X-DenseNet: Deep Learning for Garbage Classification Based on Visual Images

Sha Meng, Ning Zhang and Yunwen Ren
Dalian Maritime University, 1th Linghai Road, Dalian City, Liaoning Province, China
Email: mengsha0913@163.com

Abstract. In order to effectively solve the problem of garbage classification, this paper designs a garbage classification model based on deep convolutional neural network. Based on Xception network, combined with the idea of dense connections and multi-scale feature fusion in DenseNet, the X-DenseNet is constructed to classify the garbage images obtained by visual sensors. This paper conducts experiments through the process of "obtaining dataset-preprocessing data-building X-DenseNet model-training and testing model" and the accuracy of the model on the testing set is up to 94.1%, which exceeds some classic classification networks. The X-DenseNet automatic garbage classification model based on visual images proposed in this paper can effectively reduce manual investment and improve the garbage recovery rate. It has the vital scientific significance and application value.

1. Introduction
With the development of economy, the total amount of garbage in the world is increasing year by year. It is necessary to do garbage classification. As we all know, automatic classification which saves a lot of manpower is becoming a development trend of garbage classification. However, the traditional vision-based automatic garbage classification system uses simple manual features, and the generalization performance is unsatisfactory. There are still many problems in the automatic classification of complex and diverse garbage.

In recent years, deep learning has shown explosive development. Compared with traditional visual characteristics extraction methods, the advantage of deep learning is that it does not need to select in advance which features or design features to extract, but allows the model to learn from large-scale data. So deep learning has a stronger learning ability and adaptability.

This paper proposes a visual sensor-based method, according to deep learning, in order to achieve automatic garbage classification, which greatly improves the garbage collection rate.

2. Related Work
In the traditional method, Chan Xiang et al. [1] used Arduino UNO R3 as the main control board, used the color recognition module to realize the classification function. Yet, in this method, the characteristics of recognition are relatively simple, and the recognition object is relatively single. It is difficult to perform efficient automatic classification of complex garbage.

In the deep learning method, Ning Kai et al. [2] used YOLOv2 as the main network module, embedded deep dense modules to design a smart sweeping robot that can divide garbage into 25 subcategories including toilet paper, cans, milk, etc. according to shape and volume. Chen Yuchao et al. [3] proposed to use OpenCV to implement the background difference algorithm, cut the object to be detected from the picture, and then use the MobileNet to divide the image into syringes, hemostatic forceps, infusion bags, and gloves.
In summary, the garbage classification method based on deep learning has certain advantages. Therefore, deep learning is of certain practical significance and scientific research value to solve the problem of garbage classification and improve resource utilization.

3. Task Definition
This paper studies the problem of automatic classification of different types of garbage images in the case of using a color camera to obtain color images in the actual garbage classification scenario. This paper expects to generate the object category directly from the color image, and defines the function \( M \) as the mapping relationship between the input image \( I \) and the output category \( C \):

\[
M(I) = C \quad (1)
\]

This paper uses a deep convolutional neural network to approximate the complex function \( M \): \( I \rightarrow C \). Let \( M_\theta \) represent the neural network where \( \theta \) is the weight of the network.

This paper proves that \( M_\theta(I)=C=M(I) \), you can use the training set input \( I_T \) and the corresponding output \( C_T \) to learn and train the cross-entropy loss function \( L \), as follows:

\[
\min_\theta L(C_T,M_\theta(I_T)) \quad (2)
\]

The practical application process is as follows:

- First, use a color camera to obtain a two-dimensional color image \( I_0 = R^{H_0 \times W_0} \) with a height of \( H_0 \) and a width of \( W_0 \);
- The obtained original image \( I_0 \) is subjected to data preprocessing operations such as cropping and denoising to obtain the processed image \( I_1 = R^{H_1 \times W_1} \);
- Input the processed image \( I_1 \) to the trained deep learning model \( M_\theta \) to obtain the image category output \( C = M_\theta(I_1) \);
- Finally, according to the category \( C \) of garbage, the corresponding classification processing operation is performed.

4. X-DenseNet Network
First of all, the Xception [4] network structure is used as the basic structure. With higher accuracy and a relatively lower number of parameters, Xception is a further improvement of InceptionV3 [5]. The added ResNet [6] residual network mode also makes the speed of convergence of the network improve.

In order to make full use of image information features, this paper combines the idea of dense connection and multi-scale feature fusion in DenseNet [7] in the network structure. Among them, Dense Block is one of the main components of DenseNet. In a Dense Block, any layer is directly connected to all subsequent layers. Therefore, the layer 1 receives the feature information of all previous layers \( (x_s, \ldots, x_{i−1}) \) as layer 1 input:

\[
x_i = H_i([x_s, x_1, \ldots, x_{i−1}]) \quad (3)
\]

Among them, \([x_s, x_1, \ldots, x_{i−1}]\) refers to the connection of the feature map generated in the layer 0 to the layer \( 1−1 \).

Dense Block does not need to re-learn redundant parameter feature maps. In addition, each layer can directly access the gradient from the loss function to the original input signal, which can effectively slow down the problem of gradient disappearance. The structure of Dense Block adopted in this paper is shown in Figure 1.
Figure 1. The Structure of Dense Block

In order to merge multi-scale information and improve model performance, this paper on the basis of Xception uses Dense Block to realize feature reuse and fusion. The X-DenseNet shown in Figure 2 is proposed. The specific model structure is as follows:

- Take the garbage image (150×150×3) as the input, first extract the features through the Xception, and then transpose the convolution through 32 convolution kernels (3×3) and 64 convolution kernels (5×5) to increase the size and number of features, and then obtain feature images \( x_0 \). The \( x_0 \) is used as the input \( H_1 \) of the lower layer Dense Block.
- In Dense Block, \( H_1 \) undergoes batchnormalization and ReLu activation, and uses 32 convolution kernels (1×1) and 32 convolution kernels (3×3) for convolution, and finally obtains feature images \( x_1 \). After splicing \( x_0 \) and \( x_1 \), it is used as the input of the next layer \( H_2 \) in Dense Block, and so on. Between Dense Blocks, we use a convolution (1×1) and a maxpooling (2×2) to connect Dense Blocks.
- The output of the third Dense Block is alternately transposed and convolved with 3×3 and 5×5 convolution kernels to expand the number of features. The final output passes through the fully connected layer and finally maps to six neurons and then six types of garbage are classified.

Figure 2. The Structure of X-DenseNet

5. Experiment Preparation

5.1. Experimental Hardware Conditions

In this paper, a comparative experiment including X-DenseNet and AlexNet [8], ResNet50, InceptionV3, Vgg16 [9], and Vgg19 [9] was performed on a remote server. Hardware devices that the entire experimental process performed are shown in Table 1.

| Version                  |           |
|--------------------------|-----------|
| Operating System         | Ubuntu 16.04.6 LTS 4.15.0-91-gneric GNU/Linux |
| CPU                      | Intel® Xeon(R) CPU E5-2670 0 |
| GPU                      | GeForceGTX1080Ti |
| PyCharm                  | PyCharm 5.0.3 |
| Deep Learning Framework  | Keras 2.3.1 |
5.2. Dataset Generation
There are a total of 2527 color images in the format of .jpg in the garbage image dataset used in this paper, which are divided into 6 categories of garbage. There are 501 glass pictures, 594 paper pictures, 403 cardboard pictures, 482 plastic pictures, 410 metal pictures and 137 other trash pictures. The size of these pictures is adjusted to 512 * 384, and the camera equipment is Apple iPhone. In order to enhance the generalization performance of the model, the original image in the dataset is first subjected to data enhancement operations through shear rotation, flipping and other methods. Finally, 90% and 10% of the pictures in the dataset are used as the training set and testing set relatively.

6. Experimental Results

6.1. Training Results
This paper tested AlexNet, ResNet50, InceptionV3, Vgg16, Vgg19 and X-DenseNet in the experiment. Each batch in the training process is 32. The optimizer is SGD, the momentum is 0.9, the initial learning rate is 1e-3, and the loss function is the cross-entropy function.

The results of training accuracy and training loss are shown in Figure 3 and Figure 4. In terms of convergence speed, AlexNet training convergence speed is very slow, the training accuracy reached 65% in the 130th generation, and then stabilized at 70% afterwards. Vgg16 and Vgg19 have similar convergence rates, reaching 95% and 93% respectively at the 100th generation. After 100 generations, the training speed increases slowly. X-Dense and ResNet50 and InceptionV3 achieved 98% in training accuracy in the 25th generation. It can be seen from the training accuracy of the 125 ~ 150 generations that the training accuracy of X-Dense is almost always above 99%, and it is more stable than others.

![Figure 3. Training Accuracy](image1)

![Figure 4. Training Loss](image2)

6.2. Testing Results
248 pictures in the data set were used as the test set. After data enhancement processing, each batch of data in the test was 40 color garbage pictures. X-DenseNet is compared with AlexNet, ResNet50, InceptionV3, Vgg16, and Vgg19, then the classification accuracy is got, as shown in Table 2. X-DenseNet has the highest accuracy for the three types of garbage: cardboard, glass, and trash. Compared with other models, X-DenseNet has a higher accuracy for identifying each type of garbage. Overall, X-DenseNet's recognition efficiency is higher.

|            | Cardboard/% | Glass/% | Metal/% | Paper/% | Plastic/% | Trash/% | Average/% |
|------------|-------------|---------|---------|---------|-----------|---------|------------|
| AlexNet    | 47.50       | 52.00   | 60.98   | 86.44   | 41.67     | 38.46   | 62.10      |
| Vgg16      | 75.00       | 90.00   | 92.68   | 98.31   | 83.33     | 92.31   | 88.61      |
| Vgg19      | 82.50       | 86.00   | 95.12   | 96.61   | 97.92     | 84.62   | 90.46      |
| ResNet50   | 87.50       | 88.00   | 97.56   | 98.31   | 81.25     | 100.00  | 92.10      |
| InceptionV3| 85.00       | 90.00   | 92.68   | 96.61   | 79.17     | 100.00  | 90.58      |
| X-DenseNet | 95.00       | 90.00   | 95.12   | 94.92   | 89.58     | 100.00  | 94.10      |

6.3. Prediction Results
This paper takes 150 × 150 color garbage pictures as input, uses the X-DenseNet to classify and predict the pictures, and obtains the prediction result pictures, as shown in Figure 5. In the results, the
prediction and label are under each picture. The categories corresponding to each value in the list are cardboard, glass, metal, and paper, plastic, and other trash.

Figure 5. Prediction Results on X-DenseNet

7. Conclusion
Considering the problems of low efficiency of artificial garbage classification and poor generalization performance of traditional garbage classification algorithms, this paper designs a garbage classification model based on deep convolutional neural network X-DenseNet. The classification of domestic garbage images obtained by visual sensors can effectively classify garbage automatically, and divide the garbage into six categories: cardboard, glass, metal, paper, plastic and other garbage. And the recognition accuracy rate is higher than other advanced image classification networks.

8. References
[1] Zhan Xiang, Xue Peng, Guo Huanping. Intelligent sorting trash can based on Arduino [J]. Electronic World, 2020 (04): 160-161.
[2] Ning Kai, Zhang Dongbo, Yin Feng, et al. Rubbish detection and classification of intelligent sweeping robot based on visual perception [J]. Journal of Image and Graphics, 2019, 24 (8): 1358-1368.
[3] Chen Yuchao, Bian Xiaoxiao. Medical garbage classification system based on machine vision and deep learning [J]. Computer Programming Skills and Maintenance, 2019, (5): 108-110. DOI: 10.3969 / j.issn.1006-4052.2019. 05.042.
[4] Francois Chollet. Xception: Deep Learning with Depthwise Separable Convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567, 2015.

[6] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.

[7] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]// CVPR. IEEE Computer Society, 2017.

[8] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.

[9] Simonyan, Karen, Zisserman, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[10] garythung, “garythung/trashnet,” GitHub. [Online]. Available:https://github.com/garythung/trashnet.