Research on Automatic Garbage Detection System Based on Deep Learning and Narrowband Internet of Things

Ying Liu¹, Zhishan Ge², Guoyun Lv¹ and Shikai Wang¹

¹School of Electronic and Information, Northwestern Polytechnical University, No.127, Youyi Xilu, Xi’an, China
²Shanghai Electro-Mechanical Engineering Institute, No. 1555, zhongchun Road, Shanghai, China.

Email: lvguoyun101@nwpu.edu.cn

Abstract. This paper proposes an automatic garbage detection system based on deep learning and narrowband Internet of things. The system automatically detects and identifies decoration garbage directly in front-end embedded monitoring module, and manages thousands of monitoring front-ends through narrow-band Internet of Things and background server. In the front-end embedded module of the system, the improved YOLOv2 network model is adopted to do garbage detection and recognition. Means of target box dimension clustering and classification network pre-training are used to improve the YOLOv2 model performance; at the same time, the network is lightweight by replacing the feature extraction network and other methods, and the lightweight YOLOv2 network is optimized and ported to embedded module. The experiment test shows that compared with the traditional monitoring system, the cost is reduced by more than half, which can effectively save manpower and material resources, and the accuracy of detection and recognition has also been improved.

1. Introduction

Urban decoration garbage monitoring currently uses the traditional security monitoring mode. Thousands of monitoring points transmit video to back-office service centre through wired or wireless networks, and all kinds of video data are processed manually, some problems are as follows:

- Time limit and mobility of decoration waste, video surveillance front-end needs to be replaced constantly;
- Video transmission network has large flow and high cost, especially in the case of wireless network;
- High background processing capacity, resulting in a lot of waste of human resources;
- It is difficult to simultaneously monitor real-time monitoring of thousands of observation points.

There are very few researches on automatic detection of urban decoration rubbish based on video images in the past literature reports, and artificial intelligence and deep neural networks have been used successfully in images processing, which provides ideas for solving the above problems.

Girshick [1] and Ren [2] et al. respectively proposed Fast R-CNN and Super Fast R-CNN, which increased the detection speed and frame rate while increasing the accuracy rate. It can reach 5 frame/s. Redmon [3] et al. proposed the YOLO network, which borrows structure from the GoogleNet classification network. The difference is that it does not use the inception module, but instead uses a 1*1 convolutional layer + 3*3 convolutional layer to simply replace it, it can reach the speed at which video can be detected (45 frames/s). So YOLO has increased the detection speed and sacrificed the
accuracy rate, it has provided a new way of integrating the classification and positioning for future research. Based on YOLO, Liu [4] et al. and Redmon have successively proposed Single Shot Detector (SSD) and YOLOv2 [5], which have both improved MAP and detection speed. Mittal et al. Literature [6] have conducted research on spam detection based on deep learning. They use a full convolutional neural network as the main tool to segment the area which containing garbage in the image, achieving an accuracy of 87.69%. The boundary of the garbage area extracted by this method is more accurate, but there are more erroneous judgments, and there is a problem that the non-waste area is judged as garbage and part of the garbage is missed. For decoration waste identification, Wei Shufa carried out research on image-based urban scene garbage automatic monitoring [7], they use the Faster-RCNN method with automatic extraction characteristics, strong generalization ability, quasi real-time detection Features. They attempt to introduce the Faster-RCNN method to automatically detect garbage from images of city scenes. However, this model has the problem of insufficient generalization ability and it still needs to be improved in accuracy, and it is unable to realize the transplant of the embedded system.

Overall, great progress has been made in target detection and recognition based on vision and artificial intelligence, but it requires a lot of computing resources, and it is difficult to meet the requirements of the embedded video surveillance front-end module. This paper proposed an improved YOLOv2 model, uses optimization and acceleration algorithm to achieve a balance between real-time performance and precision. At the same time, the model is used to make target box dimensional clustering, classification network pre-training improvement and lightweight processing to facilitate the transplantation of embedded systems. Finally, narrow-band Internet of Thing is been used to do communication among thousands of monitoring terminals and background servers.

The following chapters are organized as follows. The second chapter describes the detection and recognition model of the decoration garbage based on the improved YOLOv2 network; the third chapter describes the lightweight processing of the YOLOv2 network; the fourth chapter describes experimental evaluation and summary for the system.

2. Improved YOLOv2 Network Model

2.1. YOLOv2 Network Model

2.1.1. Model description. The core idea of YOLO is to use the entire map as input to the network and return the bounding box position and bounding box's category directly to the output layer. The YOLOv2 is based on the YOLO model, using a custom network based on Googlenet, which is faster than the VGG-16, requiring only 8.52 billion operations in one forward propagation. It uses a new classification network as feature extraction. Similar to VGG, the author uses more 3*3 convolution kernels and doubles the number of channels after each pooling operation. Reference thought the network in network, a global network using pooled global average pooling. Batch normalization stable model training is also used. Adding batch normalization operations and removing operations after each convolutional layer rarely occurs.

2.1.2. Learning algorithm. Learning algorithm of YOLOv2 is mainly back-propagation algorithm, furthermore, continuous and online learning algorithm response is applied.

- Training process:
  a small batch of gradient descent methods and impulses, the impulse should allow the training process to converge faster. Based on the derivative of the YOLO loss function, the back propagation method is used to continuously update the parameters and the value of the loss function is reduced until it converges. The update rules are as follows:

\[ M_{\omega}^{(i)} = \mu M_{\omega}^{(i)}(t - 1) + \alpha \frac{\partial L(t)}{\partial M_{\omega}^{(i)}} + \lambda \omega^{(i)} \]

(1)

\[ \omega^{(i)}(t) = \omega^{(i)}(t) - M_{\omega}^{(i)} \]

(2)

- Evaluation method:
Recall = \frac{TP}{TP + FN} \quad (3)

Precision = \frac{TP}{TP + FP} \quad (4)

In the formula, TP, FP, and FN are the number of true cases, false positive cases, and false negative cases respectively.

2.1.3. Problems with detection and identification of decoration waste. Although the YOLOv2 network achieves the best detection results for many image targets, it is not entirely suitable for target detection and recognition tasks because:

- Perform dimensional clustering on the target box of the self-made data set and determine the parameters of the anchor. The anchor of YOLOv2 is determined by clustering of VOC2007 and VOC2012 datasets. The categories in the dataset are rich. The determined anchor parameters are universal, but are not suitable for specific detection tasks. Therefore, it is necessary to reconstruct the data in the decoration garbage image inspection dataset.
- Fine-tuning the network using self-made data sets with different resolutions in the classification network training process. Similar to YOLOv2, pre-training is first performed using ImageNet datasets, and the difference is the use of self-made image classification datasets with different resolutions to achieve better fine-tuning effects.
- At present, the size of the YOLOv2 training model is probably due to the limited performance of the front-end equipment, and it is necessary to transplant the model with a limited size. Therefore, it is necessary to reduce the weight of the network.

2.2. YOLOv2 Improved Network Model

2.2.1. Target box dimension clustering. Redmon et al. proposed a method of dimensional clustering. By k-means [8], the target boxes marked manually in the data set were clustered to find the target box. The statistical rule is based on the number of clusters k as the number of anchors and the width and height of the boxes of the k cluster centers as the dimensions of the anchor. The clustering result of voc and coco dataset is 5, so it is determined that the number of anchors in YOLOv2 network is 5. Using the same method, the target box corresponding to the target area in the home-made data set is clustered and analyzed to obtain the optimal number of anchors and the width-height dimension of the data set [9].

2.2.2. Classification Network Pre-training. The mainstream monitoring framework will select the pre-trained classification network on ImageNet [10] for feature extraction. For the shortcomings of the pre-training stage, the following improvements have been made:

- Pre-training Darknet-19 using ImageNet datasets.
- Using a low-resolution (224pixel*224pixel) decorative garbage image classification data set, fine-tune Darknet-19 to adapt the network to the features of the decoration garbage image.
- Modify the Darknet-19 resolution to 448pixel*448pixel. Train 10 rounds on the high-resolution image classification data set, and let the network adjust the weight of each layer to adapt to the high-resolution input. In the pre-training stage, fine-tuning is performed and the resolution of the network is improved.

3. Lightweight Processing of YOLOv2 Model and Monitor Terminal

3.1. Lightweight Processing

3.1.1. Lightweight with MoblieNet. The MobileNet model is based on a depth-decomposable convolution, which can be solved by integrating the standard volume into a deep convolution and a dot
convolution (1*1 convolution kernel). Depth convolution applies each convolution kernel to each channel, and 1*1 convolution is used to combine the output of channel convolutions. This decomposition can effectively reduce the amount of calculations and reduce the size of the model. Replace YOLOv2's original darknet-19 with MobileNet for object detection, training on coco and voc respectively to provide conditions for migration learning;

3.1.2. Pruning, quantization and Huffman coding compression model. We first trim the connections that are not important and retrain sparsely connected networks. Then use the weights to share the weights of the quantized connections, and then perform Huffman coding on the quantized weights and codebooks to further reduce the compression ratio.

3.2. Monitoring front-End Processing system
Monitoring front-end processing system is as follows:

- For mobile video surveillance front-end, real-time video capture of the monitored area;
- H265 encoding and local storage of video for reference;
- Detection and identification of decoration garbage in video images;
- The system has a narrow-band Internet of Things module and a GPS positioning module. When decoration debris is found, the decoration garbage presence information is sent to the background processing centre in real time, and a piece of video before and after the discovery of the garbage is transmitted, and the garbage dumpers therein are monitored;
- Build a network of decoration waste monitoring and control systems based on narrow-band Internet of Things.

4. Experiments and Evaluation

4.1. Space Considerations

4.1.1. Data set. This paper collects data from existing decoration garbage database and self-built database. It uses 570 urban image data containing decoration garbage. The average size of the image is 420*400 pixels, and the collected data is divided into training set and verification. Sets and test sets, including 300 training sets, 120 validation sets, and 150 test sets. In order to improve the generalization ability of the model, we fuse and expand the data. By combining the VOC2007 dataset with urban scene data containing garbage as training data, the difference in background classes can be reduced without increasing the data annotation amount, thereby enhancing the robustness of the algorithm.

4.1.2. Data annotation. The target of interest is annotated in the format specified by the VOC2007 data set. The area of interest is represented using a rectangular bounding box, and the position of the rectangle bounding box is used as a prediction target.

4.1.3. Model selection and initialization of model parameters. We choose Mobilenet as a feature extractor for extracting image features. Training deep learning models requires a large amount of data and long iterations, and we can obtain a limited number of samples, so we use the weights of pre-trained models on the Imagenet classification task. As a full-time initial value of the garbage monitoring model, fine adjustments are made on this basis.

4.1.4. A priori parameter setting of the model. The network parameters are as follows: learning rate is 0.0001; policy is steps; batch is 64; steps are taken as 100, 20000, 35000; max batches are 50000; scales are 10, 0.1, 0.1; momentu is 0.9; and decay is 0.0005.
4.2. Detect and Identify Experimental Results and Evaluations

4.2.1. Experiment configuration and training results. The graphics card is Nvidia GTX 1070, CPU is Intel Core i7-6700, clock speed is 3.40GHz, memory is 32G, operating system is ubuntu 14.04, and frame is Caffe.

4.2.2. Examples of test results. The identified decoration garbage object is marked in a rectangular box. The value above the rectangular box represents the confidence that the algorithm determines that the area is an object of interest. It shows that for different sizes of garbage in different scenes, the algorithm can be marked with a high degree of confidence.

![Identification result example.](image)

4.2.3. Evaluation of Results. Comparison of the results of using data fusion and not using data fusion. By randomly selecting pictures to construct different training sets, verification sets and test sets, 10 experiments were conducted, and the mean and variance of multiple experimental results were taken as the final results. After the test of the two groups of data, P<0.05, it is proved that the fusion of multi-data has a certain effect on improving the robustness of the algorithm.

| Number of test pictures (sheet) | Test the speed (seconds/sheet) | Before data fusion | After data fusion |
|---------------------------------|-------------------------------|--------------------|-------------------|
|                                 |                               | Average detection accuracy (%) | Detection accuracy variance | Average detection accuracy (%) | Detection accuracy variance |
| 150                             | 0.71                          | 83.02               | 0.089             | 89.71            | 0.050             |

- Comparison of Candidate Box Generation Schemes This method is compared with the candidate box generation schemes of Faster R-CNN and YOLO v2. The results are shown in Table 2. The method of dimensional clustering can guarantee a higher number of candidate boxes and less computational resources.

| Candidate Box Generation Scheme | Number of Anchors | Average Overlap Rate |
|---------------------------------|-------------------|----------------------|
| Faster R-CNN                    | 7                 | 0.77                 |
| YOLO v2                         | 5                 | 0.79                 |
| Dimension Clustering            | 4                 | 0.83                 |
4.3. System Evaluation

- Greatly reduce equipment cost
- Reduce a large number of personnel and reduce labor costs;
- Eliminating the problem of wasting network traffic caused by the transmission of large amounts of data in the traditional front-end and back-end, and adopting a low-cost narrow-band Internet of Things can accomplish the result transmission.
- Guaranteed detection accuracy

| Comparing Algorithm   | Precision Ratio/% | Recall Ratio/% | Frames Per Second f/s |
|-----------------------|-------------------|---------------|-----------------------|
| YOLOv2+MoblieNet      | 89.1              | 87.9          | 42                    |
| YOLOv2                | 89.2              | 88.2          | 32                    |
| FAST-RCNN             | 89.7              | 89.0          | 5                     |

Table 3. Comparison of table detection and recognition results

5. Conclusion

This paper proposed an improved YOLO v2 model for decoration garbage detection and recognition through the method of target box dimension clustering, classification network pre-training and other methods, which is optimized and accelerated by lightweight processing, and be ported to an embedded monitor terminal. Narrowband Internet of things can communicate among thousands of monitor terminals and background server centre. The system has the lower cost and the better performance than the traditional system.

6. Acknowledge

This work was supported in part by China shanghai space fund under project SAST2017-004, we would like to thank all master candidates in room 302 for data collection and annotation.

7. References

[1] Girshick R 2015 Fast R-CNN, Proceedings of the IEEE International Conference on Computer Vision pp 1440-1448.
[2] Ren S Q, and Girshick R, et al 2015 Faster R- CNN: towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems pp 91-99.
[3] Redmon J, Divvala S, Girshick R 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.
[4] Liu W, Anguelov D, Erhan D, et al 2016 SSD: single shot mutibox detector. European Conference on Computer Vision pp 21-37.
[5] Redmon J, Farhadi A 2017 YOLO 9000: Better, Faster, Stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition pp 1-9.
[6] Mittal G, Yagnik KB, Garg M, et al 2016 Spot Garbage: smartphone app to detect garbage using deep learning. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing pp 940-945.
[7] Wei Shufa, Cheng Zhaoling 2017 Urban scence garbage automatic monitoring based image processing. J. Journal of international technology, 6, pp 40-52.
[8] Zhang Sujie, Zhao Huaiyi 2017 Algorithm research of optimal cluster number and initial cluster center. J. Application Research of Computers, 34, no.6, pp 1617-1620.
[9] WeiYongming, Quan jicheng, Hou Yuqinyang 2017 Research on UAV aerial photography image positioning based on YOLO v2. J. Laser & Optoelectronics Progress, 54, no.11, pp 111002.
[10] Krizhevsky A, Sutskever I, Hinton G E 2012 Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems pp 1097-1105.