RecycleTrashNet: Strengthening Training Efficiency for Trash Classification via Composite Pooling

He-qun YANG, Ting LU, Wen-jing GUO, Shan CHANG and Jia-fei SONG
School of Computer Science and Technology, Donghua University, Shanghai 200000, China
*Corresponding author

Keywords: Deep residual network, Composite pooling, Trash classification, Residual block.

Abstract. In this paper, we propose RecycleTrashNet to classify house trash based on deep neural networks. In our model, we use 3x3 filter in convolution layer instead of 7x7 filter, which is a smaller filter tends to learn more features. Since single max pooling or average pooling in pooling layer can’t achieve good performance, we present composite pooling to preserve as many image features as possible. Experiments on trash dataset from Stanford University demonstrate good performance of RecycleTrashNet over other neural network models in speed and accuracy. It can improve training efficiency, which achieves classification results with 88% test accuracy at only 80 epochs.

Introduction

With global industrialization and urbanization, it is beneficial for environment protection and sustainable development to classify trash. Traditional methods of trash classification are manual classification, which are inefficient and high cost. In order to classify trash efficiently, a new method in view of neural networks is presented in this paper.

As artificial intelligence is widely used in our life, it has become a consensus to use deep neural network to classify trash, e.g., AlexNet [1], InceptionNet [2] and deep residual network [3]. AlexNet was proposed in 2012 and achieved remarkable results in ImageNet image classification competition. InceptionNet is also known as GoogleNet, which increased the depth and width in the network and optimized the quality of multi-scale processing [2]. Deep residual network got a first in the 2015 ImageNet image classification competition, which showed its superior performance [3]. Although these networks perform well in image classification, the network architecture is complex and training takes a lot of time. In most cases, trash classification is in real-time scenes, which needs to get the classification results as fast as possible. Existing trash classification methods usually need about 150 epochs of training to achieve the accuracy rate of about 85% [4, 5, 6], which is time consuming and slow. Therefore, improving the training speed and efficiency is necessary.

We introduce RecycleTrashNet, experimental results and analyze it in turn. Finally, a conclusion is given in the paper.

RecycleTrashNet

In this part, we outline the network architecture of RecycleTrashNet based on deep residual network firstly. Then, we use the small filters and residual learning to construct RecycleTrashNet. Finally, we introduce the composite pooling, our new pooling method out of which to build network architectures.

Network Architecture

Fig. 1 gives the basic architecture of RecycleTrashNet. It includes convolution layer, pooling layer and full connection layer. The convolution layer extracts feature from input image. Pooling layer is used for dimensional reduction and full connection layer gets the final prediction.
So as to decrease the number of parameters and modify computing speed, 3×3 filters, which have better discrimination and feature extraction ability, are used in convolution layer instead of the traditional large convolution filters (i.e., 7×7 or 11×11) for classification.

**Residual Learning**

With the increasing depth of layers, the problem of gradient disappearance may arise. Deep residual network presents a residual learning module such as shortcut to solve the problems of networks training [3]. Shortcut learn changed part that is known as residual, instead of learning entire output. Learning residual directly shows deeper networks are easier to optimize and more effective when the network depth is increasing.

As an important part of the convolution layer, shortcut makes cross-layer to avoid learning the same features. If we can learn the changed parts $\Gamma(x)$ from input $x$ to output $N(x)$, then it is equivalent to learn entire output, because the same part is no need to learn. After the convolution operation, the input and output dimensions may be inconsistent, we need to add $K$ for dimension matching. We expectantly make these layers satisfied a residual function

$$N(x) = \Gamma(x, \{K_i\}) + K_x x.$$  \hfill (1)

When several layers don’t learn new features, using the shortcut method to make a cross-layer connection to skip the same part. The purpose of the shortcut is to also solve the problem of repetition learning and training inefficiency.

**Composite Pooling**

At present, the main methods of pooling is single pooling, e.g., max pooling and average pooling. However, Both have their some disadvantages. Max pooling abandons all the non-maximum activation values in pooling area, which leads to the problem of serious information loss. Average pooling makes the extracted features less obvious. The two pooling methods are combined in this article, making them complementing each other, and complementing advantages, obviously improved the efficiency of classification.

In this paper, pooling area is viewed as a grid area. The row and column of the grid area are denoted $i$ and $j$ (i=1,2,…, j=1,2,…), respectively. Each grid has a weight $\lambda$, which is corresponding to the characteristics of image feature. Let $R_\rho$ denote the set of all grid locations. The composite pooling $F(x)$ is defined as follows:

$$F(x) = \alpha \times g(x) + \beta \times h(x),$$  \hfill (2)

in which $\alpha$ and $\beta$ are weight factors, $g(x)$ is max pooling, and $h(x)$ is average pooling.
\[ g(x) = \max_{i,j \in R_p} \lambda_{i,j}, \] in which \( \lambda_{i,j} \) is the weight of \( i \)th row and \( j \)th column. \[ h(x) = \frac{1}{N_{R_p}} \sum_{i,j \in R_p} \lambda_{i,j}, \] in which \( N_{R_p} \) is the total number of grids in the grid area.

Composite learning is shown in Fig. 3. First, composite pooling assigns pre-training weights (\( \alpha \) and \( \beta \)) to both ways. The value range of \( \alpha \) and \( \beta \) is 0-1, and the initial value is 0.5. According to classification accuracy of each round of training, evaluate the contribution of maximum pooling and average pooling to result respectively. Increase or decrease the max or average pooling weights by 0.01, and then pooling them again until achieve the best results.

![Composite Pooling](image)

**Figure 3. Composite Pooling.**

**Experiments**

We introduced the datasets used in this article. We then show the compare results to a number of popular models and traditional pooling methods. We implement the proposed architecture of RecycleTrashNet. All of our experiments are evaluated on a Ubuntu16.04 LTS machine with 32G RAM. The experiments were performed using NVIDIA GeForce GTX 1080Ti GPU.

**Dataset**

The dataset used in the paper is from Stanford University students [7]. We provide the image classification results achieved by the described RecycleTrashNet architectures on the trash dataset.

![Sample Trash Images](image)

**Figure 4. Sample Trash Images.**

This dataset contains some of the most common trash in our daily lives. There are 6 categories, a total of 2,527 photos in jpg format. Dataset is divided into training sets (2026 images) and testing sets (506 images). The classification performance is mainly evaluated by using test accuracy. Considering the impact of sunlight or room lighting, the background was a white poster board. The original size of each image is about 1.5M, most of the size is about 8032 x 3024 pixels. The raw data size is about 3.5GB.

In the training, so as to ensure that the model can learn more abundant and diverse input images, we used data augmentation such as: random slicing, random flipping, random adjustment of image brightness, color and other operations to ensure that the model has better generalization performance. Fig. 4 shows sample trash images: cardboard, glass, metal, paper, plastic, and other trash.
Comparison with Other Neural Networks

We conducted a comparative analysis of classification in the same trash set. In the context of our proposed network model, RecycleNet [4], Xception [5], DenseNet201 [5], Inception-V4 [6], DenseNet169 [6], and MobileNet [6] were trained by a great many epochs to achieve a relatively high accuracy rate. In Table 1, efficiency evaluation index provides the relationship between classification accuracy and training epochs. The higher efficiency evaluation index, the higher the training efficiency of the model can be achieved, which can reduce the training time and improve the efficiency of the model.

\[
\text{Efficiency Evaluation Index} = \frac{1}{\text{Accuracy} \times \text{Epochs}} \times 100 \tag{3}
\]

| Model                | Epochs | Data Aug. | Test Accuracy | Efficiency Evaluation Index |
|----------------------|--------|-----------|---------------|-----------------------------|
| RecycleNet [4]       | 200    | +         | 81%           | 0.62                        |
| Xception [5]         | 100    | +         | 85%           | 1.18                        |
| DenseNet201 [5]      | 200    | -         | 85%           | 0.59                        |
| Inception-V4 [6]     | 100    | +         | 89%           | 1.12                        |
| DenseNet169 [6]      | 150    | +         | 84%           | 0.79                        |
| MobileNet [6]        | 150    | +         | 84%           | 0.79                        |
| RecycleTrashNet      | 80     | +         | 88%           | 1.42                        |

A symbol + indicates the data augmentation.

Comparison with Traditional Pooling

The experiment shows that results of three pooling methods work. The new pooling method proposed in this paper adopts both small convolution filter and composite pooling, while the traditional pooling method does not use these two strategies. We compare them with each other and train in different epochs. We can see from Fig. 5 that in the training set, composite pooling showed good performance. Loss converges very quickly in the early stage. When the value of epochs is 80, the accuracy rate has reached a relatively high level.

Figure 5. Train accuracy (%) and loss of composite pooling results on trash classification.

The aim of this paper is to improve the running speed while ensuring the accuracy. If we can learn more features in early training stage, we could get more features, which will reduce training time. It can be found that with the increase of training epochs, the accuracy of composite pooling is getting higher and higher, and slightly higher than traditional pooling methods from Table 2.
### Table 2. Accuracy (%) of composite pooling results on the trash classification.

| Pooling          | Epochs | Loss  | Accuracy |
|------------------|--------|-------|----------|
| max pooling      | 10     | 0.56  | 84.25%   |
| avg pooling      | 10     | 0.56  | 83.46%   |
| Composite pooling| 10     | 0.53  | 85.24%   |
| max pooling      | 20     | 0.67  | 84.65%   |
| avg pooling      | 20     | 0.55  | 85.23%   |
| Composite pooling| 20     | 0.66  | 85.43%   |
| max pooling      | 80     | 0.79  | 87.01%   |
| avg pooling      | 80     | 0.67  | 86.81%   |
| Composite pooling| 80     | 0.71  | 88.19%   |

### Conclusion

We proposed a new network model architecture named RecycleTrashNet with composite pooling. It can extract more features in the early training phase, which can reduce training time and get the classification results faster. At the same time, the whole network uses small convolution filters, which reduces the network parameters and improves the speed of operation. We then compared different neural networks to our models demonstrating speed and accuracy characteristics. For future work, we will incorporate more complex and multiple object classification images.

### Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 61402101, 61672151), Shanghai Municipal Natural Science Foundation (Grant No. 18ZR1401200). (Corresponding author: Ting Lu).

### References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[2] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[4] Bircanoglu, C., Atay, M., Beseer, F., Genç, Ö., & Kizrak, M. A. (2018, July). RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks. In 2018 Innovations in Intelligent Systems and Applications (INISTA) (pp. 1-7). IEEE.

[5] Ozkaya, U., & Seyfi, L (2018). Fine-Tuning Models Comparisons on Trash Classification for Recyclability.
[6] Aral, R. A., Keskin, Ş. R., Kaya, M., & Hacıömeroğlu, M. (2018, December). Classification of TrashNet Dataset Based on Deep Learning Models. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 2058-2062). IEEE.

[7] Yang, M., & Thung, G. (2016). Classification of trash for recyclability status. CS229 Project Report, 2016.