clausius Scientific Press, Canada
DOI: 10.23977/cpcs.2025.090109
ISSN 2371-8889 Vol. 9 Num. 1

Application of Lightweight CNN in Garbage Image Classification

Peng Yin

School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China
3473792229@qq.com

Keywords: Lightweight Convolutional Neural Network; Garbage Image Classification; Transfer Learning; ResNet Simplification; TrashNet Dataset

Abstract: With urbanization accelerating, domestic waste output has surged, and garbage classification is crucial for alleviating environmental pressure and recycling resources. Traditional manual classification is inefficient, costly, and subjective, so automated garbage classification technology is necessary. To solve problems of existing CNN models like large parameter size, slow inference, and difficulty in edge - device deployment, this paper proposes a lightweight CNN based on a simplified ResNet for garbage image classification. The public TrashNet dataset with 6 common domestic waste categories is used. Data augmentation and transfer learning are employed to optimize the model's adaptation to garbage image features. Experimental results show the simplified ResNet model achieves 91.2% classification accuracy on the TrashNet dataset, with precision of 90.8%, recall of 90.5%, and an F1 - score of 90.6%. Its parameter number is only 48% of the traditional ResNet18, and the per - image inference time is shortened to 12.3ms. Compared with mainstream models, it reduces computational complexity and storage requirements while ensuring performance, making it more suitable for edge - computing devices like smart trash cans and classification robots, and providing an efficient solution for practical automated garbage classification.

1. Introduction

With urbanization accelerating, domestic waste output has surged, and garbage classification is crucial for alleviating environmental pressure and recycling resources. Traditional manual classification is inefficient, costly, and subjective, so automated garbage classification technology is necessary. To solve problems of existing CNN models like large parameter size, slow inference, and difficulty in edge-device deployment, this paper proposes a lightweight CNN based on a simplified ResNet for garbage image classification[1]. The public TrashNet dataset with 6 common domestic waste categories is used. Data augmentation and transfer learning are employed to optimize the model's adaptation to garbage image features. Experimental results show the simplified ResNet model achieves 91.2% classification accuracy on the TrashNet dataset, with precision of 90.8%, recall of 90.5%, and an F1-score of 90.6%. Its parameter number is only 48% of the traditional ResNet18[1], and the per-image inference time is shortened to 12.3ms. Compared with

mainstream models, it reduces computational complexity and storage requirements while ensuring performance, making it more suitable for edge-computing devices like smart trash cans and classification robots, and providing an efficient solution for practical automated garbage classification.

The rapid urbanization and improved living standards have increased global domestic waste output, making "garbage siege" a major environmental issue. Garbage classification, a prerequisite for waste reduction, recycling, and harmless treatment, can cut disposal costs, reduce pollution, and promote resource reuse. China plans to implement a domestic waste classification system, but current manual sorting has low efficiency, unstable accuracy, and high labor costs, making it hard to meet large-scale needs. Thus, developing efficient automated garbage classification technology is significant.

Image recognition technology offers a new solution for automated garbage classification. Convolutional neural network-based image classification has advanced in natural image classification and object detection and is being applied to garbage image classification. However, traditional CNN models like VGG and ResNet[1] have large parameters and high complexity, making them difficult to deploy in edge computing. Lightweight convolutional neural networks, by simplifying structures and optimizing operations, are suitable for edge devices.

In existing garbage image classification research, some use lightweight models like MobileNet[2][3] and ShuffleNet[4] but lack consideration of garbage image feature differences, and others improve traditional CNN models[1] without targeted lightweight optimization. The TrashNet dataset, commonly used in this field, has limited samples and sample imbalance, which may cause overfitting.

To solve these problems, this paper uses the TrashNet dataset for 4-6 categories of domestic waste classification. It designs a simplified ResNet model[1] by reducing residual blocks and optimizing convolution kernels. A transfer learning strategy[6] initializes the model with pre-trained ImageNet parameters[7] and fine-tunes the feature extraction layer to adapt to garbage image features, alleviating overfitting. Data augmentation expands sample diversity and improves generalization. In experiments, the proposed model's performance, complexity, and inference speed are verified by comparing with traditional models[1], providing support for lightweight models' practical application in garbage image classification.

2. Related Work

2.1 Research Status of Garbage Image Classification

Research on garbage image classification technology began with traditional machine learning methods. Early scholars mainly achieved garbage classification by manually extracting image features combined with classifiers. For example, some studies used Histogram of Oriented Gradients (HOG) to extract texture features of garbage images and combined with Support Vector Machine (SVM) classifiers to achieve binary classification of recyclable and non-recyclable garbage, with an accuracy of about 75%. Other studies used color histograms and morphological features to classify 3 categories of garbage such as glass, plastic, and metal through random forest classifiers, achieving an accuracy of 82%. However, manual feature extraction relies on human experience, making it difficult to capture the deep semantic features of garbage images. It has poor adaptability to changes in garbage shape, differences in illumination, and background interference, resulting in limited classification performance.

With the development of deep learning technology, automatic feature extraction methods based on CNN have gradually replaced manual feature extraction and become the mainstream method for garbage image classification. Some studies used the VGG16 model for classification on the

TrashNet dataset, achieving an accuracy of 86.7%, but the model contains 138 million parameters with high computational complexity. Other scholars improved the ResNet50 model by adding an attention mechanism to enhance the targeting of feature extraction, increasing the classification accuracy to 89.5%, but the number of parameters is still as high as 25.6 million, making it difficult to meet the deployment requirements of edge devices. In recent years, the application of lightweight CNN models in garbage image classification has attracted attention. Some studies used the MobileNetV2 model for classification on the TrashNet dataset, achieving an accuracy of 88.3% with only 3.4 million parameters, but the classification accuracy of this model in complex backgrounds needs to be improved.

2.2 Research on Lightweight Convolutional Neural Networks

The core design goal of lightweight convolutional neural networks is to minimize the number of parameters and computational load while ensuring model performance. At present, the design ideas of lightweight CNN mainly include three directions: network structure simplification, convolution operation optimization, and model compression. Network structure simplification streamlines the model by reducing the number of network layers and the number of convolution kernels. For example, ResNet18 reduces the number of parameters from 25.6 million of ResNet50 to 11.7 million by reducing the number of residual blocks. SqueezeNet[5] reduces the number of parameters to 1/50 of AlexNet while maintaining classification performance through a "squeeze-and-excitation" structure.

Convolution operation optimization reduces computational complexity by improving traditional convolution methods. For example, depthwise separable convolution decomposes standard convolution into depthwise convolution and pointwise convolution. Under the condition of the same feature map output, the computational load is only 1/(input channel number × output channel number) + 1/output channel number of standard convolution. The MobileNet series models achieve lightweight based on depthwise separable convolution. Group convolution divides input feature maps and convolution kernels into multiple groups, performs convolution operations separately, and then concatenates the outputs. ShuffleNet solves the problem of information isolation between channels caused by group convolution through channel shuffling technology, further improving model performance. Model compression technologies include parameter pruning, quantization, knowledge distillation, etc., which reduce model storage requirements and computational load by removing redundant parameters and reducing parameter precision. For example, quantizing 32-bit floating-point parameters to 8-bit integers can reduce the model storage capacity by 75%.

2.3 Application of Transfer Learning in Image Classification

Transfer learning effectively solves the problems of small target task dataset size and high annotation cost by transferring knowledge learned from source tasks to target tasks. In image classification tasks, transfer learning usually uses pre-trained models as feature extractors or fine-tunes model parameters to adapt to target tasks. Since the ImageNet dataset contains 1000 categories and more than 1.2 million natural images, CNN models pre-trained on this dataset can learn general image features, which have good transferability to other image classification tasks.

In garbage image classification tasks, due to the limited number of samples in public datasets, transfer learning has been widely used. Some studies used VGG16 as a pre-trained model, froze the first 10 convolutional layers, and fine-tuned the latter convolutional layers and fully connected layers, increasing the classification accuracy on the TrashNet dataset by 12.3% compared with no transfer learning. Other studies based on the ResNet18 pre-trained model achieved a classification accuracy of 89.8% through a layer-wise unfreezing fine-tuning strategy. Transfer learning can not

only improve the classification performance of models on small datasets but also accelerate model training speed and reduce overfitting, providing important support for the application of lightweight models in garbage image classification.

3. Research Methods

3.1 Dataset Introduction and Preprocessing

This paper adopts the public TrashNet dataset as the basic data for model training and verification. The dataset includes 6 categories of common domestic waste samples: cardboard, glass, metal, paper, plastic, and miscellaneous waste. Among them, miscellaneous waste mainly includes non-recyclable domestic waste such as food residues and discarded clothing. The total number of samples in the dataset is 2527, and the distribution of the number of samples in each category is as follows: 403 cardboard samples, 501 glass samples, 410 metal samples, 594 paper samples, 482 plastic samples, and 137 miscellaneous waste samples. It can be seen that the number of miscellaneous waste samples is small, indicating a certain degree of sample imbalance.

To meet the model input requirements and improve the generalization ability of the model, the following preprocessing operations are performed on the dataset. First, image size normalization is carried out, and all images are resized to 224×224 pixels. This size is consistent with the input size of the pre-trained model, facilitating the implementation of transfer learning. Second, pixel value normalization is performed, and the RGB channel pixel values of the images are normalized from the range [0,255] to [0,1] to reduce the impact of pixel value differences on model training. To address the problems of sample imbalance and overfitting, data augmentation technology is used to expand the training set samples. Specific operations include random horizontal flipping, random vertical flipping, random rotation (-15 °to 15 °), random cropping (cropping ratio of 0.8 to 1.0 of the original image), and brightness adjustment (brightness coefficient of 0.8 to 1.2). Through data augmentation, the number of training set samples is expanded to 3 times the original, effectively alleviating the problems of sample imbalance and overfitting.

The dataset is divided into training set, validation set, and test set in a ratio of 7:2:1. The training set is used for model parameter training, the validation set is used to monitor the overfitting during model training and adjust hyperparameters, and the test set is used to evaluate the final classification performance of the model. The distribution of the divided data is as follows: 1769 samples in the training set, 505 samples in the validation set, and 253 samples in the test set. The proportion of each category in the training set, validation set, and test set remains consistent to ensure the rationality of data division.

3.2 Design of Lightweight ResNet Model

Based on the residual connection structure of ResNet, a lightweight ResNet model (referred to as Lite-ResNet) suitable for garbage image classification is designed. The traditional ResNet18 includes 4 residual block stages, each stage containing 2 residual blocks. Each residual block consists of 2 3×3 convolutional layers, a Batch Normalization (BN) layer, and a ReLU activation function. Lite-ResNet reduces model complexity by reducing the number of residual blocks and optimizing convolution kernel configuration while ensuring feature extraction ability.

The network structure of Lite-ResNet is as follows: the input layer receives 224×224×3 RGB images; the first layer is a 7×7 convolutional layer with 32 convolution kernels (64 for the traditional ResNet18), a stride of 2, and a padding of 3, outputting feature maps of size 112×112×32; then through a BN layer and a ReLU activation function, it enters a max-pooling layer with a pooling kernel size of 3×3, a stride of 2, and a padding of 1, outputting feature maps of size

56×56×32.

There are 3 residual block stages in total (4 for the traditional ResNet18), each stage containing 2 residual blocks. The residual blocks adopt a bottleneck structure for simplified design, and each residual block includes 1×1 convolutional layers, 3×3 convolutional layers, and 1×1 convolutional layers. The number of convolution kernels in the 1×1 convolutional layers of the first residual block stage is 32, 32, and 128 respectively, outputting feature maps of size 56×56×128; the number of convolution kernels in the 1×1 convolutional layers of the second residual block stage is 64, 64, and 256 respectively, with a stride of 2, outputting feature maps of size 28×28×256; the number of convolution kernels in the 1×1 convolutional layers of the third residual block stage is 128, 128, and 512 respectively, with a stride of 2, outputting feature maps of size 14×14×512. A BN layer and a ReLU activation function are set after each convolutional layer. The residual connection adopts identity mapping or 1×1 convolution to adjust the dimension, ensuring that the input and output feature map dimensions are consistent.

After the residual block stages, a global average pooling layer is used to compress the feature map size to 1×1×512, followed by a fully connected layer to map the feature vector to 6-dimensional output (corresponding to 6 categories of garbage), and finally, the probability distribution of each category is output through the Softmax activation function. The total number of parameters of Lite-ResNet is 5.62 million, which is only 48% of that of the traditional ResNet18, and the computational load is 1.8 GFLOPs, reducing by 52% compared with ResNet18, realizing the lightweight design of the model.

3.3 Transfer Learning and Model Training

A transfer learning strategy is adopted to improve the classification performance of the Lite-ResNet model. The pre-trained model parameters of ResNet18 on the ImageNet dataset are used as initial weights to initialize Lite-ResNet. Due to the difference in network structure between Lite-ResNet and ResNet18, only the parameters of the corresponding convolutional layers in ResNet18 are reused, and for the newly added or adjusted convolutional layers, the He normal distribution is used for parameter initialization.

A layered fine-tuning strategy is adopted during model training: first, the first two convolutional layers and the first half of the residual block stages are frozen, and only the fully connected layers and the second half of the residual block stages are trained with a learning rate of 0.001 for 10 epochs to make the model quickly adapt to the feature distribution of garbage images; then all frozen layers are unfrozen, the learning rate is adjusted to 0.0001, and training is continued for 40 epochs to fine-tune the parameters of all layers and further improve model performance.

The model training adopts the PyTorch deep learning framework, and the optimizer selects the Adam optimizer, which combines momentum gradient descent and adaptive learning rate strategy, enabling faster training convergence speed and avoiding local optimal solutions. The loss function adopts the cross-entropy loss function, which is suitable for multi-classification tasks and can effectively measure the difference between the model's prediction results and the true labels. To prevent overfitting, in addition to data augmentation technology, a dropout layer is added before the fully connected layer with a dropout probability of 0.5 to randomly discard some neurons and reduce the overfitting risk of the model. During training, the classification accuracy of the validation set is used to monitor the model performance. When the accuracy of the validation set does not improve for 5 consecutive epochs, an early stopping strategy is adopted to terminate the training and save the model parameters with the best performance.

4. Experimental Results and Analysis

4.1 Experimental Environment and Evaluation Metrics

The experimental hardware environment includes an Intel Core i7-10700F CPU, 16GB DDR4 memory, and an NVIDIA GeForce RTX 3060 graphics card (12GB video memory); the software environment includes a Windows 10 operating system, Python 3.8 programming language, PyTorch 1.12 deep learning framework, OpenCV 4.5 image processing library, and Scikit-learn 1.0 machine learning library.

Multi-dimensional evaluation metrics are used to evaluate the model performance, including classification accuracy (Accuracy), precision (Precision), recall (Recall), F1-score, as well as the number of model parameters (Params) and inference time per image (Inference Time), to comprehensively measure the classification performance, computational complexity, and inference speed of the model. The calculation formulas of each evaluation metric are as follows:

Accuracy = (Number of correctly classified samples / Total number of samples) × 100%

Precision = (Number of true positive samples / (Number of true positive samples + Number of false positive samples)) × 100%

Recall = (Number of true positive samples / (Number of true positive samples + Number of false negative samples)) × 100%

F1-Score = 2 × Precision × Recall / (Precision + Recall)

Among them, precision, recall, and F1-score are calculated using the macro-average method, that is, the metric values of each category are calculated first, and then the average value is taken to comprehensively reflect the classification performance of each category and avoid the impact of sample imbalance.

4.2 Comparative Experimental Results

To verify the performance advantages of the Lite-ResNet model, 5 mainstream models are selected for comparison, including the traditional deep learning model ResNet18, the lightweight model MobileNetV2, the classic machine learning model SVM (combined with HOG features), as well as the Lite-ResNet without transfer learning (Lite-ResNet w/o TL) and the Lite-ResNet without data augmentation (Lite-ResNet w/o DA). All models are trained and tested in the same experimental environment and dataset, and the comparison results are shown in Table 1.

Table 1 Performance Comparison of Different Models on the TrashNet Dataset

Model	Params (M)	Inference Time (ms)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM+HOG	0.12	8.5	72.3	71.8	70.6	71.2
MobileNetV2	3.4	10.2	88.3	87.9	87.5	87.7
ResNet18	11.7	25.6	90.5	90.2	89.8	90.0
Lite-ResNet DA	5.62	12.1	85.6	85.2	84.8	85.0
Lite-ResNet TL	5.62	12.2	82.3	81.9	81.5	81.7
Lite-ResNet (Proposed)	5.62	12.3	91.2	90.8	90.5	90.6

It can be seen from Table 1 that the classification accuracy of the traditional machine learning model SVM+HOG is only 72.3%, which is much lower than that of deep learning models, indicating that manual feature extraction is difficult to capture the deep features of garbage images and has limited classification performance. The lightweight model MobileNetV2 has a parameter size of only 3.4M and an inference time of 10.2ms, showing a significant lightweight effect, but its classification accuracy is 88.3%, which is lower than that of ResNet18 and the proposed Lite-

ResNet model, indicating that there is still room for improvement in its feature extraction ability.

The ResNet18 model achieves a classification accuracy of 90.5%, but it has a large parameter size of 11.7M and an inference time of 25.6ms, with high computational complexity and storage requirements. The proposed Lite-ResNet model achieves a classification accuracy of 91.2% with a parameter size of only 5.62M (48% of ResNet18) and an inference time of 12.3ms (48% of ResNet18), which is 0.7 percentage points higher than that of ResNet18. Its precision, recall, and F1-score are also superior to those of ResNet18, realizing the dual optimization of classification performance and lightweighting.

Comparing different versions of Lite-ResNet, it can be found that the accuracy of the version without data augmentation is 85.6%, which is 5.6 percentage points lower than that of the complete model, indicating that data augmentation technology can effectively expand sample diversity and improve model generalization ability; the accuracy of the version without transfer learning is only 82.3%, which is 8.9 percentage points lower than that of the complete model, verifying the significant advantages of transfer learning on small datasets, which can effectively improve the feature extraction ability and classification performance of the model.

4.3 Confusion Matrix Analysis

To further analyze the classification effect of the Lite-ResNet model on each category of garbage, a confusion matrix of the model on the test set is drawn, and the results are shown in Table 2. The rows of the confusion matrix represent the true categories, and the columns represent the predicted categories. The diagonal elements represent the number of correctly classified samples of each category, and the off-diagonal elements represent the number of misclassified samples.

Table 2 Confusion Matrix of the Lite-ResNet Model on the Test Set (Unit: Number of Samples)

True Category \ Predicted Category	Cardboard	Glass	Metal	Paper	Plastic	Miscellaneous Waste
Cardboard	48	2	1	3	1	0
Glass	1	47	2	0	3	0
Metal	0	1	49	0	2	0
Paper	2	0	0	56	4	1
Plastic	1	3	2	5	42	1
Miscellaneous Waste	0	0	0	2	1	12

It can be seen from Table 2 that the Lite-ResNet model has the best classification effect on metal garbage, correctly classifying 49 samples and misclassifying 3 samples, with an accuracy of 94.2%; followed by cardboard and glass garbage, with accuracies of 90.6% and 88.7% respectively; the number of correctly classified paper samples is 56, and the number of misclassified samples is 7, with an accuracy of 89.7%; the number of correctly classified plastic samples is 42, and the number of misclassified samples is 9, with an accuracy of 82.4%; due to the smallest number of samples (only 15 test samples), the miscellaneous waste has 12 correctly classified samples and 3 misclassified samples, with an accuracy of 80.0%.

The misclassification of the model is mainly concentrated between categories with similar shapes, such as the confusion between plastic and glass, and between paper and cardboard. Both plastic and glass have a certain degree of transparency, and their feature differences are small when the illumination conditions change, leading to 3 glass samples being incorrectly predicted as plastic and 5 plastic samples being incorrectly predicted as glass; paper and cardboard have similar material and texture features, resulting in 3 cardboard samples being incorrectly predicted as paper and 2 paper samples being incorrectly predicted as cardboard. The misclassification of

miscellaneous waste is mainly due to the small number of samples, and the model fails to fully learn its features, leading to 2 samples being incorrectly predicted as paper and 1 sample being incorrectly predicted as plastic. Overall, the model has a good classification effect on each category of garbage with a small number of misclassified samples, which can meet the needs of practical garbage classification.

5. Discussion

The Lite - ResNet model proposed simplifies ResNet structure, introduces transfer learning and data augmentation for efficient garbage image classification on TrashNet dataset. With 5.62M parameters and 12.3ms inference time, it has 91.2% accuracy, offering lightweight advantages and slightly better performance than ResNet18 and MobileNetV2, suitable for edge - computing devices.

Model lightweighting comes from two aspects. First, simplifying network structure by reducing residual block stages and convolution kernels cuts parameters and load. Second, using bottleneck residual block and replacing 3×3 with 1×1 convolution layers reduces complexity. Transfer learning solves limited - sample problem as pre - trained features adapt to garbage image distribution, enhancing extraction and classification ability. Data augmentation expands sample diversity, alleviating imbalance and overfitting, and improving generalization.

Experimental results show the model classifies metal, cardboard, and glass well but has lower accuracy for plastic and miscellaneous waste, indicating class imbalance and feature similarity. Future optimizations include expanding dataset for minority categories and complex - background images, introducing attention mechanism to focus on main features, and fusing multi - modal features.

Currently, the model classifies 6 TrashNet categories. In real - world, domestic waste has more and complex categories. Future work involves expanding dataset coverage, optimizing model structure, and deploying on embedded devices like NVIDIA Jetson Nano for practical - scenario testing and optimization to verify practicality and stability.

References

[1] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

[3] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[4] Zhang, Xiangyu, et al. "Shufflenet: An extremely efficient convolutional neural network for mobile devices." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[5] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." arXiv preprint arXiv:1602.07360 (2016).

[6] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2009, 22(10): 1345-1359.

[7] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2009: 248-255.