

Detection of Domestic Waste Based on YOLO

Yaohui Hou

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
houyh_0817@bupt.edu.cn

Abstract: Deep learning has increasingly permeated every aspect of our lives. Currently, the economy develops rapidly and the population living in cities grows quickly. In the meanwhile, the amount of domestic waste is constantly rising, and the serious shortage of waste treatment competence has become increasingly prominent. Many cities can only use landfill or incineration to generate electricity for most of the waste, but this is not a long-term solution. In this paper, the YOLO-based domestic waste classification and treatment method are proposed to improve the efficiency of waste treatment.

Keywords: YOLO; Domestic waste; Object detection; Deep learning.

1. Introduction

In the domain of artificial intelligence and pattern recognition, machine learning especially deep learning is an enduring research hotspot, and its theories and techniques have been widely used to settle complicated engineering applications and science problems. Among them, the YOLO algorithm is one of the representative algorithms. YOLO,

short for *You only look once*, as the name suggests, you just need to look once. We humans can know what objects in an image are, where they are, and how they relate to each other with a glance. In the first generation of YOLO, the input was resized to a specific size, then passed through a single convolutional neural network, and the detection results were thresholded by the confidence of the model. Figure 1 shows the working principle of a simple YOLO detection system.

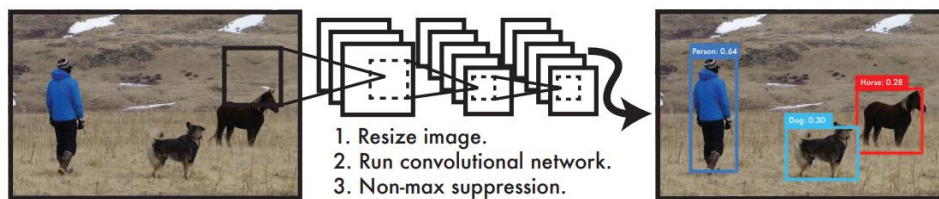


Figure 1. The YOLO Detection System

The second-generation YOLO algorithm improves some of the shortcomings of the first generation. The author used batch normalization, direct location prediction, multi-scale training and other methods to solve the problem of inaccurate localization and low recall. In addition to this, the training network is also modified. The author uses Darknet-19 as the base network, which reduces the amount of computation to some extent.

By comparison with the previous generation, the third-generation algorithm has mainly made two improvements. Firstly, drawing on the method of ResNet and adding some short connections, Darknet-19 is improved to Darknet-53. Secondly, using Feature Pyramid Networks for Object Detection enabled a better detection for small object. However, unfortunately, the detection performance for large objects degraded slightly.

The authors have done quite a bit of work to get YOLOv4 to high performance. The main work is reflected in the integration and parameter adjustment of the network model. YOLOv4 makes everybody can train a fast and accurate object detector by applying a common GPU, like 1080 Ti.

2. The reason for choosing YOLO

Prior to YOLO, algorithms such as R-CNN(Region-CNN) used the region proposal method which can generate potential bounding boxes on images. Each image would generate about 2000 category-independent potential bounding boxes, extract

a fixed-length feature vector from every proposal by using a convolutional network and then perform classification and detection on the bounding boxes. The box is then refined to remove reduplicate detections. The problem with this approach, however, is that the process is overly complex, slow and difficult to optimize. In contrast, the YOLO algorithm is much simpler, a simplex neural network can predict multiple bounding boxes at the same time, and it trains on the entire image and optimizes detection performance directly. Compared with traditional detection models, YOLO has a couple of advantages. First of all, it is really fast. The object detection is served as a simple regression problem, which means we do not need a complicated procedure. Secondly, YOLO can detect the image as a whole during prediction. It is different from the R-CNN series of algorithms, since it sees larger context and makes less than half of background errors. Finally, YOLO learns abstract features of images. When ported to a new field or unexpected input emerges, it is less likely to break down.

3. Experimental work

In this experiment, the writer uses the YOLOv5 algorithm, the Windows operating system, the GPU model MX250, the Pytorch framework, and the python programming language. The VOC dataset used in the experiment contains 44 classes common domestic wastes and more than 15,000 image, such as disposable snack box, washing product, stained plastic and

so on. The YOLOv5 source code is available at <https://github.com/ultralytics/yolov5>. In addition, the source code was also modified as follows:

(1) Create a `split_train_val.py` file in the dataset directory VOCData to divide the training dataset and validation dataset. The training dataset accounts for 90%.

(2) Create a `xml_to_yolo.py` file in the VOCData directory to convert name, width, height, and other information into `yolo_txt` format.

(3) Create a `my.yaml` file in the data folder of the source code and write the training dataset and validation dataset paths and the types and names of the waste.

(4) The number of classes(nc) is changed to 44 in the configuration file in the model's directory.

(5) In this experiment, the `yolov5s` model is used, so the model is supposed to be downloaded in advance and added to the directory.

4. Result analysis

The representation of an object detection model is mainly reflected by the precision ratio(P), recall ratio(R), and mean average precision (mAP). The calculation method is as follows:

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$recall = \frac{TP}{TP+FN} \quad (2)$$

$$P_{smooth}(r) = \max_{r' \geq r} P(r') \quad (3)$$

$$AP = \int_0^1 p(r)dr \quad (4)$$

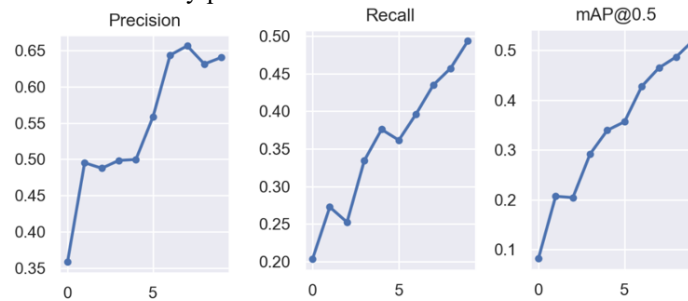
Among them, TP (True Positive) represents a sample that is accurately predicted as a positive sample, FP (False Positive) represents a sample that is not accurately predicted

to be positive, and FN (False Negative) represents a sample that is not accurately predicted to be negative. Precision shows the ratio of accurately predicted positive samples to all detected positive samples. Recall shows the ratio of accurately predicted positive samples to all positive samples. Taking the recall value as the horizontal axis and the precision value as the vertical axis, we can get the Precision-Recall curve (P-R curve). We will find that the values of precision and recall are negatively correlated, and will fluctuate up and down in the local area. In practical applications, since this P-R curve is not convenient to calculate directly, we need to smooth it, as shown in the above formula. Then the performance of the model can be evaluated by the P-R curve. The area between it and the coordinate axis represents the average precision (AP), and mAP is the mean value of AP.

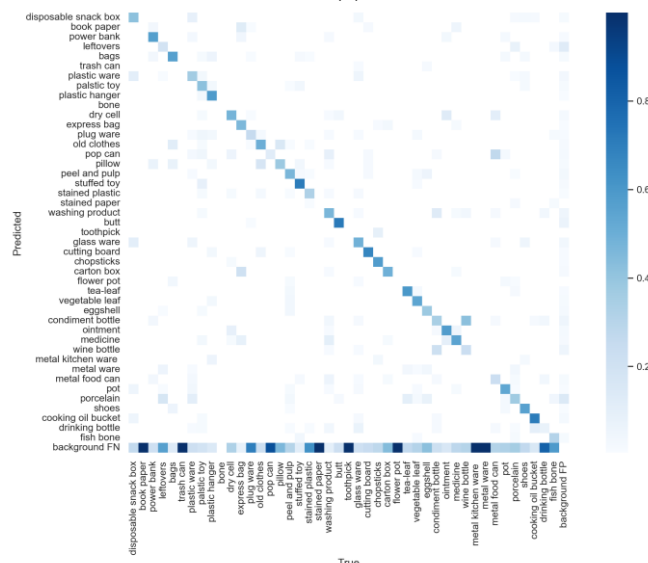
As the Figure2(a) shows, this experiment has a total of 10 rounds of training, the precision reaches 65%, and the recall is about 50%. The mAP also increases with the add in experimental rounds.

From Figure2(b) above, we can see that the model performs well on some such as stuffed toy, cooking oil bucket and washing product. However, it does not perform well on others, for instance, the detections of book paper, metal kitchen ware and stained paper are not very good.

Limited by hardware conditions, the results of this experiment still have room for improvement. From the result on test dataset, it can be seen that there are still some items that have not been detected, which may be due to the shortcomings of YOLO. For example, it is really difficult to detect objects which are next to each other because each grid can propose just 2 bounding boxes and it is also prone to positioning errors of objects.



(a)



(b)

Figure2. Visualization of experimental results, (a) Precision, recall and mAP, (b) Confusion matrix

5. Conclusion

This paper proposes a domestic waste detection model based on YOLO. It has a simple structure and the accuracy of the detected items is really high, which can meet daily use. In the follow-up research, the focus will be on improving the recall rate to enhance the practical value.

References

- [1] Redmon J, Divvala S, Girshick R, et al. (2016) You only look once: Unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas. pp. 779-788.
- [2] Redmon J, Farhadi A. (2017) YOLO9000: better, faster, stronger. In: IEEE conference on computer vision and pattern recognition. Honolulu. pp. 7263-7271.
- [3] Redmon J, Farhadi A. (2018) Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
- [4] Bochkovskiy A, Wang C Y, Liao H Y M. (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
- [5] Girshick R, Donahue J, Darrell T, et al. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. Columbus. pp. 580-587.
- [6] Y. He, J.W. Tian, Z. Zhang, Q. Wang and P. Zhao (2022) Lightweight Research of YOLOv5 Target Detection. Computer Engineering and Applications., 1-9 (In Chinese).
- [7] Pawangfg. (2022) YOLO: You Only Look Once – Real Time Object Detection. <https://www.geeksforgeeks.org/yolo-you-only-look-once-real-time-object-detection/>.