Waste Classification Based On Yolov4

Yongchuang Yang*

Department of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, Henan, 454003, China
*Corresponding author: Yongchuang Yang

Abstract: The importance of recycling to the environment and economy is well known. The existing manual labor and traditional industrial sorting technology have low efficiency, poor environment and easy to make mistakes. In recent years, there are many researches on waste recycling, but there is less recognition of recycling in restaurant environment. Therefore, this paper proposes a method based on deep learning to classify restaurant recycling. The experimental platform based on yolov4, SSD and fast RCNN is built. The experimental results show that yolov4 has the highest detection effect, with an accuracy of 77.78% and a detection speed of 39.3 FPS.

Keywords: recyclable items; object detection; image recognition; yolov4

1. Introduction

With the development of the times, a large amount of garbage will be produced in production and life. The industrial revolution has led to the rapid development of science and technology and the subsequent garbage is also growing exponentially. The resulting waste disposal problem has also become a thorny problem that people must face and solve. It is imperative to reasonably dispose of waste, realize resource reuse and reduce waste pollution People focus on detecting and identifying the types of garbage in daily life.

The United Nations issues *Transforming our World: The 2030 Agenda for Sustainable Development*[1] want to ensure environmental sustainability. This requires us to strengthen the utilization of recyclable resources in our daily life, and use science and technology to reduce human duplication of labor. Compared with the traditional recycling methods, the intelligent object recognition software can better deal with the large number of objects that are identified with efficiency.

Deep learning methods are being used in different fields with its high efficiency and accuracy, such as autonomous driving and defect detection with remarkable results on object detection. We can use the advantages of deep learning in object detection and apply it to the detection of objects on restaurant trays. Applying these methods to waste sorting can increase type and quantity of recycled material.

In this work, we solve two problems in the traditional recycling algorithm, the problems as follows:

1)The traditional methods of waste recycling are based on the materials of articles, which can be roughly divided into recyclable, unrecyclable and harmful. Traditional algorithms ignoring the shape of the object.

2)The traditional methods of waste recycling, a picture contains only one recycled object, but in real life, a picture often contains multiple recycled items.

In order to solve these problems, this paper uses some new methods as follow:

1) A data set of recycled waste based on restaurant environment is created, and the recycling types are classified according to the shape and material of object.

2) Multi-target convolution neural network YOLO v4 is used to detect multiple targets in a single picture

An overview of research for recyclable articles is addressd in section 2. Section 3 describes the methods mentioned. Section 4 explains the process and results of the whole experiment. Section 5 summarizes the research work and the future research direction.

2. Related Work

In recent years, with the increasing awareness of environmental protection, the identification of recycled items by scientific and technological means has gradually developed.

Maher Arebey et al.[2] use gray level co-occurrence matrix (GLCM) method to identify the types of objects. The characteristics obtained from GLCM are processed by MLP and KNN methods to obtain the types of recyclable articles. Comparing the classification results of GL and KNN, it is found that KNN is better than MLP. Sakr et al.[3] use convolution neural network (CNN), the recognition accuracy can reach 83%. Sakr found that the accuracy can be achieved 94.8%by using SVM is better than CNN.

With the development of deep learning, the first classification network AlexNet appeared. Mindy Yang et al.[4] compares SVM with AlexNet[5], and optimizes SVM with SIFT. Experiments show that the optimized SVM algorithm is better than AlexNet, and its accuracy is 63%.

Lyu Dong, Wang Ping et al[6] proposed a metal waste classification based on deep learning. A data set of solid metal waste is created and named GX trashnet. Resnet-101 is used as the classification standard model, and the channel attention module[7] is used to improve the classification accuracy. Experimental results show that the classification accuracy of the improved model is 97.00%.

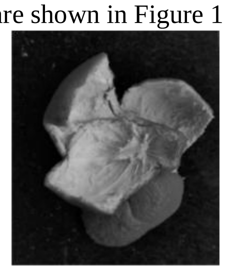
Yang et al. [8] proposed a lightweight model wasnet, and compared the classification performance of AlexNet, vgg-19 and inception-ResNet on the data set in the experiment. Among them, the classification accuracy of wasnet is the best, 96.10%; In addition, the author also embedded the trained model into the mobile terminal for testing. Wang Xiaoyan, Xie Wenhao et al.[9] proposed three classification and recognition methods based on yolov3, faster-rcnn and retinanet. Through the comprehensive analysis and comparison of the training process and test results, it is found that the faster-rcnn algorithm has better performance.

3. The Propose Method

3.1. Dataset Preprocessing

Although the existing literature is based on self-made data sets, there are still some problems, such as (a) there are no datasets available that have multiple types of objects; (b) The scenes captured by all data sets are in a controllable environment. Our datasets is composed of food tray images obtained from different places, such as shopping centers, canteens or homes, and does not control the lighting and type of shooting. The data sets in the literature and in this paper are shown in Figure 1:

(a)Oters_Plastic



(b) Oters_Paper

(c)Ours_Complex

Figure 1. Comparison between literature and this dataset

In this dataset, we use the shape and material of each object to classify objects. Materials can be divided into glass, paper, metal and plastic. According to the shape classification, it can be divided into boxes, plates, trays, waste, cutlery, bottles, can, cups, paper and plastics.

3.2. Yolov4 Algorithm

Yolov4 algorithm, as the fourth version of Yolo series of object detection algorithms, is optimized based on the original yolov3 target detection network architecture, The main structure of yolov4 can be divided into four part: backbone feature extraction network(CSPDarknet53), spatial pyramid pooling structure (SPP), path aggregation network (PaNet), Network detection output(Yolo head). The specific network structure is shown in Figure 2.

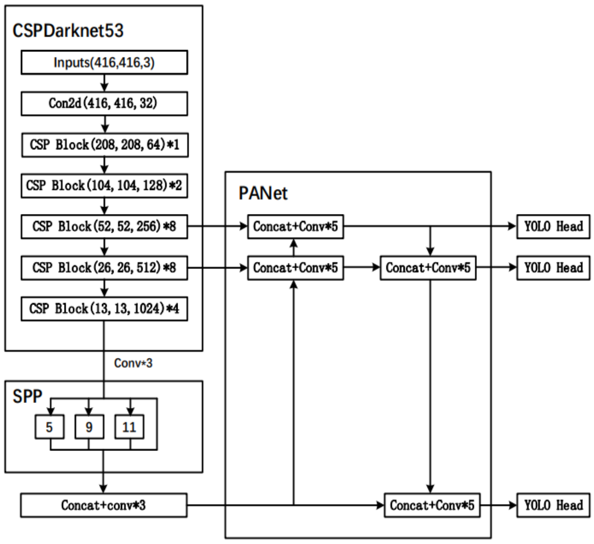


Figure 2. Yolov4 network structure

Yolov4 replaced the backbone feature extraction network of yolov3 with CSPDarknet53. As shown in Figure 2, after the input image passes through the backbone network, yolov4 will output feature maps of 13*13, 26*26, 52*52, and feature maps of different scales contain semantic information of different dimensions. In the feature fusion part, the 13*13 size feature map will enter the SPP structure, and the SPP will stack and convolute the obtained new feature map and the feature map before entering the network and output it to the feature fusion network PaNet. PaNet up samples the feature maps of 13*13 and convolutes the results with the feature maps of 26*26 and 52*52 respectively, then performs a series of similar down sampling and stacking convolution from bottom to top, fully integrates the features of three different scale feature maps, and finally outputs three Yolo heads of 13*13, 26*26, 52*52.

3.3. Transfer Learning

Transfer learning is a strategy of supervised learning for small samples. The optimized weight is obtained by pre training in similar large samples, and then the trained network model is embedded into other task models as the main feature extractor. As shown in Figure 3, the model is trained from the large sample source domain, and the knowledge learned from the source domain is frozen, trained, and the classification layer parameters are updated to achieve the effect of knowledge transfer. In this paper, multi-target recognition is carried out in complex environment, so the pre trained weights in VOC dataset are selected, and finally the selected model is fine tuned to achieve better recognition effect in this dataset.

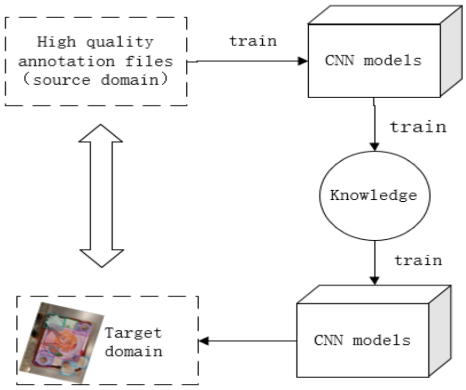


Figure 3. Transfer learning process

4. Experiments and Analysis

4.1. Experimental Environment

During this experiment, Using GPU to train the neural network can accelerate the convergence speed of the network. The configuration information of GPU hardware and programming language used in this experiment is shown in the Table 1.

Table 1. Experimental environment

Hardware\ Software	Parameter
CPU	Inter Xeon 2274
RAM	10G
GPU	NVIDIA GeForce GTX 3080
Python	3.7 version
Deep learning framework	Tensorflow 2.2/keras

4.2. Evaluation Criterion

The experiment compares the performance of the model by analyzing the precision rate, recall rate, average precision rate and average detection time before and after the model improvement in the test. Among them, precision(P) is a measure of accuracy, recall(R) is a measure of coverage, The average accuracy is a function of the accuracy as a function of the recall rate.

$$P = \frac{TP}{TP+FP} \quad R = \frac{TP}{TP+FN} \tag{1}$$

$$AP = \int_0^1 P\,(R)dR \quad mAP = \frac{AP}{N} \tag{2}$$

In the formula (1), TP is quantity with correct classification, FP quantity is of classification errors. FN is Classify non target and correct quantity. In the formula (2). P(r) is the function of precision rate with respect to recall rate; R is the recall rate.

4.3. Training Process

Label the current data set with Labelimg tool. The specific categories can be divided into tray, plastic plate, glass plate, ceramic plate, paper, napkin paper, mixed waste, mental cutter, plastic cup, paper cup, glass cup, metal can, plastic box, paper box, plastic bottle, glass bottle. Moreover, we define the yolov4 model to extract and recognize the features of 16 types in the dataset. The loss of the model reflects the degree of network convergence in the training process and the recognition ability of the network in the training process. Recording the change of model loss value can dynamically reflect the advantages and disadvantages of the model.

During the training, 8 pictures were taken as the size of one epoch. Using transfer learning, first train the VOC dataset as the source domain to obtain an initial weight information. Then the feature extraction network parameters of yolov4 are frozen and the classification layer of the network starts training. Although the feature extraction layer of transfer parameter weights can extract image features, there are great differences between ours data set and VOC dataset. Only training the classification layer can not achieve the ideal effect.

The experiment uses Adam optimizer, and the initialization learning rate is 0.001, The batch size is set to 8. When only the classification layer is trained in the transfer learning, the initialization learning rate is 0.001. When the classification layer training of 50 epochs is completed, the whole model is trained. At this time, the initialization learning rate is 0.0001. In order to prevent over fitting, early stop is used in the experiment. When the loss value of 10 epochs does not decrease, the training will be stopped automatically. During training, the iterative diagram of model loss is shown in Figure 4.

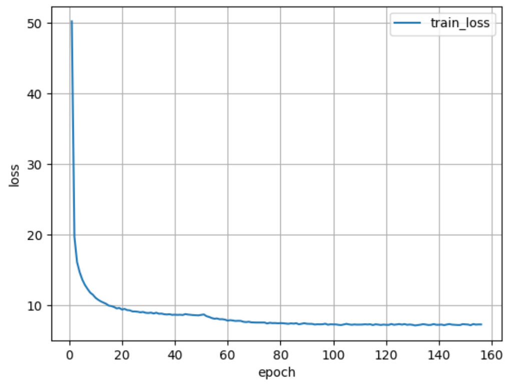


Figure 4. Loss curve

5. Results

We compare the performance of datasets on different algorithms, and compare the algorithm results with other object classification networks, such as literature 8. As shown in Table 2:

Table 2. Comparative study on recognition of recyclable datasets

<table>
<tr><th colspan="2">Models</th><th>Dataset</th><th>Types</th><th>mAP</th><th>FPS</th></tr>
<tr><td colspan="2">Yolov4</td><td>Ours</td><td>16</td><td>77.78%</td><td>39.3</td></tr>
<tr><td colspan="2">SSD</td><td>Ours</td><td>16</td><td>62.75%</td><td>45.4</td></tr>
<tr><td colspan="2">Faster-rcnn</td><td>Ours</td><td>16</td><td>74.1%</td><td>13.6</td></tr>
<tr><td rowspan="3">Paper 8</td><td>yolov3</td><td rowspan="3">Ga dataset</td><td>2</td><td>82.1%</td><td>10.6</td></tr>
<tr><td>RetinaNet</td><td>2</td><td>83.1%</td><td>8.1</td></tr>
<tr><td>Faster-rcnn</td><td>2</td><td>84.4%</td><td>6.5</td></tr>
</table>

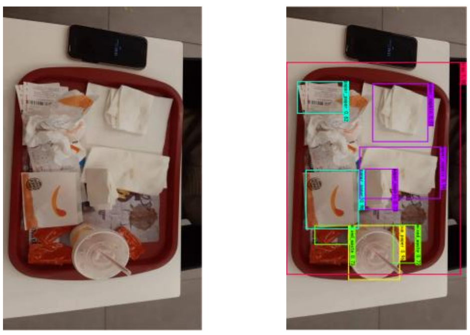
As shown in Table 1, we compared three different algorithms under the same conditions and the same dataset. Among the three algorithms, yolov4 performs better than the other two algorithms, the accuracy reaches the highest, and FPS is at the medium level. Compared with literature 8, our experimental accuracy is lower than that of garbage dataset, but the dataset in this paper is better in type and speed, and the dataset in this paper is more complex. Taking one training result as an example, the weight obtained is used to test the accuracy of the test set to obtain the accuracy of each type of items. These results are shown in Table 3.

Table 3. Experimental results of Yolov4 algorithm on test set

Model	Types	Material	Acc
Yolov4	bottle	glass	42.03%
		plastic	91.80%
	plate	plastic	89.93%
		glass	33.64%
		ceramic	98.24%
	paper	napkin	86.64%
		paper	58.52%
	cup	plastic	87.72%
		paper	94.73%
		glass	80.09%
	box	paper	82.79%
		plastic	80%
	tray		97.9%
	metal cutter		78.50%
	metal can		76.97%
	mixed waste		65.13%

As shown in Table 3, the accuracy of 16 categories under yolov4 algorithm can be seen that according to the shape of the object, the accuracy of larger object is higher than that of smaller object, but under the same shape, according to different materials, the accuracy of object with rich characteristics such as plastics and ceramics is higher than that of glass materials.

Taking the optimal experimental results as an example, some recognition results are selected from the test samples for prediction. The detection effect is shown in Figure 5 and the detection results are shown in Table 4.



(a)Image1_Original (b)Image1_ Classified

(c) Image2_Original (d) Image2_ Classified

Figure 5. Comparison of partial model test results

Table 4 Experimental prediction results

Image	Image1	Image2
Original label type	Cup_paper paper*2 paper_napkin*4 Tray mixed_waste*3	Cup_plastic Cup_paper Cutlery_plastic Plate_ceramic Paper_napkin Paper_napkin Tray mix_Waste
Classified prediction type	Cup_paper mixed*2 paper*2 tray paper_naokin*4	Cup_plastic Cup_paper Cutlery_plastic paper_napkin Paper_napkin Plate_ceramictray

6. Conclusion

Deep learning is an important field in machine vision. In traditional machine learning, sift and SVM algorithms are used to classify the types of recyclable items, but only one kind of items can be identified in a picture. Deep learning can use the rich feature information in an image to identify a variety of items in a picture. Therefore, this paper uses yolov4 algorithm to classify multiple types of items in the restaurant data set, and compares it with the other two algorithms to obtain good experimental results. The accuracy of the test set reaches 77.78%, which is 15% higher than SSD algorithm, and the detection speed is 20fps faster than fast RCNN algorithm, which proves that yolov4 algorithm has strong applicability in complex scenes. As future work, it is planned to add data sets in different scenarios, improve the variability of data sets, and improve the network to enhance its recognition ability in items with unclear features.

References

[1] Nations U. “About the sustainable development goals,”un.org.

[2] Arebey M.; Hannan M.; Begum R. A.; and Basri H. Solid waste bin level detection using gray level co-occurrence matrix feature extraction approach. *Journal of environmental management*, 2012; vol. 104, pp. 9–18.

[3] Sakr G. E.; Mokbel M.; Darwich A.; Khneisser M. N.; and Hadi A. Comparing deep learning and support vector machines for autonomous waste sorting, *in Multidisciplinary Conference on Engineering Technology (IMCET), IEEE International, IEEE*, 2016; pp. 207–212.

[4] Yang M. and Thung G. Classification of trash forrecyclability status, *CS229 Project Report*, 2016; vol. 2016.
[5] Krizhevsky A.; Sutskever I.; and Hinton G. E. Ima-genet classification with deep convolutional neural net-works. *Advances in neural information processing systems*, 2012; pp. 1097–1105.
[6] Lyu D.; Wang P. Research on a deep learning method for solid metal waste classification. *Journal of guangxi university of science and technology*, 2021; vol. 32, pp. 104–110.
[7] Hu J.; Shen L.; Sun G. Squeeze-and-excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020; vol. 42, pp. 2011–2023.
[8] Yang Z. H.; Li D. WasNet: a neural Network-based garbage collection management system. *IEEE Access*, 2020; 8: pp. 103984–103993.
[9] Wang X. Y.; Xie W. H.; Yang Y. F.; Hu R. Garbage classification and detection method based on deep learning. *Modern Electronics Technique*. 2021; vol. 44, pp. 110-113.