Paradigm Academic Press
Innovation in Science and Technology
ISSN 2788-7030
AUG. 2022 VOL.1 NO.1

SSD-based Lightweight Recyclable Garbage Target Detection Algorithm

Jiahui Fang[1]

[1] College of Horticulture, Gansu Agriculture University, Lanzhou, Gansu,730000, China

Correspondence: Jiahui Fang, College of Horticulture, Gansu Agriculture University, Lanzhou, Gansu,730000, China.

doi: 10.56397/IST.2022.08.05

Abstract

Reducing garbage pollution to the environment will improve the quality of people's living environment and achieve a green cycle of life. In response to the existing traditional garbage recycling and classification, which requires a lot of wasted human and material resources, this paper proposes a garbage detection network with higher accuracy for garbage identification and localization by improving the SSD network. Firstly, the backbone network is improved to generate multiple scales of feature maps through the Resnet network as the basis for subsequent detection. In addition, the SE module is introduced to optimize the features needed for the model and weaken the environment's interference in the detection model. At the input image size of 320×320, the map of the improved model in terms of accuracy is 90.18%, which is an improvement of 1.82% compared to the original network. It can also meet the demand for real-time detection in terms of detection rate. Finally, compared with the common model, the improved model has better results in terms of detection accuracy and detection rate.

Keywords: garbage detection, attention mechanism, feature enhancement, target detection

1. Introduction

With the development of the economy and logistics industry, the amount of waste generated per capita is increasing dramatically. Different types of waste have different impacts on the environment. In recent years, waste separation has been gradually promoted in most regions, but it still needs some development time. For the environment, we need to make the garbage harmless and increase the recyclable rate. Therefore, rational recycling of waste is one of the best methods available. Continuing manual waste sorting is inefficient, and the proposed deep learning approach helps to achieve sorting by machine, improving efficiency and saving labour cost (Gu Xiangyu, 2022; Zheng Caiyun, Cao Danhua & Hu Cheng, 2022).

Target detection, based on convolutional neural networks to identify object categories in images and the regression of positions, is an important part of artificial intelligence and a method for intelligent robots to perceive the outside world. In recent years, with the strong development of computer hardware, the arithmetic power that the device can provide has increased significantly, and target detection algorithms based on deep learning have become mainstream. According to the implementation steps of the method are divided into two categories: two-stage algorithms (RCNN series) (REN, Shaoqing, et al., 2015), divided into two stages, first artificially generate the region of interest, in the detection of the target by convolutional neural networks. Two-stage target detection algorithms, because in the first stage are artificially filtered features, so the training process reduces a lot of negative sample information, so the second-stage algorithms usually have a higher detection accuracy, but the model has a higher complexity compared to the first-stage algorithms. One-stage algorithms, based on regression, return prediction results and anchor frame positioning directly on the input image. Therefore, the one-stage algorithm usually has lower model complexity, and the detection rate of the model can meet the real-time detection compared with the two-stage algorithm, while the global tuning of the model can be achieved. In recent years,

most studies have been based on implementing one-stage algorithms.

Garbage detection has certain specificities, firstly, the adverse weather environment has a great impact on the detection algorithm, and secondly, the data samples of garbage belong to the category of small samples. Both of these adverse conditions have a great impact on the accuracy of detection. In addition, it can lead to a higher false and missed detection rate of the model.

2. Related Algorithms

2.1 SSD Target Detector

As a single-stage target detection algorithm, the SSD network has a simple structure usually with a fast detection rate. As a regression-based target detection algorithm, it is an end-to-end target detection algorithm that identifies objects on the original image and position regression in a single computation. SSD adapts to multi-scale garbage detection by performing convolution operations on the input image to derive multiple scales of feature maps separately. Similar to the YOLO (REDMON, Joseph, et al., 2016) target detection algorithm, SSD (LIU, Wei, et al., 2016) maps the points on the feature map to the input image in an N × N square. Each point on the feature map is used to detect the presence of a target in the corresponding square, and the multi-scale detection design allows for better detection of targets at different scales.

2.2 Lightweight Feature Extraction Network

Given realizing real-time garbage detection, Resnet34 is used as the feature extractor for the feature extraction network, compared with the VGG (SIMONYAN, Karen & ZISSERMAN, Andrew, 2014) network. Resnet34 solves the problem of gradient disappearance caused by the depth of the network, and the learning ability of the network is greatly improved due to the design of the residual module. In addition to the initial convolutional pooling layer and the final pooling full connection layer, most of the network belongs to the stacking of modules, and these stacked parts have a shortcut, which can realize the reuse of features and make the richness of the feature map increase (HE, Kaiming, et al., 2016). As shown in Figure 1.

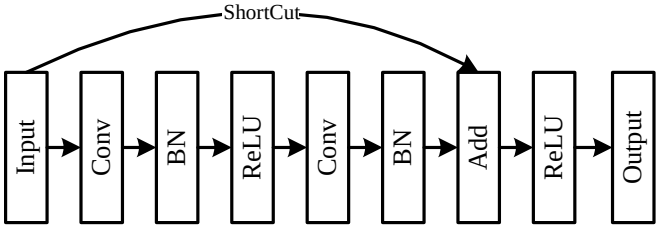

Figure 1. Residuals module

2.3 Improved Network Structure

As shown in Figure 2 below, the SE module first passes through the residual module and forms a one-dimensional feature map by global pooling, then compresses the channels a certain number of times through the fully connected layer, and returns to the original number of channels after ReLU, and finally derives the weighted feature information through the activation function. Suppose the input is a feature map of h * w * c, firstly, a global average pooling is performed, and a feature map of 1 * 1 * c is obtained by the global pooling (pooling size is h * w), then there are two fully connected layers, the number of neurons in the first fully connected layer is c/16, which is a The first fully connected layer has c/16 neurons, which is a dimensionality reduction method, and the second fully connected layer is dimensioned up to C neurons, which has the advantage of adding more nonlinear processing to fit the complex correlations between channels. Then a sigmoid layer is added to obtain a feature map of 1 * 1 * c. Finally, it is a full multiplication of the original h * w * c and 1 * 1 * c feature maps. The reason why it is a full multiplication instead of matrix multiplication is that it can get a feature map with different importance for different channels, and the above feature information is fully multiplied with the input feature X to get X', which is a feature map with enhanced weights and is integrated with certain global information and can filter certain interference information. This enables the model to focus more on global feature learning (HU Jie, SHEN Li & SUN Gang, 2018).

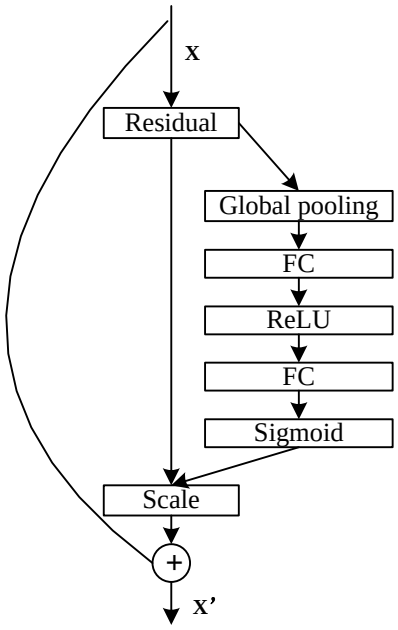

Figure 2. SE module

Figure 3 below shows the skeleton diagram of the network as a whole. The feature extraction of the original image is first performed by Resnet34 to generate multiple scales of feature maps. The attention module is introduced during this period, and its purpose is to solve the loss problem caused by the different importance occupied by different channels of the feature map in the convolutional pooling process. In the traditional convolutional pooling process, each feature channel is equally important, but in the actual problem, different channels have different importance, so the SE module is introduced to realize the weighting of different channels. In garbage detection, there are many interferences in the model training, making the model unable to focus on the learning of target features. The introduction of the SE module forces the channels to be scaled and adds the extra learned features to the original feature map, making the feature map richness enhanced, which is helpful to filter the interference factors and improve the detection accuracy of the model.

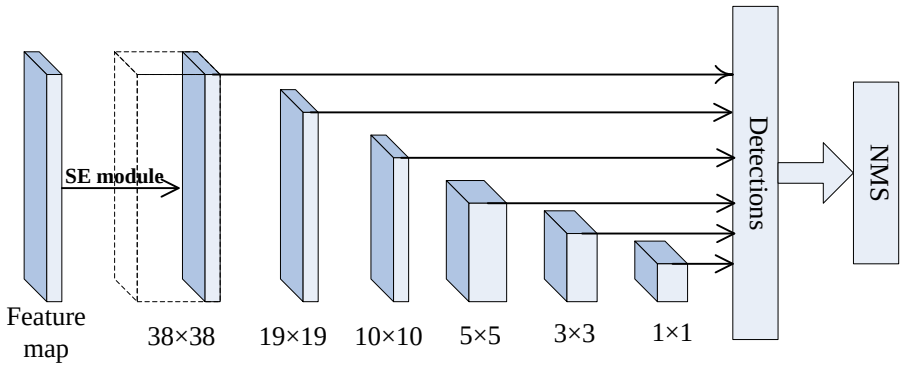

Figure 3. Network overall structure diagram

3. Experimental Dataset and Experimental Results

3.1 Dataset

The dataset uses an open source garbage detection dataset, which contains five categories, respectively, paper, cup, citrus, bottle and battery; as the detection dataset, the detection objects have been labelled, and these five categories are common garbage, both recyclable and non-recyclable. Theoretically, the more the number of training sets is beneficial to the training of the model, and the data can be augmented by data augmentation. For example, rotation transformation, mirror transformation, etc.

3.2 Experimental Environment and Parameter Settings

The anchor-based target detection algorithm, usually in the training will be derived from the prediction frame, after the IoU(As shown in equation 1) function prediction frame and the real frame will compare and calculate the error, will backpropagate the loss, so that the data set used more fit the real data, until approximating the optimal solution.

In the training process of the model, the learning rate is set with respect to the learning density of the model on the data, too large or too small will have an impact on the final results of the model. On the other hand, the performance of the training device used also affects the learning rate setting, and a larger learning rate can be set when the memory of the graphics card is relatively large. The learning rate is set to 0.001 and the Batch_size is set to 24. 150 epochs are iterated on the data set, and the optimizer uses Adam's algorithm, and the learning rate is decayed in the middle and later stages of training to make the training of the model more refined and to make it reach the optimal solution.

3.3 Experimental Results and Evaluation Metrics

As the evaluation metrics of deep learning, precision (as shown in equation 2), recall (as shown in equation 3), the average accuracy (Average Precision, AP) (as shown in equation 4) and the accuracy of each category (as shown in equation 5) are usually used as indicators to judge the performance of the model. In addition, as a real-time target detector, the value of FPS is also an indicator to evaluate the inference rate of the detector.

$$\mathrm{IOU} = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A is the prediction box and B is the true box. The cross-merge ratio is obtained as an estimate of the model error.

$$Precision(P) = \frac{TP}{TP + FP} \tag{2}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{3}$$

$$AP = \int_0^1 P dR \tag{4}$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \tag{5}$$

Where TP is true positive, FP false positive, FN false negative, and TN true negative, AP is the precision of one category, and mAP is the average precision of multiple categories.

Table 1. Multiple model comparison table

Model	Size(M)	mAP(%)	FPS
Faster-RCNN	195.0	89.79	2.5
YOLO v3	237.0	86.41	36.1
SSD	100.0	88.36	34.8
Our	33.5	90.18	32.4

4. Detection Test Results

Figure 4. Test results

As shown in the figure below, the improved SSD detector has a high confidence level for the determination of small-sized targets, and there is no leakage or false detection, indicating that the introduction of the new feature extraction network provides an excellent feature map for the subsequent detection stage, and in addition, the introduction of the SE module makes the noise of the model's environment weakened, which can ensure that the model can focus more on what it should learn, and then for the overall network, the fit to the dataset will become better and better.

5. Summary

For the problem of poor detection of garbage by SSD network and the problem of low inference rate of detection network, the original network is improved. First, for the feature extraction network, the original VGG network has the phenomenon that the gradient disappears too deep in the network, so it will affect the accuracy of the model, this paper uses the Resnet network to achieve feature extraction, and the residual module can effectively reduce the problem of gradient disappearance caused by too deep in the network layers. For garbage detection, the noise of the environment has a great impact on the detection of the model, through the channel attention mechanism to filter the channel, the algorithm will be assigned relatively lower weights to the channels with lower weights, then the overall perceptiveness of the model to the object to be detected will be enhanced, so the detection accuracy of the model will be improved. Finally, the use of some data augmentation methods to supplement the data set and the use of some novel training means can also improve the accuracy of the model, making the improved model more beneficial for real-time garbage detection.

References

Gu Xiangyu, Zhu Lijia, Liu Hu Nan, Li Guilin, Bu Wenping, Liu Guihua. (2022). Water surface litter saliency detection based on space-time domain information fusion. *Electronic Measurement Technology, 45*(11), pp.154-160.

Zheng Caiyun, Cao Danhua, Hu Cheng. (2022). A similarity-guided segmentation model for garbage detection under road scene. *Frontiers of Optoelectronics, 15*(1).

REN, Shaoqing, et al. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 28*.

REDMON, Joseph, et al. (2016). *You only look once: Unified, real-time object detection.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788.

LIU, Wei, et al. (2016). *Ssd: Single shot multibox detector*. European conference on computer vision. Springer, Cham, p. 21-37.

SIMONYAN, Karen & ZISSERMAN, Andrew. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556.

HE, Kaiming, et al. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

HU Jie, SHEN Li, SUN Gang. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition, pp. 7132-7141.