Yolov4 Algorithm for Target Detection in Existing Intelligent Waste Sorting Systems

Pengshu Liu [1], Xinyi Zhang [2] and Zhenyu Xu [3, *]

[1] School of Guangda, Sanhe, China

[2] School of Huangguang, Guangzhou, China

[3] Majestic International College, Foshan, China

* Corresponding Author Email: lvaqbb@ldy.edu.rs

Abstract. The problem of garbage sorting has caused a lot of trouble. Some target detection algorithms are being developed for the current situation of garbage sorting in China. In order to understand the current state of development, we investigate several algorithms suitable for object detection. This paper mainly analyzes three parts of yolov4 network model. In the backbone network, CSPDarkNet53 achieves dimensionality reduction through five convolution kernels, and MobileNetV3 replaces the activation function based on MobileNetV2 and uses MnasNet and Platform-Aware for optimization. Ghostbottleneck is obtained by superposition Ghost module, and then build GhostNet lightweight neural network.The principles of the three attention mechanisms are also explained，SENet using a single-channel convolution kernel，CBAM contains two plates Mc and Ms and CA has a length and width of the two channels. Finally, we explain a data enhancement where four random images are cropped and spliced together.

Keywords: YOLOv4, deep learning, attention, monocular vision, waste sorting.

1. Introduction

China's current waste disposal methods are mainly incineration and landfill, and the sorting of waste before disposal requires a lot of manpower and time. Long hours of manual work are not only inefficient, but the accuracy of sorting is also low, which leads to serious environmental pollution caused by the incineration and landfill of incompletely sorted waste.

Since 2017, China has introduced standardised use and treatment methods for different types of waste, and the importance of waste sorting has increased significantly. In order to solve the problems of waste sorting efficiency, pollution and labour, intelligent waste sorting systems may become an important technology to solve the problem in the future. [1]

The rise of convolutional neural networks in recent years has also promoted the research of rubbish sorting system, and YOLOv4 [2] network is quite popular among researchers due to its ability to greatly improve the recognition accuracy under the premise of guaranteeing the detection speed. Hu et al. [3] obtained 58.57% of the recognition accuracy among 14,964 sample maps by using CSPDarknet53 as the backbone network and adding Triplet attention mechanism. mAP values, and the recognition accuracy increased by 1.91%. Wang et al. [4] replaced CSPDarknet53 with MobileNetV3 [5] to make the network lighter and used CA attention mechanism to obtain about 44% mAP values in a richer variety of spam species and reduced the number of parameters by 89% compared to the original YOLOv4 network and improved the FPS by 51%. Yao et al. [6] used GostNet, also a lightweight neural network, to obtain 94.28% of the mAP values in a dataset of 35 types of rubbish with 10846 images. Li et al. [7] added two attention mechanisms, SENet [8] and CBAM [9], to the network with Mosaic data augmentation, and used the backbone network CSPDarknet53 for the recyclable rubbish of 13 types of rubbish and 36,572 frames of images, a dataset of 36,572 frames of images was used for target detection and a mAP close to 100% was obtained.

With the continuous development of Convolutional Neural Networks, based on the target detection system for rubbish sorting based on the Yolov4 algorithm is also advancing. In this paper, we will

briefly analyse the individual network structures and their important algorithms used in the above literature.

2. Backbone Network

2.1. CSPDarkNet53

CSPDarknet53 is the core of the algorithm and is used to extract the target features. From the figure, it can be seen that this backbone network structure contains five CSP modules. The downsampling of each CSP module can be achieved by a convolutional kernel of size 3×3. The YOLOv4 network model defines the size of the input image as 608×608, and after the five CSP models in the backbone network carry out the feature extraction, the feature map size changes five times, and the feature map finally changes from 608×608 to 19×19, which achieves the fast dimensionality reduction of the feature map. The advantages of using the CSPDarknet53 network structure as the backbone network for YOLOv4 include two aspects: firstly, the ability of the convolutional network for feature extraction is improved without loss of detection accuracy, which improves the detection speed; secondly, the computational loss of the whole model is reduced, which makes it possible to achieve the training of the YOLOv4 model on the CPU with simple configuration.

2.2. MobileNetV3

MobileNetV3 improves on the MobileNetV2 base block by introducing the Squeeze-Excited module between its deep convolutional and separated convolutional layers for feature compression. The hswish activation function is used instead of the swish activation function to improve the efficiency of the model. In terms of network structure, a platform-aware neural architecture search (MnasNet) is used to optimise each lightweight network module and the connectivity between different modules to generate an initial network structure. The initial network structure generated by MnasNet is then optimised using the Platform-Aware Neural Network Adaptation (NetAdapt) algorithm to optimise the parameters of the initial network structure at a hierarchical level with the optimisation goal set, and then the globally optimal model is obtained. It should be noted that [10] research has shown that the MobileNetV3 heavy spindle structure is not suitable for gradient backpropagation, which may lead to early overfitting of the network.

2.3. GhostNet

GhostNet follows the basic architecture of MobileNet-v3 and then replaces the MobileNet-v3 bottlenecks with Ghost bottlenecks, which consist of a series of Ghost bottlenecks built on top of Ghost modules. The first layer is a standard convolutional layer with 16 convolutional cores, followed by a series of Ghost bottlenecks with increasing channels. These Ghost bottlenecks are divided into different stages depending on the size of their input feature maps. All Ghost bottlenecks have a step size of 1 except for the last Ghost bottleneck in each stage, which has a step size of 2. Finally, the feature maps are converted to 1280-dimensional feature vectors using global average pooling and convolutional layers for final classification.

3. Attention Mechanism

3.1. SENet (squeeze-and-excitation networks)

The global average pooling make the original feature map with dimension H×W×C compresse to 1×1×C and the compressed one-dimensional parameter has a global sense field, then two fully connected layers with activation function are used to obtain the weights of each feature channel and show the correlation between feature channels and the importance of each feature channel, and finally the weights that have been obtained are weighted on each original feature map channel by using multiplication operation.

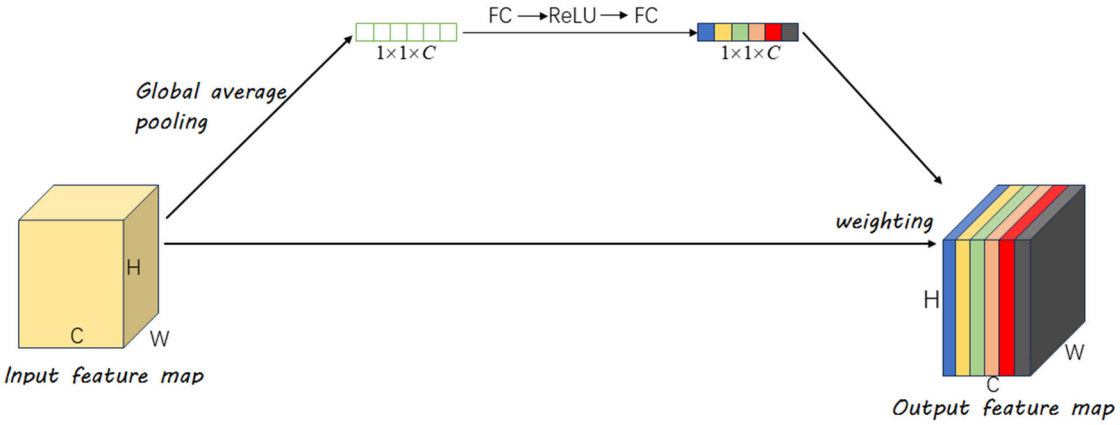


Figure 1 The structure of the SENet attention [7]

3.2. CBAM (Convolutional block attention module)

3.2.1 Mc

The original feature map with dimension H×W×C is processed by global average pooling in addition to global maximum pooling to obtain two sets of 1×1×C feature maps, which are then fed into the shared two-layer neural network MLB for conversion, and the outputs of the two MLBs are subjected to the summation operation, which is finally activated by the sigmoid function to generate the weights of each channel-the channel attention module Mc, and the weights are weighted by the multiplication operation.

3.2.2 Ms

The input of the spatial attention module uses the output of the channel attention module, channel-based average pooling and maximum pooling are performed, the two output H×W×1 feature maps are concatenated according to the channel, then converted to a single channel by convolution, Finally spatial attention module Ms will activate to gain by using a sigmoid function.

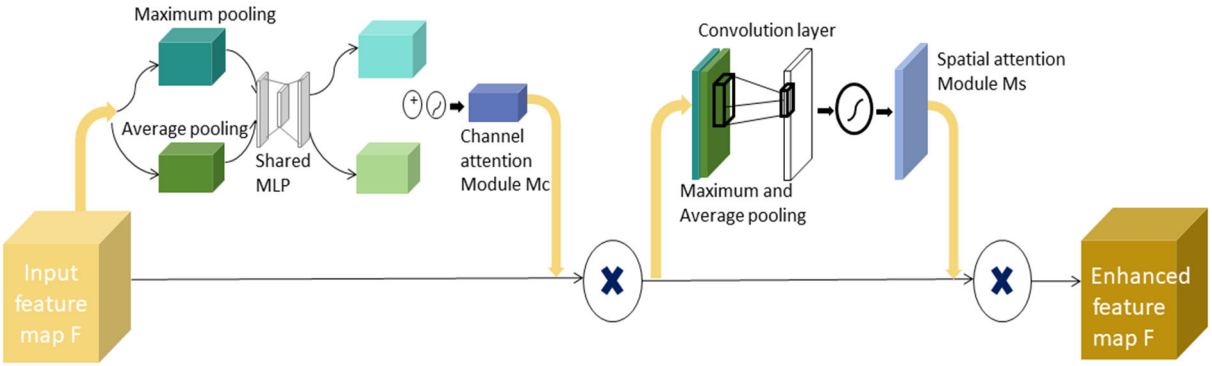


Figure 2 The structure of the CBAM attention [7]

3.3. TA (Triplet Attention)

Different from the patterns of CBAM and SENet, they require a certain number of learnable parameters to establish dependencies between channels, triplet attention consists of three parallel branches, two of them is responsible for capturing channel C and space across the dimensions of the interaction between H/w. The last branch is used to establish spatial attention. Finally a similar

CBAM branch to build spatial attention. The output of the last three branch using the average for summary.

3.4. CA (Coordinate attention)

First of all, the input figure characteristics of the global average pool tumble the height and width of two direction, have the characteristics of the two directions after pooling figure, and then thecharacteristics of the width and height direction graph joining together with global receptive field, then the batch input to the characteristics of the normalized graph F1 sigmoid activation function, Shape of 1 x (W + H x C/r) characteristics of figure f. Then the functional map f is the convolution with the convolution kernel 1×1 according to that time, sigmoid activation function is used to transform the output one-dimensional feature map into 1×(W + H)×C/r, then the feature map f is convolve with the original height and width using a 1×1 convolution kernel to obtain feature maps Fh and Fw with the same number of channels as the original feature map. Obtain the attention weights of the feature map in the height and width directions gh and the attention weights of the feature map in the width direction gw from the sigmoid activation function gw.

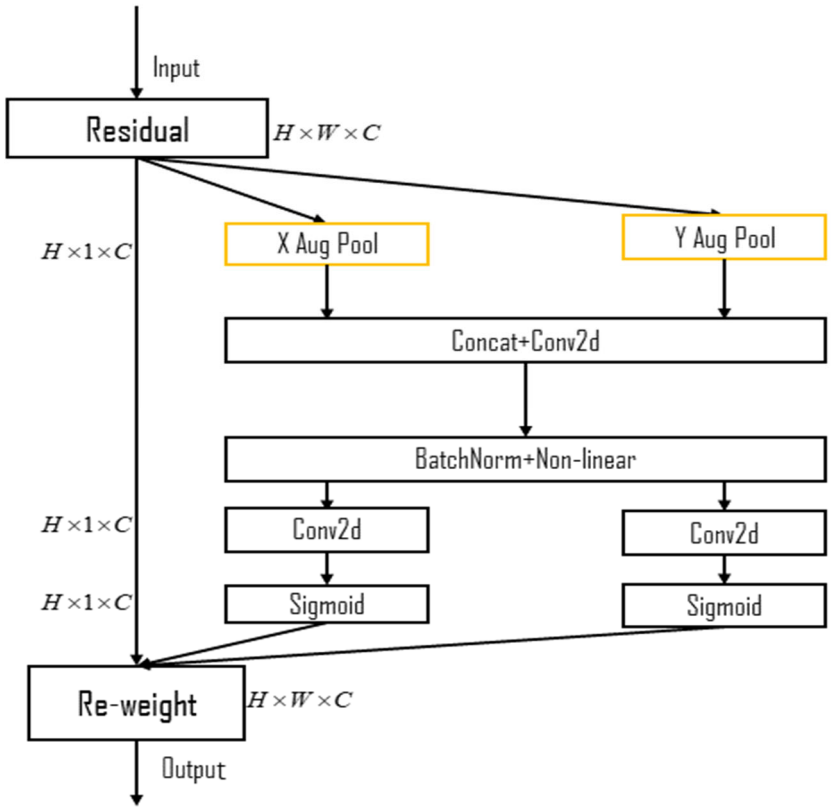


Figure 1 The structure of the CA attention (original)

4. Data Enhancement

4.1. mosaic

Four random original images are subjected to random data enhancement operations such as rotation, scaling, etc. and they are randomly cropped and stitched together into a single image, and the operations are repeated to train the neural network's target recognition ability, resulting in a significant increase in the model's robustness and generalisation ability.

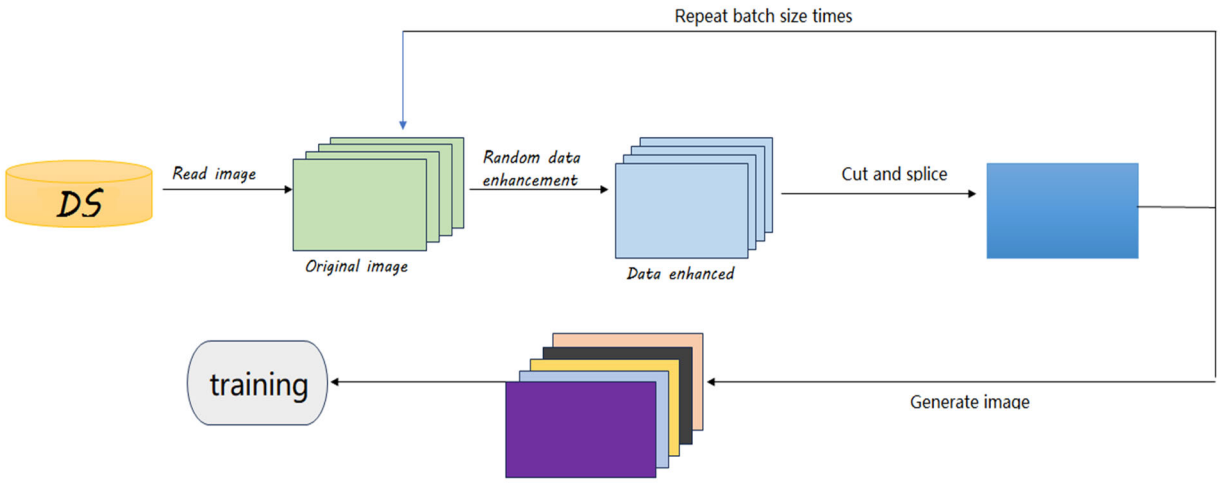


Figure 2 The structure of the Mosaic data enhancement [7]

5. Conclusion

For the time being, YOLOv4-based target detection algorithms have not been sufficiently studied in the field of rubbish sorting, and they have weak detection ability for small targets, limited classification accuracy and are not friendly to mobile with low arithmetic power [5]. Li et al. have achieved better results in the study of sorting recyclable rubbish only, but the algorithms that detect only a single category of rubbish are cannot be applied to the huge scale and variety of rubbish nowadays. In the study of target detection for a comprehensive range of rubbish types, the large size and variety of data sets lead to mismatches in detection accuracy, computation volume, and speed of detection and robotic sorting, which are far from enough to support daily rubbish sorting work.

Nowadays, with the improvement of people's living standard, the type and quantity of rubbish are increasing, incomplete sorting work not only requires a huge labour force but also leads to the special way of dealing with a certain type of rubbish mixed with some other types of rubbish, and incorrect treatment will cause serious pollution of air, water and land, and many recyclable rubbish will be wasted due to sorting problems. Therefore, waste sorting is a major issue for economy, environment and human survival. It is believed that in this deep learning boom, target detection algorithms will be reformed and innovated to be used in future intelligent waste sorting systems.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Gao Genshu, Song Kaiyu, Weng Qinwen. New progress and bottleneck of municipal solid waste separation technology. Environmental sanitation engineering, 2016, 24(6): 11–13, 17.

[2] Alexey Bochkovskiy, Wang Yao, Liao Hongyuan.Yolov4: optimal speed and accuracy of object detection[J]. Computer Vision and Pattern Recognition, 2020.

[3] Hu Yalan, Ke Minyi. Research on Spam Identification Based on Improved YOLOv4. School of Computer Science, Hubei University of Technology, 2021.

[4] Wang Lin, Liu Jing, et al. A lightweight detection algorithm for intelligent waste sorting. Computer System Applications, 2023, 32(4): 231-240.

[5] Andrew G. Howard, Menglong Zhu, Bo Chen, et al. MobileNets: Efficient-convolutional neural networks for mobile vision applications. Computer Vision and Pattern Recognition, 2017.

[6] Yao Lintao, Geng Zhiqing. Lightweight rubbish detection system based on YOLOv4 improvement. Hebei University of Engineering [A]. 2022(11).

[7] LI Qing , GONG Yuanqiang , ZHANG Wei, et al. Attention YOLOv4 Algorithm for Intelligent Waste Sorting [A]. 2022, 58(11): 260-268.

[8] Hu Jie, Shen Li, Gang Sun. Squeeze-and-excitation networks[C]. Computer Vision and Pattern Recognition, 2018.

[9] Woo Sanghyun, Park Jongchan, Lee Joon-Young, et al. CBAM: convolutional block attention module[C]. Computer Vision ECCV,2018.

[10] Zhou Daquan, Hou Qibin, Chen Yunpeng, et al. Rethinking bottleneck structure for efficient mobile network design. proceedings of the 16th European Conference on Proceedings of the 16th European Conference on Computer Vision. 2020: 680-697.