ICIC Express Letters
Part B: Applications
Volume **13**, Number **6**, June **2022**
ICIC International ©2022 ISSN 2185-2766
pp. 631–638

APPLICATION OF SMALLER CNN INTEGRATED WITH IG-MODULE IN RECYCLABLE GARBAGE CLASSIFICATION

Changjiang Li[1], Meng Xu[1,*], Xikun Xie[1] and Zhong Zhang[2]

[1]School of Mechanical and Power Engineering
Chongqing University of Science and Technology
Huxi University Town, Shapingba District, Chongqing 401331, P. R. China
{ lcjgdm; shelldon_xikun }@163.com; *Corresponding author: Xu_Meng1hao@163.com

[2]Department of Intelligent Machinery
Hiroshima Institute of Technology
2-1-1, Miyake, Saeki-ku, Hiroshima 731-5193, Japan
t.sho.g4@cc.it-hiroshima.ac.jp

Received October 2021; accepted December 2021

Abstract. *Due to a wide range of shapes, colors, and other characteristics of recyclable garbage, the classification based on machine vision shows poor real-time capability and lower accuracy, making it hard to meet the needs of practical application in industry. In this paper, we propose the Slight-SqueezeNet-IGN by integrating the improved SqueezeNet with IG-module. First, to avoid excessive loss of image features caused by the SqueezeNet, the Slight-SqueezeNet is constructed by changing first Convolutional layer and Maxpooling layer of the SqueezeNet. Second, the IG-module is proposed to increase the model's ability of feature extraction by concurrently designing the IN (Instance Normalization) and GN (Group Normalization). Finally, we integrate the IG-module into Slight-SqueezeNet to form Slight-SqueezeNet-IGN. The experimental results demonstrate that compared with AlexNet and VGG16, Slight-SqueezeNet-IGN can achieve higher accuracy with fewer parameters, showing strong practicability in recyclable garbage classification.*
Keywords: IG-module, Slight-SqueezeNet-IGN, Recyclable garbage classification

1. **Introduction.** With the rapid development of machine vision and artificial intelligence, using machines to replace manual work will be a trend in the domain of garbage classification. Due to the variety and shape of recyclable garbage, there is no mature visual processing product that meets the requirements of sorting accuracy, especially the visual recognition technology collaborated with industrial mechanical arm in real time. Recent years have witnessed remarkable achievements of deep learning technology in the field of image recognition. Therefore, many scholars have applied this technology to garbage classification with some problems remaining unsolved. On the TrashNet dataset, Rabano et al. [1] applied a smaller network MobileNet [2] to achieving an accuracy of 87.2%; Ruiz et al. [3] used ResNet [4] to achieve an average accuracy of 88.66%; Aral et al. [5] employed DenseNet121 and DenseNet169 [6] to increase accuracy to an average of 95%; and Ozkaya and Seyfi [7] adopted GoogleNet [8] with Support Vector Machine (SVM) [9] to achieve an accuracy of 97.86%. However, there are several problems. For one thing, the large convolutional neural network is applied to garbage. It is difficult to meet the requirements of timeliness. At the same time, the smaller convolutional neural network cannot meet the accuracy requirements. And, the above experimental results are all obtained from the TrashNet dataset with a single background, which is difficult to meet the actual requirements of the industrial application under complex background.

DOI: 10.24507/icicelb.13.06.631

Our contributions can be summarized as follows.

1) We form the recyclable garbage classification dataset with complex background by integrating the Huawei dataset with complex background into the TrashNet dataset. Compared with the TrashNet dataset, this dataset can more meet the needs of industrial applications.

2) An improved method of SqueezeNet is proposed, which can effectively improve the recognition ability of SqueezeNet.

3) A new model IG-module is proposed. It can improve the performance of the model without adding too many parameters.

4) In combination with the above two improvements, we propose a new smaller model called Slight-SqueezeNet-IGN. It can solve the industrial application difficulties caused by the large model and the insufficient generalization ability.

2. Theory Introduction and Related Work.

2.1. SqueezeNet. SqueezeNet is a smaller Convolutional Neural Network (CNN) proposed by Iandola based on the Inception model. The number of parameters is only 0.72 M, which is equivalent to 1/50 of the number of AlexNet [11,12] parameters, but the accuracy of Top-1 and Top-5 in ImageNet is similar to AlexNet. The model structure is shown in Figure 1.

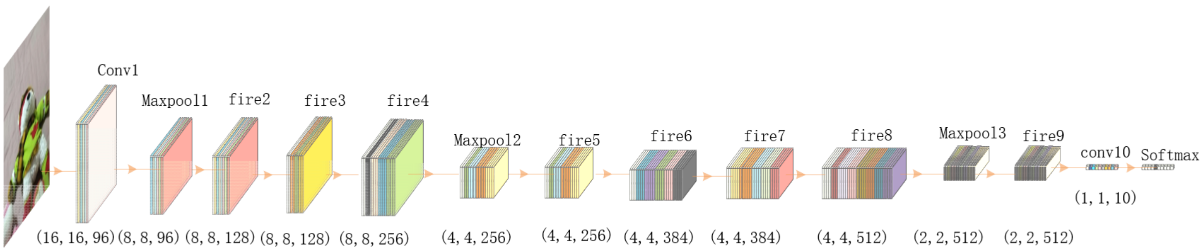

Figure 1. Structure of the SqueezeNet

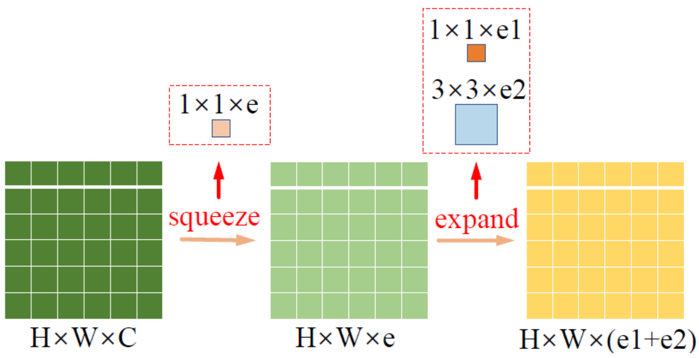

Figure 2. Structure of the fire module

In Figure 2, H represents the height of the feature image, W represents the width of the feature image, and C represents the number of channels of the feature image. Compared with the traditional CNN, the most important change is the proposed fire module. The structure of the fire module is shown in Figure 2, which contains two parts: the squeeze layer and the expand layer. In the squeeze layer, e convolution kernels of size 1 are used to decrease the dimensions of input images, and in the expand layer, $e1$ convolution kernels of size 1 and $e2$ convolution kernels of size 3 are used to increase the dimensions of output images.

2.2. **IG-modules.** IG-module is a normalization method that combines instance normalization [13,14] and group normalization [15]. I is the abbreviation of IN (Instance Normalization) layer. IN can accelerate model convergence while maintaining the independence between images. Therefore, IN is used to preserve most of the features of the input image. In the early stage, it was mostly used in image style transfer. G is the abbreviation of GN (Group Normalization) layer. The GN can normalize the feature image without being affected by the parameter of batch. So GN is used to learn new features and to distinguish individual samples easily. However, it also makes the model vulnerable to the influence of image appearance transformation. The use of the IN weakens individual differences while reducing the amount of useful information in the image. Because both approaches have their limitations, IN and GN are combined to form the IG-module in this paper. The IG-module is used to reduce the loss of image characteristic information because it maintains a certain degree of appearance invariance while maintaining the differences of the input images. Therefore, the use of IG-module is conducive to the improvement of model generalization ability.

The calculation formulas for IN and GN are as follows.

$$\hat{x} = \frac{1}{\sigma_i}(x_i - \mu_i) \tag{1}$$

Here x is the feature computed by a layer, and i is an index. In the case of 2D images, $i = (iN; iC; iH; iW)$ is a 4D vector indexing the features in $(N; C; H; W)$ order, where N is the batch axis, C is the channel axis, and H and W are the spatial height and width axes. μ and σ in (1) are the mean and standard deviation (std) computed by

$$\mu_i = \frac{1}{m}\sum_{k \in S_i} x_k \tag{2}$$

$$\sigma_i = \sqrt{\frac{1}{m}\sum_{k \in S_i}(x_k - \mu_i)^2 + \epsilon} \tag{3}$$

where ϵ is a small constant, S_i is the set of pixels in which the mean and std are computed, and m is the size of the set.

GN refers to the standardization of several channels on the same layer of the same picture, among which these channels are called a group. Its S_i is calculated in this way.

$$S_i = \left\{k_N = i_N, \left\lfloor \frac{k_C}{C/G} \right\rfloor = \left\lfloor \frac{i_C}{C/G} \right\rfloor \right\} \tag{4}$$

IN standardizes the individual channels of a single picture. The formula for calculating S_i is as follows.

$$S_i = \{k_N = i_N, k_C = i_C\} \tag{5}$$

Finally, both IN and GN learn a per-channel linear transform to compensate for the possible loss of representational ability.

$$y_i = \gamma \widehat{x_i} + \beta \tag{6}$$

where γ and β are trainable scale and shift (indexed by iC in all case, which we omit for simplifying notations).

2.3. **Improvement strategy and integration strategy.** The Slight-SqueezeNet-IGN in this paper is proposed after the improvement and integration of the smaller SqueezeNet.

Step1: The error rate of the model will increase due to excessive loss of image information caused by the SqueezeNet. Therefore, to reduce the loss of image information and increase the generalization performance of the model, slight modification and deletion of the SqueezeNet are made to form a novel Slight-SqueezeNet model in this paper. The improvement strategies of SqueezeNet are as follows.

1) Slight-SqueezeNet-1: proposed by removing the first Maxpooling layer.

2) Slight-SqueezeNet-2: proposed by removing the Maxpooling layer after the eighth fire module in SqueezeNet.

3) Slight-SqueezeNet-3: proposed by changing the step size of the first Convolutional layer of SqueezeNet to 1 and deleting its first Maxpooling layer.

4) Slight-SqueezeNet-4: proposed by changing the step size in the first Convolutional layer of SqueezeNet to 1 and removing the Maxpooling layer after the eighth fire module.

5) Slight-SqueezeNet-5: The kernel size in the first Convolutional layer of the SqueezeNet is changed to 5, the step size is changed to 1, and the zero-padding is deleted. Then a zero-padding of the last Maxpooling layer is deleted.

6) Slight-SqueezeNet-6: Firstly, the kernel size in the first Convolutional layer of SqueezeNet is changed to 5, the step size is changed to 1, and the zero-padding is deleted. Secondly, a zero-padding of the last Maxpooling layer is deleted. Thirdly, the first Maxpooling layer in the SqueezeNet is removed.

Step2: To increase the capability of feature extraction and generalization of the model, IG-module is fused into the Slight-SqueezeNet formed after Step1. The structures formed by different fusion strategies are shown in Figure 3.

1) Slight-SqueezeNet-IGN-1: proposed after fusing the IG-module into the squeeze layer of the fire module.

2) Slight-SqueezeNet-IGN-2: proposed after fusing the IG-module into the branch of kernel size 1 of the expand layer of the fire module.

3) Slight-SqueezeNet-IGN-3: proposed after fusing the IG-module into the branch of kernel size 3 of the expand layer of the fire module.

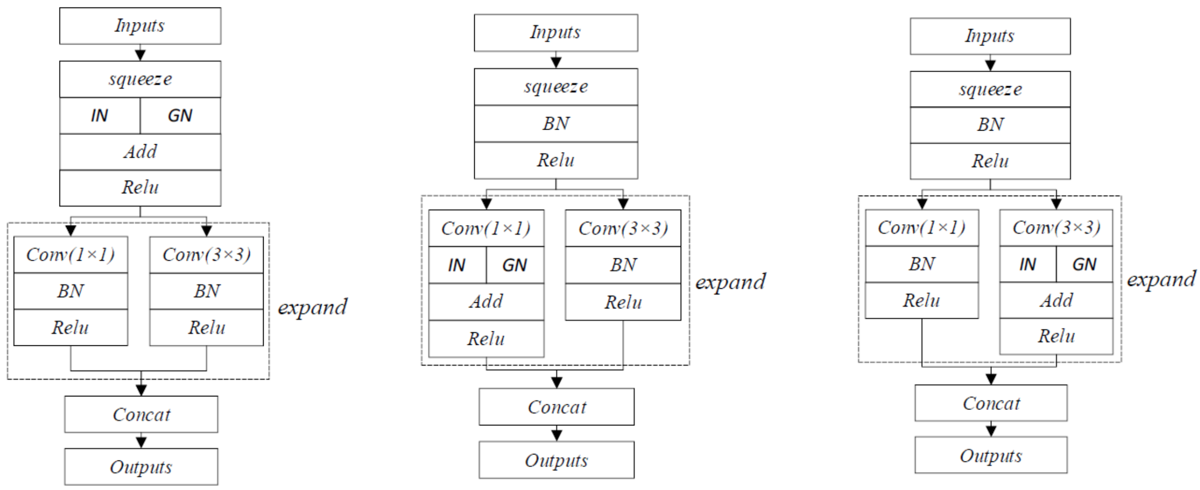

FIGURE 3. Structures of different mix strategies of the Slight-SqueezeNet-IGN

3. Experimental Results and Analysis.

3.1. **CIFAR10.** It contains 10 categories: airplane, car, bird, cat, dog, deer, frog, horse, boat, and truck. Each category contains 6000 color pictures, and the size of each picture is 32×32, totaling 60000 images. Training set, validation set, and test set constitute the dataset, of which 40000 are training set, 10000 validation set, and 10000 test set. On the training set, the image is standardized, flipped horizontally, translated and so on. However, on the validation set and the test set, only normalization is done.

3.2. **Recyclable Waste Dataset.** The dataset in this paper comes from the TrashNet dataset [16] and the Huawei dataset. Because the TrashNet cannot meet the practical application needs of the industry due to a small amount of data and a single background, Huawei dataset and TrashNet are integrated to form the Recyclable Waste Dataset with the complex background [17]. And five categories of images (fabric, metal, glass, paper, and plastics) are filtered out.

3.3. **Experimental environment settings and model parameter settings.** The Keras and Tensorflow are used as the deep learning open-source framework in this paper. The Central Processing Unit (CPU) is Inter(R) Core (TM) i7-7700. The GPU (Graphics Processing Unit) is NVIDIA GeForce GTX 1050Ti (4 GB). The Stochastic Gradient Descent (SGD) is chosen as optimize function, the epochs are setting to 300 times, the initial learning rate to 0.1 and minimum learning rate to 0.001.

3.4. **Experimental analysis.** In this section, we carry out experiments with three different purposes to prove the effectiveness of the improved models and their practicability in the field of recyclable garbage classification. The results are as follows.

3.4.1. *Experiment 1 result analysis.* On the CIFAR10 dataset, experiment 1 compares six models proposed according to the improvement strategy with the SqueezeNet from four aspects: accuracy, parameter quantity, floating point operations (FLOPs), and time/one iteration. The distribution of the four aspects of each model is shown in Table 1. In Figure 4, the variety rule of accuracy is shown.

TABLE 1. Comparison of models of the Slight-SqueezeNet

CNN architecture	Accuracy	Parameters	FLOPs	Time/one iteration
SqueezeNet	79.12%	741088	3687435	47 ms
Slight-SqueezeNet-1	87.06%	741088	3687435	63 ms
Slight-SqueezeNet-2	84.13%	741088	3687435	46 ms
Slight-SqueezeNet-3	87.44%	741088	3687435	174 ms
Slight-SqueezeNet-4	85.91%	741088	3687435	74 ms
Slight-SqueezeNet-5	87.22%	745696	3710475	64 ms
Slight-SqueezeNet-6	87.06%	745696	3710475	136 ms

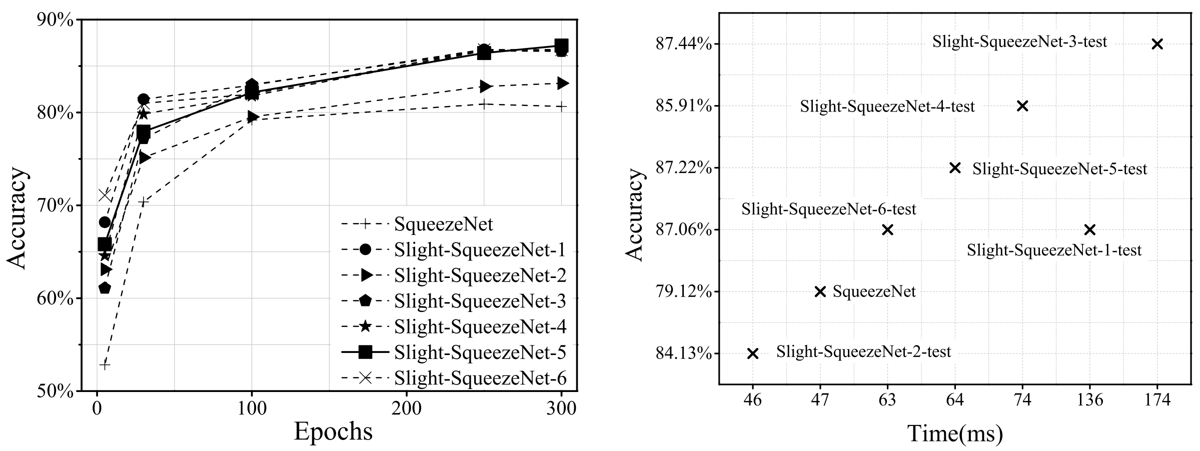

FIGURE 4. **Left:** The accuracy of the models on the validation set; **Right:** Accuracy and time distribution of the models

From Table 1, the accuracy of the SqueezeNet is only 79.12%, and six Slight-SqueezeNet formed after Step1 exceed the SqueezeNet. Meanwhile, both the number of parameters and the complexity of the model do not increase significantly. The result verifies the effectiveness of the improvement strategy Step1 proposed.

Among all the models, the Slight-SqueezeNet-3 has the highest accuracy, but the training time increases by more than three times compared to the original one. Therefore, if the improvement strategy is adopted, it will consume a lot of training time, which is not conducive to the follow-up work. It can be seen that from Table 1 Slight-SqueezeNet-5 shows great advantages with the accuracy of 87.22%. Firstly, the accuracy of Slight-SqueezeNet-5 only decreases by 0.22% compared with Slight-SqueezeNet-3. Secondly, the training time only increases by 17 ms, compared with the SqueezeNet, and it decreases by 110 ms compared with the Slight-SqueezeNet-3. Slight-SqueezeNet-5 can achieve high accuracy in a relatively short training time, so it is more conducive to the development of subsequent experimental work. Therefore, the improved strategy 5 is adopted in this paper and named Slight-SqueezeNet, which functions as the basis for subsequent experimental work.

3.4.2. *Experiment 2 result analysis.* On the CIFAR10 dataset, experiment 2 compares six models, AlexNet-like, VGG16-like, Slight-SqueezeNet, Slight-SqueezeNet-IGN-1, Slight-SqueezeNet-IGN-2, and Slight-SqueezeNet-IGN-3 from four aspects: accuracy, parameter, FLOPs, and time spent to predict a picture. The experimental results are shown in Table 2. The variety rule of accuracy is shown in Figure 5.

TABLE 2. Comparison of models of the Slight-SqueezeNet-IGN

CNN architecture	Accuracy	Parameters	FLOPs	Time/predict one picture
AlexNet-like	81.28%	3631946	8833472	42.88 ms
VGG16-like	88.88%	15142858	74406852	85.80 ms
Slight-SqueezeNet	87.22%	745696	3710475	92.75 ms
Slight-SqueezeNet-IGN-1	87.53%	745072	3704315	87.73 ms
Slight-SqueezeNet-IGN-2	87.03%	743152	3686075	86.76 ms
Slight-SqueezeNet-IGN-3	87.59%	743152	3686075	84.88 ms

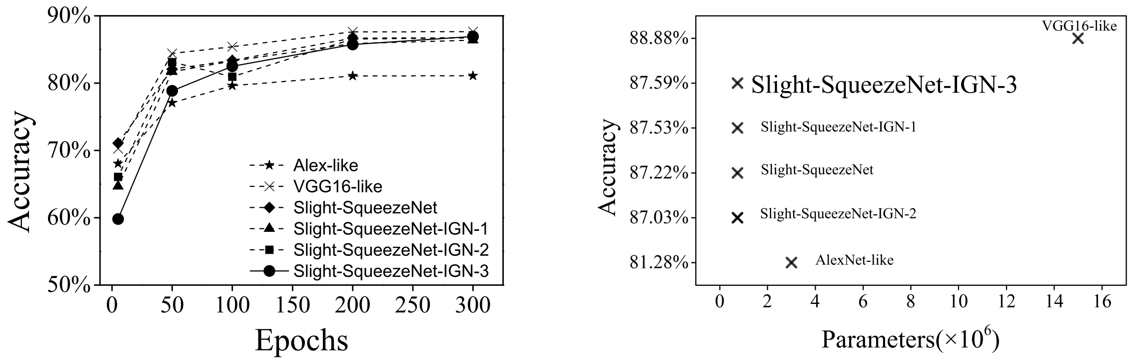

FIGURE 5. **Left:** The accuracy of the models on the validation set; **Right:** Parameters and accuracy distribution of the models

Firstly, it can be concluded from the comparison of data and curves of Slight-SqueezeNet, Slight-SqueezeNet-IGN-1, Slight-SqueezeNet-IGN-2, and Slight-SqueezeNet-IGN-3 in Table 2 and the two models, Slight-SqueezeNet-IGN-1 and Slight-SqueezeNet-IGN-3, achieve better results with an increase in accuracy of 0.31%-0.37% compared with Slight-SqueezeNet without much change in model parameters and model complexity. The validity of IG-module is verified by experiments.

Secondly, the model data and curves of Slight-SqueezeNet-IGN-1, Slight-SqueezeNet-IGN-2, and Slight-SqueezeNet-IGN-3 in Table 2 are compared. Adding IG-module to the branch with a kernel size of 3 in the fire module can achieve the best accuracy of 87.59%

without increasing the number of parameters and the complexity of the model. Slight-SqueezeNet-IGN-3 also has a slight advantage in predicting the total time consumed by an image. Therefore, this paper uses the fusion strategy 3 proposed by Step2, and names the model as Slight-SqueezeNet-IGN. Then we use the model for subsequent experiments.

Finally, the data and curves of the Slight-SqueezeNet-IGN-3, the AlexNet-like, the VGG16-like in Table 2 and Figure 5 are compared. The accuracy of the AlexNet-like model on CIFAR-10 is about 81.28%, which performs poorly, but the time is shorter. VGG16-like performs well, but its parameters amount to more than 15 million, which is difficult to meet the actual application of industrialization. The Slight-SqueezeNet-IGN proposed has only 1/5 of the parameters and model complexity of the AlexNet-like model, and its accuracy can exceed by 6.31%. Compared with the VGG16-like model, although the accuracy is reduced by 1.29%, its model complexity and parameter amount are only 1/20 of it, and the time consumed is reduced by 0.92 ms. In brief, experiments show that the Slight-SqueezeNet-IGN proposed has better recognition ability.

3.4.3. *Experiment 3 result analysis.* In this experiment, the Slight-SqueezeNet-IGN, the AlexNet-like, and the VGG16-like are applied to the Recyclable Waste Dataset for comparison. Table 3 presents the experimental results from four aspects of accuracy, parameter amount, FLOPs, and training time required for one iteration. The variety rule of accuracy is shown in Figure 6 on the Recyclable Waste Dataset.

Table 3. Comparison of various models in the Recyclable Waste Dataset

CNN architecture	Input size	Accuracy	Parameter	BFLOPs	Time/one iteration
AlexNet-like	32×32	93.18%	4413253	0.0088	205 ms
VGG16-like	32×32	95.41%	14717253	0.074	210 ms
Slight-SqueezeNet-IGN	32×32	95.78%	743097	0.0037	200 ms

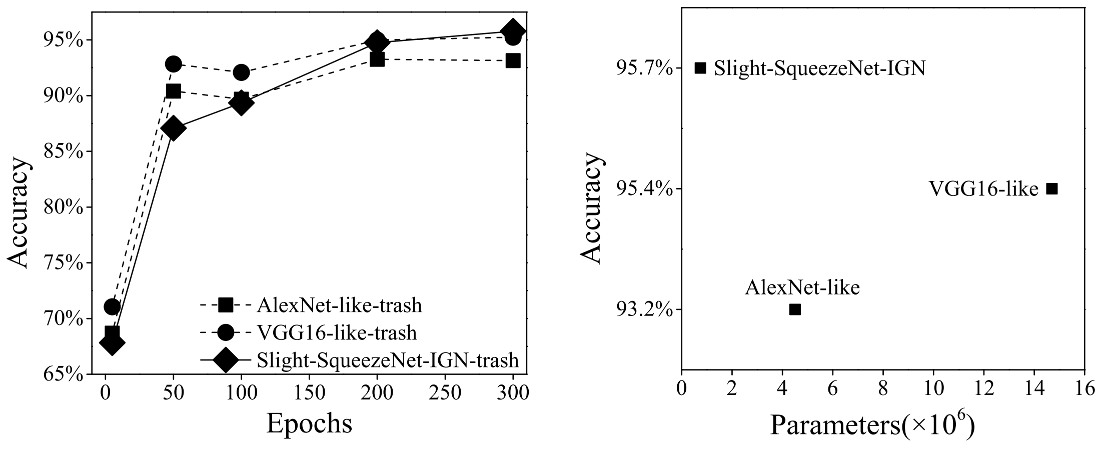

Figure 6. **Left:** The accuracy of the models on the validation set; **Right:** Parameters and accuracy distribution of models

We can see this in Table 3. The Slight-SqueezeNet-IGN converges at approximately 90 iterations, with 99% accuracy on the training set and 95.78% accuracy on the test dataset. The accuracy increases by 2.60% compared with the AlexNet-like, and by 0.37% compared with the VGG16-like. However, its parameters only need 0.74 M, and the complexity of the model is far less than VGG16-like. It takes about 200 ms to train an iteration. Therefore, the Slight-SqueezeNet-IGN not only meets the needs of a smaller model, but also strong recognition and generalization ability, and it can be better applied to the industrial field.

4. **Conclusion.** In this paper, based on the SqueezeNet, we propose Slight-SqueezeNet-IGN, a novel smaller model. First, the Slight-SqueezeNet is proposed by improving SqueezeNet. Then, IN (Instance Normalization) and GN (Group Normalization) are concurrently designed to form an IG-module, which is merged into Slight-SqueezeNet to form Slight-SqueezeNet-IGN. Experiments on the CIFAR10 dataset show that the accuracy of Slight-SqueezeNet-IGN increases by 8.47% over the SqueezeNet, by 6.31% over the AlexNet-like model without significant increase in parameters and model complexity. Furthermore, the parameters of Slight-SqueezeNet-IGN are only 0.74 M, and are far less than VGG16-like and AlexNet-like. And the accuracy of Slight-SqueezeNet-IGN can achieve 95.78% on the self-made recyclable garbage classification dataset with complex background. Experiments prove that the Slight-SqueezeNet-IGN prides itself in good recognition ability, generalization ability, and stronger practical value, and it can be better embedded in hardware devices to meet industrial applications.

In the future, we will pay close attention to the development of smaller models and further expand the experimental scale and application scenarios of waste classification. We will explore the continuous improvement of the model and determine the optimal experimental parameters.

REFERENCES

[1] L. Rabano, M. K. Cabatuan et al., Common garbage classification using MobileNet, *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, Baguio City, Philippines, 2018.

[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang et al., MobileNets: Efficient convolutional neural networks for mobile vision applications, *arXiv.org*, arXiv: 1704.04861, 2017.

[3] V. Ruiz, N. Sánchez et al., Automatic image-based waste classification, *International Work-Conference on the Interplay between Natural and Artificial Computation*, pp.422-431, 2019.

[4] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp.770-778, 2015.

[5] R. A. Aral, Ş. R. Keskin, M. Kaya and M. Hacıömeroğlu, Classification of TrashNet dataset based on deep learning models, *2018 IEEE International Conference on Big Data*, Seattle, WA, USA, pp.2058-2062, 2018.

[6] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp.4700-4708, 2016.

[7] U. Ozkaya and L. Seyfi, Fine-tuning models comparisons on garbage classification for recyclability, *arXiv.org*, arXiv: 1908.04393, 2019.

[8] C. Szegedy, W. Liu et al., Going deeper with convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015.

[9] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning*, pp.273-297, 1995.

[10] F. N. Iandola, S. Han, M. W. Moskewicz et al., SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and <0.5 MB model size, *arXiv.org*, arXiv: 1602.07360, 2016.

[11] D. Jia, D. Wei, R. Socher, L. Li, L. Kai and F. Li, ImageNet: A large-scale hierarchical image database, *Advances in Neural Information Processing Systems*, pp.248-255, 2009.

[12] A. Krizhevsky, I. Sutskever and G. E. Hinton, Image network classification with deep convolution neural network, *Advances in Neural Information Processing Systems*, pp.1097-1105, 2012.

[13] D. Ulyanov, A. Vedaldi and V. Lempitsky, Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis, *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp.4105-4113, 2017.

[14] D. Ulyanov, A. Vedaldi and V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, *arXiv.org*, arXiv: 1607.08022, 2016.

[15] Y. Wu and K. He, Group normalization, *International Journal of Computer Vision*, pp.742-755, 2020.

[16] M. Yang and G. Thung, *Classification of Trash for Recyclability Status*, Corpus ID: 27517432, 2016.

[17] Y. T. Hafiyan, Afiahayati, R. D. Yanuaryska, E. Anarossi, V. M. Sutanto, J. Triyanto and Y. Sakakibara, A hybrid convolutional neural network-extreme learning machine with augmented dataset for DNA damage classification using comet assay from buccal mucosa sample, *International Journal of Innovative Computing, Information and Control*, vol.17, no.4, pp.1191-1201, 2021.