Domestic garbage recognition and detection based on Faster R-CNN

Zhifeng Nie¹, Wenjie Duan¹*, Xiangdong Li²

¹College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao Shandong 266590, China
²College of Information and Control Engineering, Qingdao Technological University, Qingdao Shandong 266033, China

*Corresponding author’s e-mail: dduanwenjie@163.com

Abstract The core of intelligent garbage sorting is target identification and detection. In order to achieve effective garbage sorting, on the basis of deep learning, the Faster R-CNN target detection model and ResNet50 image classification model are used to identify and train 3984 garbage images, and predict 3552 images. The results show that the accuracy of garbage recognition is 89.681%, the average accuracy of each garbage prediction is 91.68%, and the accuracy of each category of garbage image prediction is over 93.3%. Through the identification, detection and classification prediction of garbage images, it provides data support for the intelligent classification of domestic garbage.

1. Preface

The emergence of new things has caused the diversity of domestic garbage. Taking China as an example, the total amount of domestic waste generated in a year is about 400 million tons[1]. In order to prevent waste of resources and environmental pollution, garbage classification is an inevitable trend of social development, and the classification accuracy of existing garbage classification methods is not ideal. Therefore, smart garbage classification methods play an important role in environmental protection and resource recovery.

Paul R.Wilson[2] analyzes and researching garbage collection through single-processor technology to improve the efficiency and locality of garbage collection. Huang, B [4] used NYU Depth V2 scene database to convert the single-channel depth information of convolution recognition into three-channels to improve the accuracy of garbage classification. Kashif Ahmad [3] uses the feature and score fusion method to optimally combine multiple deep learning models to solve the problem of automatic garbage classification.

According to the collected 14587 four-category garbage images, the Faster R-CNN target detection model and the ResNet50 classification model are used to identify and train 3984 randomly selected images, and 3552 randomly selected images are predicted to obtain higher recognition. The prediction accuracy is improved by adjusting the loss function to improve the robustness and generalization of the model, providing data support for the calculation of garbage identification and classification in the future.
2. Data analysis and model selection

2.1 Data collection
The garbage images are collected mainly through mobile phone camera records and network downloads. In the actual garbage classification work, it is not guaranteed that all the images recognized by the sorting system have high definition. In order to make the image training applicable to diverse occasions, the collection of different definitions, different sizes and JPG format. There are 14,587 garbage images, of which 23 types of recyclable garbage; 6 types of other garbage; 4 types of hazardous garbage; 8 types of kitchen waste.

Due to the wide variety of garbage types, there is a gap in the amount of garbage collected. The specific amount of each garbage collected is shown in Table 1.

| Recyclable Trash | Name          | Amount | Name        | Amount | Name          | Amount | Name      | Amount |
|------------------|---------------|--------|-------------|--------|---------------|--------|-----------|--------|
| Plastic box      | 213           | Clothing | 375        | Seasoning bottle | 434    | shoes      | 375    |
| Plastic bag      | 370           | Can     | 309        | Carton box | 318    | board      | 417    |
| Canvas bag       | 419           | Pillow  | 318        | Wine bottle | 280    | Plastic bottle | 215  |
| Toy              | 308           | Plug    | 594        | Milk can   | 322    | Pot        | 395    |
| Washbasin        | 347           | Doll    | 550        | Barrel     | 357    | Carton     | 225    |
| Hanger           | 301           | Cup     | 536        |           |        |            |        |

| Other Garbage    | Name          | Amount | Name        | Amount | Name          | Amount | Name        | Amount |
|------------------|---------------|--------|-------------|--------|---------------|--------|-------------|--------|
| Butts            | 277           | Sign   | 84          | Tiles  | 386           | Chopsticks | 288            |
| Shampoo bottle   | 350           | Cosmetic bottle | 353     |         |            |        |             |        |

| Hazardous Waste  | Name          | Amount | Name        | Amount | Name          | Amount | Name      | Amount |
|------------------|---------------|--------|-------------|--------|---------------|--------|-----------|--------|
| Portable charger | 355           | Battery | 321        | Ointment | 391    | Pill      | 435    |

| Kitchen Waste    | Name          | Amount | Name        | Amount | Name          | Amount | Name      | Amount |
|------------------|---------------|--------|-------------|--------|---------------|--------|-----------|--------|
| Leftovers        | 395           | Bone   | 361        | Fruit peel | 378   | Rotten fruit | 388    |
| Tea              | 386           | Vegetables | 728    | Egg skin   | 325    | Fish bone | 408    |

2.2 Faster R-CNN model
The Faster R-CNN model is a network model with fast calculation speed and real-time positioning and detection of objects. It is a combination of Fast R-CNN model and full convolutional network RPN to make the output rectangular area nomination and classification regression shared convolution feature. The Faster R-CNN model structure is shown in Figure 1.

![Figure 1 Faster R-CNN model structure](image)

2.3 ResNet50 model
The ResNet50 model is a major breakthrough in the development of CNN [5], and the accuracy of image classification can be as high as 96.6%. The ResNet50 model mainly introduces the residual
network structure, and the core of its operation is to establish a "Shortcut Connection" between the front layer and the back layer. The structure of the residual network is shown in Figure 2.

![Figure 2: Residual network structure](image)

The residual network unit can be expressed as:

\[ y_i = h(x_i) + F(x_i, w_i) \]  

(1)

\[ x_{i+1} = f(y_i) \]  

(2)

Learning features from the shallow layer to the deep layer of the neural network.

\[ X_L = x_i + \sum_{i=1}^{k} F(x_i, W_i) \]  

(3)

The gradient of back propagation in the image training process is:

\[ \frac{\partial \text{loss}}{\partial x_i} = \frac{\partial \text{loss}}{\partial x_i} \frac{\partial x_i}{\partial x_i} + \frac{\partial \text{loss}}{\partial x_i} (1 + \frac{\partial}{\partial \sum_{i=1}^{k} F(x_i, W_i)}) \]  

(4)

The ResNet50 neural network model solves the problem of inaccurate classification accuracy caused by the deepening of neural network layers. It can use deeper neural networks and fewer parameters to improve the accuracy of image classification and the speed of image training.

3. RPN loss function

3.1 RPN loss function

RPN is to add a conv+relu layer after the original feature map extracted from the image, which can directly generate the feature map of the local target and K anchor frames. After each anchor frame is connected a dichotomize softmax for detecting the probability of the object in the anchor frame as the detection target and a bbox regressor for adjusting the coordinate values of the four points of the anchor frame. After RPN training, the target detection result is finally output.

The loss function of RPN includes classification loss and regression loss, and its expression is:

\[ L(p_i, t_i) = \frac{1}{N_{\text{cls}}} \sum_{i=1}^{N_{\text{cls}}} L_{\text{cls}}(p_i, p_i^*) + \frac{\lambda}{N_{\text{reg}}} \sum_{i=1}^{N_{\text{reg}}} p_i^* L_{\text{reg}}(t_i, t_i^*) \]  

(5)

The classification loss is a typical cross-entropy loss. The log loss is calculated and summed for each anchor box first, and finally divided by the number of anchor boxes. The formula is:

\[ L_{\text{cls}}(p_i, p_i^*) = -\log[p_i^*p_i + (1-p_i^*)(1-p_i)] \]  

(6)

The dichotomous cross entropy loss function is:

\[ L_{\text{cls}} = -y\log y - (1-y)\log (1-y) = \begin{cases} 
-\log y^*, & y = 1 \\
-\log (1-y^*), & y = 0 
\end{cases} \]  

(7)

Offset of anchor box prediction in regression loss function

\[ t_i = \{t_x, t_y, t_w, t_h\} \]  

(8)

The actual offset of the anchor frame relative to gt is:

\[ L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*) \]  

(9)

R is the smoothL1 function. Let x=ti-ti*, then the function can be expressed as:
The loss function is one of the criteria for measuring the quality of the model prediction. In the actual image training, the classification loss and regression loss of the image detector need to be considered. As the number of training increases, the iterative curve of RPN network loss and Detector loss function is shown in Figure 3.

![Figure 3: Loss function iteration curve.](image)

Figure 3 Loss function iteration curve.

It can be concluded from Figure 3 that the overall trend of the accuracy of the loss function is decreasing, but the regression loss of the RPN network and the Detector network both have two large sudden changes, indicating that the model is instability during training and will affect the training accuracy.

### 3.2 Improve RPN loss function

It can be seen from Table 1 that the number of garbage images is unbalanced, resulting in poor robustness of the RPN network loss function during training, and the anchor size and positioning accuracy are not accurate enough. In order to improve the stability of the loss function and recognition accuracy of the RPN network model, the *focal* loss function is used to improve the classification loss function (2.3). The *focal* loss function model is:

\[
\text{smooth}_{\alpha} = \begin{cases} 
0.5x^2 \times \frac{1}{\sigma^2}, & |x| < \frac{1}{\sigma^2} \\
|x| - 0.5, & \text{otherwise}
\end{cases}
\]

(10)

![Equation](image)

The loss function is one of the criteria for measuring the quality of the model prediction. In the actual image training, the classification loss and regression loss of the image detector need to be considered. As the number of training increases, the iterative curve of RPN network loss and Detector loss function is shown in Figure 3.

3.2 Improve RPN loss function

It can be seen from Table 1 that the number of garbage images is unbalanced, resulting in poor robustness of the RPN network loss function during training, and the anchor size and positioning accuracy are not accurate enough. In order to improve the stability of the loss function and recognition accuracy of the RPN network model, the *focal* loss function is used to improve the classification loss function (2.3). The *focal* loss function model is:

\[
L_{\beta} = \begin{cases} 
-(1 - y') \log y', & y = 1 \\
-y' \log(1 - y'), & y = 0
\end{cases}
\]

(11)

Add a balance factor Alpha to balance the uneven proportion of garbage images.

\[
L_{\beta|\alpha} = \begin{cases} 
-\alpha(1 - y') \log y', & y = 1 \\
-(1 - \alpha)y' \log(1 - y'), & y = 0
\end{cases}
\]

(12)

Through the improvement of the loss function model, the iterative accuracy of the loss function will decrease steadily with the increase of the number of training. The iterative curve of the improved loss function is shown in Figure 4.

![Figure 4: Iterative curve of the improved loss function.](image)
4. Analysis of experimental and prediction results

4.1 Analysis of training results
We use the data set in VOC2007 format as the training and testing data set. The computer is configured with Intel(R) Core(TM) i7-10750H CPU 2.60 GHz, RAM 16GB, running environment is Python 3.6, tensorflow-gpu: 1.13.2. Under keras: 2.1.5, numpy: 1.17.4. Using Faster R-CNN target detection model and ResNet50 image classification model to conduct a round of training on the 3984 garbage pictures identified to obtain the initial accuracy of the garbage training is 84.138%. After 50 rounds of training on 3,984 garbage images through the target detection model, a data set is obtained. Although the number of training is relatively low, it contains a variety of garbage image data, which makes the model training results have a certain degree of robustness and universality. The accuracy analysis of the garbage recognition training process is shown in Figure 5.

![Figure 5 Accuracy change of 50 garbage recognition training.](image)

It can be seen from Figure 5 that the recognition accuracy can reach 89.61%, and the accuracy of the iterative process of recognition training is relatively stable, which can meet the classification and sorting of most garbage.

4.2 Training result prediction
After the training is completed, when there are multiple targets in the same image, the classification target in the image is automatically identified to determine the category of domestic waste and the probability of the category. Take leftovers in kitchen waste as an example, garbage identification the prediction results are shown in Figure 6.

![Figure 6 Schematic diagram of multi-target image recognition.](image)

Figure 6 shows that the classification prediction for all the leftovers in the identification frame is accurate. Due to the types of images collected are limited and the object types detected by the target have errors, but classification accuracy and recognition accuracy of garbage can be well.

In order to verify the accuracy of the model for predicting each type of garbage, 3552 sheets were randomly selected for classification accuracy evaluation. The calculation method of classification accuracy prediction: the ratio of the number of accurately identified garbage pictures in each type of garbage to the number of all garbage pictures identified in this type of garbage. The test of garbage classification accuracy is shown in Table 2.
Table 2 Test results of garbage classification accuracy.

| Recyclable Trash | Name          | Accuracy | Name          | Accuracy | Name          | Accuracy | Name          | Accuracy |
|------------------|---------------|----------|---------------|----------|---------------|----------|---------------|----------|
|                  | plastic box   | 0.903    | Clothing      | 0.922    | Seasoning bottle | 0.837 | shoes        | 0.945    |
|                  | Plastic bag   | 0.952    | Can           | 0.902    | Carton box     | 0.954 | board        | 0.946    |
|                  | canvas bag    | 0.897    | pillow        | 0.923    | Wine bottle    | 0.957 | plastic bottle | 0.886    |
|                  | toy           | 0.91     | plug          | 0.891    | Milk can       | 0.933 | Pot          | 0.94     |
|                  | washbasin     | 0.934    | doll          | 0.91     | barrel         | 0.968 | Carton        | 0.979    |
| Others Waste     | Butts         | 0.902    | sign          | 0.91     | Tiles          | 386    | chopsticks    | 0.989    |
| Hazardous Waste  | Shampoo bottle | 0.897   | Cosmetic bottle | 0.923 |                    |        |               |          |
| Kitchen Waste    | Portable charger | 0.958 | battery       | 0.937    | ointment      | 0.912 | pill          | 0.831    |

| Name  | Accuracy | Name  | Accuracy | Name  | Accuracy | Name  | Accuracy |
|-------|----------|-------|----------|-------|----------|-------|----------|
| Leftovers | 0.905 | bone | 0.905 | fruit peel | 0.833 | Rotten fruit | 0.913 |
| tea     | 0.924    | vegetables | 0.891 | Egg skin    | 0.917 | fish bone   | 0.9    |

Through classification prediction of 40 kinds of garbage images, it is shown that the classification prediction accuracy of garbage that is not easy to distinguish, such as fruit peel, is low, but the accuracy of classification prediction for most garbage images is more than 90%, which is greater than the training of target detection accuracy.

In the image prediction process, there are different characteristics in the same garbage image, such as identifying fruit skins as fruits or identifying fruit as fruit skins. These two types of garbage belong to kitchen waste. When the manipulator identification and sorting, sort these two types of garbage to the same garbage bin. In order to verify the accuracy of the classification prediction when the same category of garbage appears, according to the calculation method in Table 2. The recognition accuracy of the four categories of garbage is shown in Table 3.

Table 3 Recognition accuracy of four types of garbage.

| Name  | Recyclable Trash | Other Garbage | Hazardous Waste | Kitchen Waste |
|-------|------------------|---------------|-----------------|---------------|
| Amount | 1996             | 504           | 345             | 707           |
| Accuracy | 0.969          | 0.933         | 0.954           | 0.955         |

It can be learned from Table 3 that increasing the base of training pictures can improve the overall accuracy of garbage classification and recognition. Comparing Table 2 and Table 3, it can be concluded that when the training image data of each type is sufficient, the prediction accuracy of image classification is more stable, and the accuracy of garbage classification can reach a higher level, which will be provided data support for future garbage classification research.

5. Conclusion

By collecting a large number of garbage images, using the Faster R-CNN target detection model to locate and detect the images, the ResNet50 neural network performs feature extraction and classification of the target, and improves the RPN loss function to make the image training more stable, with a training accuracy of 89.61%, the prediction accuracy of the four categories of garbage classification is above 93.3%.

Recognition training is performed by mixing 40 types of garbage. Under the condition that the amount of each type of garbage is small, the accuracy obtained by 50 rounds of training is higher. This
method can improve the robustness and generalization of garbage recognition, and can be applied to garbage types complex and diverse occasions.

Acknowledgments
Thank you very much for the support of the National Natural Science Foundation of China, thank you Professor Nie for my guidance and help in the research process of this project, and thanks classmates for my support and suggestions in the research of the project, so that the research content is more rigorous and enriched.

References
[1] Sun, J.S., Xie, X.H., et al.(2019)Investigation of Community Domestic Waste Classification Behavior and Research on Optimization Strategy. Renewable resources and circular economy,12: 20-25.
[2] Paul R. Wilson. (1992) Uniprocessor Garbage Collection Techniques. Lecture Notes in Computer Science,637:1-42.
[3] Kashif, Ahmad., Khalil, Khan.,et al.(2020) Intelligent Fusion of Deep Features for Improved Waste Classification. IEEE Access,8:495-504.
[4] Huang, B., Lu, J.J., et al. (2016) Object recognition algorithm based on deep convolutional neural network, Computer application,12:333-340.
[5] Szegedy, C., Liu, W., Jia, Y., et al.(2014)Going deeper with convolutions. Conference on Computer Vision and Pattern Recognition,1-9.
[6] Dong, Z.Y. (2020)Design and Implementation of Waste Classification System Based on deep Learning.Chinese Academy of Sciences University,Shenyang.