Garbage Image Classification based on Deep Learning

Wanqing Wang

College of Century, Beijing University of Posts and Telecommunications, Beijing, China
wangwanqing@stu.ahu.edu.cn

Abstract. Today, there are many disadvantages to using manual sorting for refuse classification. How can we solve the problem of garbage classification efficiently and correctly? It is necessary to solve it at present. In order to solve this problem, researchers have begun to use deep learning technology to sort waste in recent years and have come up with some effective methods. The application and development of different deep learning models in garbage classification are introduced from the aspects of methods and principles. In order to avoid duplication of work by other researchers, improve the significance and value of research, help other researchers to be familiar with and understand the existing research results of deep learning garbage classification, find out the frontier problems of these models, and expand the research ideas and methods of deep learning garbage classification. In this paper, three models are analyzed: ResNet50 model, transfer learning model, Inception-v3 model, and network model combining ResNet and SENet, and a new feasible model is proposed.

Keywords: ResNet50, Learning model, Inception v3, SENet, Garbage image classification

1 Introduction

Garbage categorization is the process of dividing trash into categories based on whether it can be recycled or not. Humans produce a lot of trash each day, and a lot of that trash is recycled carelessly and dumped without classification, which pollutes the environment. Therefore, to classify garbage, maximize the use of garbage, and save energy. However, at present, a large number of garbage classifications in my country still use manual classification, and most people lack knowledge of garbage classification, so the difficulty of garbage classification in my country is increasing. There are many applications in Tao Hang et al. that developed a method for classifying garbage images based on the ResNet50 model and transfer learning and designed an automatic garbage classification system based on deep learning, which realized the automatic classification of common garbage in daily life [1-3]. Chen Feiyu et al used the Inception-v3 model to migrate the training results to achieve the correct classification of cardboard, plastic, metal, glass, waste paper, and alternatives [4, 5]. Wang Yu et al. used a network model combining ResNet and SENet to

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_42

implement a composite image classification algorithm based on a deep residual shrinkage network [6-8]. This paper adopts a clearer model for analysis.

2 Methods

A sophisticated machine learning technique known as "deep learning" discovers the internal rules and representational depths of sample data. The knowledge gained during these learning processes is extremely helpful when interpreting data like text, visuals, and sounds. Deep learning has achieved many results in many fields such as data mining [9], speech recognition [10], image recognition, and so on [11]. Therefore, many garbage classification methods based on deep learning have been developed. This section describes the deep learning garbage classification methods for the combination of the ResNet50 model and migration learning model, the Inception-v3 model, and the combination of ResNet and SENet.

2.1 ResNet50 Model Based on Transfer Learning

ResNet50 Model. The ResNet network realizes the modeling of the deep network by introducing a residual module. This structure can effectively alleviate the deep network's training process issues with gradient explosion and gradient vanishing.

The residual block is the most important part of ResNet50, and its structure is shown in Figure 1. Each residual module consists of three convolutional layers and a skip connection, where the convolutional kernel size of the first convolutional layer and the third convolutional layer are both 1x1 and the convolutional kernel of the second convolutional layer size is 3x3. Between each convolutional layer and skip connection, The addition of a ReLU activation function and a batch normalization layer. Figure 1 depicts the ResNet50 network structure. The five sections of the network structure are shown here as I, II, III, IV, and V. In the first part, the input image is convolved with $7\times7\times64$, and the convolution of the nonlinear activation function ReLu and the maximum pooling 3×3 is used to output the feature map of $65\times56\times56$. The following 4 parts All have a residual module structure, and each residual module contains three convolution operations. There are a total of $1+(3+4+6+3)\times3=49$ convolution operations in the network, and then a total of 50 convolution operations through the fully connected layer. Layer, after each part, the size of the picture is reduced to one-half of the original, and the output feature map is $2048\times7\times7$. Finally, the average pooling is performed to obtain a one-dimensional feature vector, which is calculated and input after entering the classifier. Probabilities for each category.

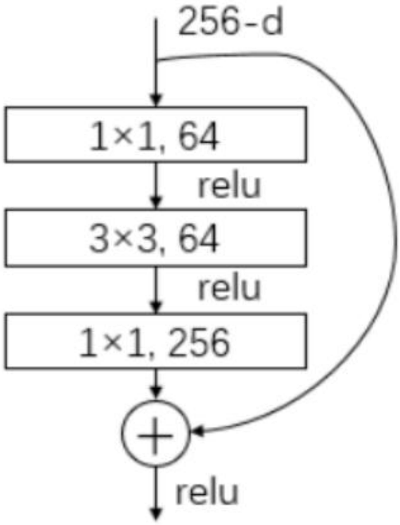

Fig. 1. Residual learning: a building block [2]

Transfer Learning. By adding a fully connected layer to the improved network model in accordance with the number of classification targets, transfer learning is a machine learning technique that uses the model created by task A as the starting point and reuses it in task B. As seen in Figure 2, trials with various categorization purposes exhibit strong generalizability. Although it takes a lot of data and computing power to train a deep learning model, the amount of data utilized in most circumstances is quite limited, direct training will result in model non-convergence or over-fitting, and transfer learning has a detrimental impact on deep learning models. The modest requirement for training data effectively resolves this issue. However, a crucial problem with transfer learning is that job A from the source domain and task B from the target domain need to be somewhat comparable or related.

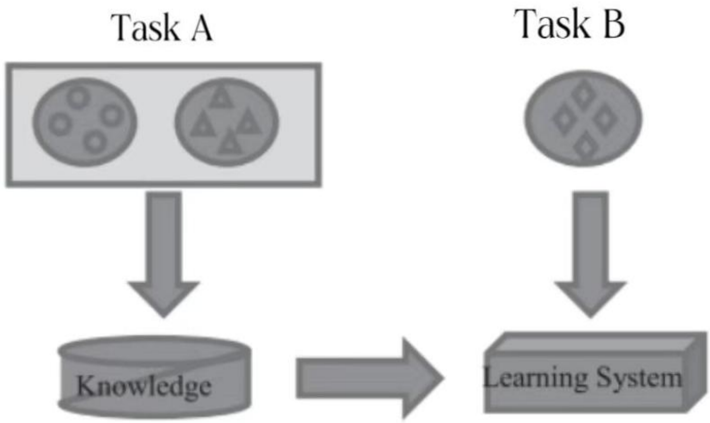

Fig. 2. A Schematic diagram of the transfer learning process [12]

Combination of ResNet50 Model and Transfer Learning. The model is first loaded into the ResNet50 network model, and the training weights based on the ImageNet dataset are imported to obtain a pre-trained model. Because the total number of garbage classifications for this project is 40, the output of the modified classifier is 40. Secondly, retrain on the self-made garbage data set. During the training process, the structural parameters of other layers are frozen, and only the structural parameters of the fully connected layer are changed. The input of the last fully connected layer is connected to a linear layer with 256 output units, followed by a ReLU layer, then a 256 × 40 linear layer and the output is a 40-channel softmax layer. The selection of Softmax classifier to realize the classification of garbage, and finally the cross entropy loss function and Adam gradient optimization method to train the model. The selection of Softmax classifier to realize the classification of garbage, and finally the cross entropy loss function and Adam gradient optimization method to train the model.

2.2 Method based on Inception-v3 Model

Inception v3 model. In order to create different-sized receptive fields, Inception employs convolution kernels of various sizes. Splicing is then used to combine features at various scales. Figure 3 depicts the entire model architecture of Inception-v3. The model comprises 11 Inception modules and has a total of 46 layers. The Inception model combines different convolutional layers in parallel, a total of 96 convolutional layers [13]. Convolution decomposition and auxiliary classifier design are introduced in this model in contrast to the conventional convolutional neural network. Convolution decomposition is the process of breaking down a big convolution factor into smaller, more manageable components. This reduces the number of parameters and increases computational performance. In order to increase the convergence of deep neural networks, auxiliary classifiers are used. Its function is to act as a regularizer. The primary goal is to move helpful gradients to lower levels so that they are instantly accessible. The vanishing gradient problem is solved in the network to improve the convergence during training.

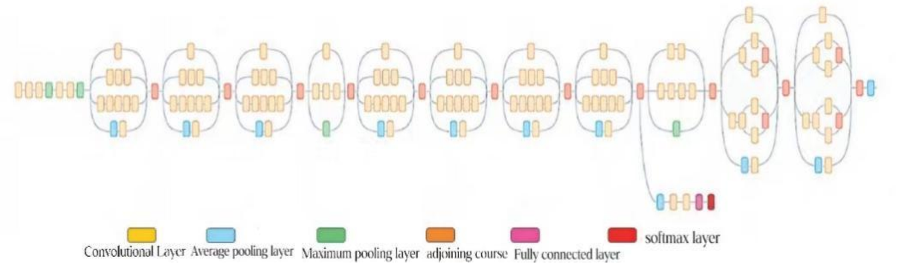

Fig. 3. Inception v3 model [4]

Inception v3 combined with transfer learning [14]**.** The Inception-v3 model is used to pre-train the parameter model trained on the large-scale image classification data ImageNet. The feature extraction model uses the network structure and parameters of the pre-trained model. By inputting and processing a garbage image, extracting the 2048-dimensional vector features of the image, freezing the convolutional layer before the fully connected layer and the Softmax layer of the Inception-v3 model, and training a new the fully connected layer and Softmax layer are used to extract image features in-depth to complete garbage image classification. The domestic waste classification model requires less data, and the training time is much shorter than that of directly using the deep learning model. Considering the size of the image data set and computing GPU resource factors, the parameter selection adopts 16 random selection methods as batch training evaluation, adjusts the weight of model parameters through the backpropagation algorithm, and completes the final training of the model through iteration. The migration learning method in the Inception-v3 pre-training model can only retrain from the updated layer and keep the original model's weights and biases, which can save a lot of time.

2.3 Combining ResNet and SENet

SENet. After SENet undergoes a series of convolution operations on the feature x, there is a skip connection directly connected to the operation of the next layer. SENet adds Squeeze, Excitation, and Reweight operations between the two layers, and implements the attention mechanism through these three operations [15]. The SENet module's schematic is shown in Figure 4. The first is the Squeeze operation[16], which turns each two-dimensional feature channel into a real number while compressing features along the spatial axis. This one-dimensional feature broadens the global receptive field and has a global vision.

The Excitation operation, a mechanism akin to gates in recurrent neural networks, is the second. The one-dimensional channel feature obtained by the Squeeze operation adds a fully connected layer to the operation and creates weights for each feature channel using the parameter w, where w is a multi-layer perceptron's equivalent. A reweight operation comes last. Following feature selection, the relevance of each feature channel is determined by the weight of the Excitation output, which is then multiplied channel by channel to complete the original channel dimension. Modification of a feature.

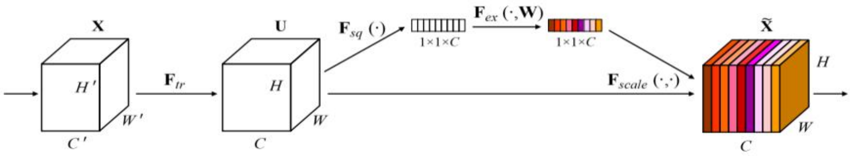

Fig. 4. Schematic of the SENet module [16]

Deep residual shrinkage network. Two traditional neural networks (ResNet and SENet) make up the deep residual shrinkage network. In this network, soft thresholding takes the place of SENet's Reweight function, and a branch line is added to learn the thresholds needed for soft thresholding, giving each feature its own weight. Figure 5 displays the basic modular structure of it.

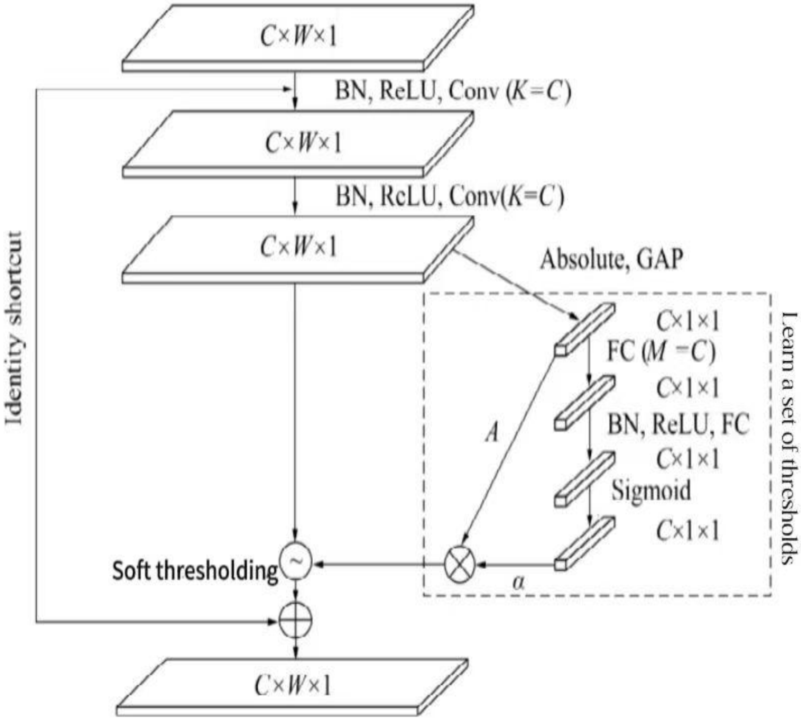

Fig. 5. The identity block of deep residual shrinkage network [16]

Implemented a deep residual shrinkage network-based image classification algorithm for school trash [17]. The model used in the algorithm consists of four parts, namely the convolutional layer, the residual layer combined with ResNet and SENet, the pooling layer, and the fully connected layer. Compared with traditional convolutional neural networks, this model has the advantages of avoiding overfitting and fast convergence. The SENet network reduces the impact of noise data on the accuracy rate and finally achieves a higher recognition rate. It is feasible for garbage classification tasks. The trained model has achieved better performance when applied to the campus garbage recognition app.

The overall process of garbage classification is shown in Figure 6. Adjust the size of the image before classification to achieve image preprocessing. Preliminary features are obtained after a convolution operation, and more refined features are obtained through many basic modules of the residual shrinkage network. After a completely connected layer and a global average pooling layer, the most likely classification result is finally obtained through the softmax classifier [18].

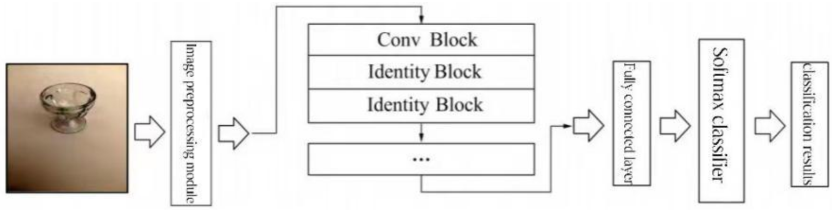

Fig. 6. Flow chart of intelligent waste classification [6]

3 Research and Analysis

3.1 Three Model Analysis

Based on the ResNet50 model, combined with the transfer learning method, only part of the parameters are trained during the training process. Reduce the number of parameters for retraining, and on the basis of the original model parameters, it can be used as a pre-classification model of the network, and a classification model with higher accuracy can be obtained under the action of the Softmax function and the fully connected layer. The transfer learning method reduces the training parameters and improves the training efficiency.

The method based on the Inception-v3 model can reduce misclassifications caused by subjective factors, and reduce image imaging requirements through data enhancement methods. It is not sensitive to illumination, distance, and size, and has a high robustness and generalization ability. Compared with the traditional convolutional neural network, the model based on the combination of ResNet and SENet has the advantages of avoiding over-fitting and fast convergence. The SENet network reduces the impact of noise data on the accuracy rate and finally achieves a higher recognition rate. It is feasible for garbage classification tasks. The trained model has achieved better performance when applied to the campus garbage recognition app.

3.2 New model ResNeXt

The ResNeXt structure is shown in Figure 7 and Figure 8. ResNeXt proposes aggregated transformations, which replace the initial ResNet block with three layers of convolution and a parallel stack of blocks with the same structure. As shown in the data in Figure 9, ResNeXt is better than ResNet and Inception-ResNet-v2 and has a lower error rate under the same calculation amount [19].

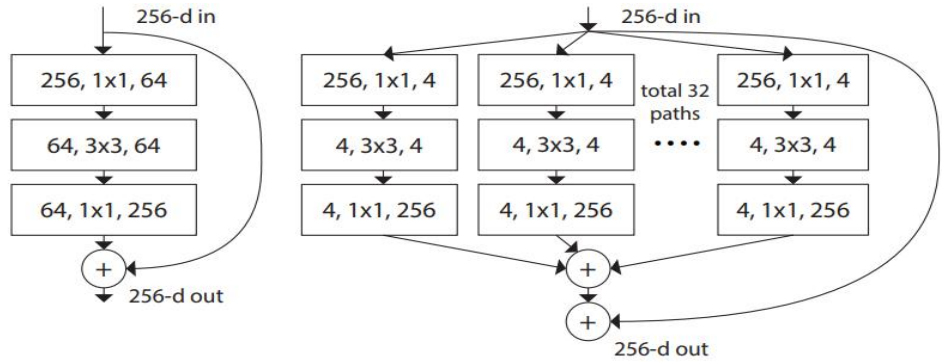

Fig. 7. ResNet and ResNeXt [19]

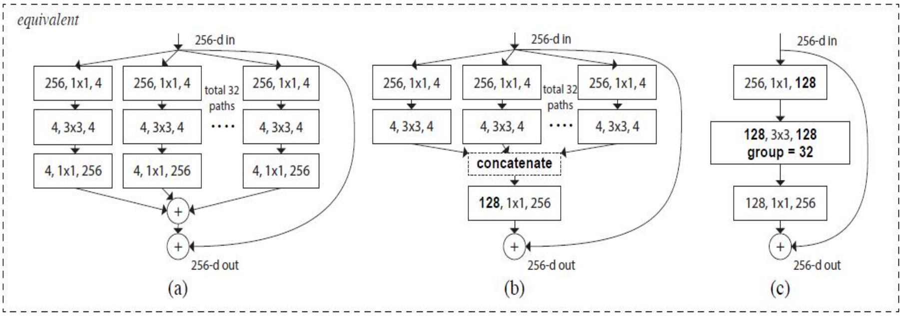

Fig. 8. ResNeXt's equivalent building blocks [19]

	224×224		320×320/299×299	
	top-1 err	top-5 err	top-1 err	Top-5 err
ResNet-101[14]	22.0	6.0	-	-
ResNet-200[15]	21.7	5.8	20.1	4.8
Inception-v3[39]	-	-	21.2	5.6
Inception-v4[37]	-	-	20.0	5.0
Inception-ResNet-v2[37]	-	-	19.9	4.9

ResNeXt-101(64×4d)	20.4	5.3	19.1	4.4

Fig. 9. State-of-the-art models on the ImageNet-1K validation set [19]

4 Conclusion

This section summarizes the research status of three deep learning models in the field of garbage classification, introduces the classical garbage classification method ResNet50 model and transfer learning model Inception-v3 model and their structure, based on ResNet50 model, combined with transfer learning method, only some parameters are trained in the training process, the training efficiency is improved, the method based on the Inception-v3 model can reduce misclassifications caused by subjective factors， the model based on the combination of ResNet and SENet has the advantages of avoiding over-fitting and fast convergence. Moreover, this paper a new executable model ResNeXt is proposed. ResNeXt proposes aggregated transformations, which replace the initial ResNet block with three layers of convolution and a parallel stack of blocks with the same structure. As shown in the data in Figure 9, ResNeXt is better than ResNet and Inception-ResNet-v2 and has a lower error rate under the same calculation amount. Putting ResNet and Inception-ResNet-v2 on the same amount of calculation has a lower error rate. although deep learning-based garbage classification methods are the future development trend, there are still many challenges in the face of high accuracy and strong real-time requirements. Therefore, researchers should identify the shortcomings of numerous methods and conduct deeper research.

References

1. Tao H, Jiang X, Zhang J, etc. A Garbage Classification System Based on Deep Learning. Journal of Hubei Automotive Industries Institute,2022,36(02):36-39+44.
2. He K, Zhang X, Ren S, etc. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
3. Chen X. Research on Deep Learning Algorithms and Applications Based on Convolutional Neural Networks. Zhejiang Gongshang University, 2014.
4. Chen F, Yang T, Jiang M. Design of a Domestic Waste Classification Model Based on Deep Learning Technology. Design of Household Garbage Classification Model for Electronic Components and Information Technology [J]. Electronic Components and Information Technology,2020, 4(07): 94-96.
5. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna, Rethinking the Inception Architecture for Computer Vision, Submitted on 2 Dec 2015 (v1), last revised 11 Dec 2015 (this version, v3)6

6. Wang Y, Zhang Y, Zhou Y, etc. Campus Garbage Image Classification Based on Deep Residual Shrinkage Network [J]. Journal of Jilin University (Information Science Edition), 2023,41(01):186-192.
7. Jie H，Li S，Gang S，etc．Squeeze-and-Excitation Networks．IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020, 42(8): 2011-2023．
8. Zhao M, Zhong S, Fu X, etc. Deep Residual Shrinkage Networks for Fault Diagnosi. IEEE Transactions on Industrial Informatics. 2020, 16(7): 4681-4690．
9. Jaisal E K. Data-mining and analytics： rising concerns over privacy and people's security. Proceedings of the 42nd Indian Social Science Congress Conference. New Delhi， Indian：2018：346-455.
10. Hinton G E,Deng L,Yu D,etc.Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups[J].IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
11. Gong M, Yang H, Zhang P, etc. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images[J].Isprs Journal of Photogrammetry and Remote Sensing, 2017: 212-225.
12. Li Y, Mu L. Research on Garbage Image Classification Model Based on Transfer Learning [J]. Computer and Information Technology,2021,29(04):17-21.
13. Li H, Guo H. "Design of Bionic Robotic Hand Gesture Recognition System Based on Machine Vision", 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), 2022
14. Dong N, Zhao L, Wu H, Chang J."Inception v3 based cervical cell classification combined with artificially extracted features", Applied Soft Computing, 2020
15. Li Q， Miao N， Zhang X， etc. Image Recognition of Maize Disease Based on Attention Mechanism Asymmetric Residual Network and Transfer Learning. Science Technology and Engineering, 2021, 21(15): 6249-6256．
16. Yang B, Tang M, Chen S ,Wang G, Tan W,Li B. "A vehicle tracking algorithm combining detector and tracker", EURASIP Journal on Image and Video Processing, 2020
17. Zhao J, Chen W. "Research on image classification algorithm based on depth residuals shrinkage network in Commercial Image Library", Journal of Physics: Conference Series, 2021
18. Wan L, Tong X, Sheng M, etc.Overview of Softmax Classifier Deep Learning Image Classification Methods Application. Navigation and Control, 2019, 47(6): 1-9．
19. Xie S, Ross Girshick, Piotr Dollár, Tu Z, He K, Aggregated Residual Transformations for Deep Neural Networks, Submitted on 16 Nov 2016 (v1), last revised 11, 2017.