Road Pedestrian Detection Based on Deep Learning

Siyu Jia

School of Information and Intelligent Engineering, Tianjin Ren'ai University, Tianjin 301636, P. R. China

ABSTRACT

Aiming at the problem that deep learning-based road pedestrian detection algorithms are difficult to be deployed in low-performance devices, this paper utilizes the deep learning-based target detection algorithm YOLO V5s, optimizes and updates the network model, changes the preprocessing method of the input side, and adjusts the model structure of the YOLO V5 algorithm, and proposes an improved lightweight pedestrian detection model algorithm, YOLO V5s-LW. The final results show that the YOLO V5s-LW model has a mAP of 95.1%, which is 2.3% lower than that of the YOLO V5s model, but it can reduce the difficulty of deploying devices for pedestrian detection in road environments while maintaining a good detection rate.

KEYWORDS

Target detection; Pedestrian detection; Deep learning; YOLO model

DOI: 10.47297/taposatWSP2633-456904.20230402

1 Introduction

Pedestrian detection as a hot research project currently applied to intelligent transportation, unmanned driving, dense place identification, etc. Pedestrian detection is to identify and locate the human body in the image live or video, and identify the human body, so the recognition accuracy of the algorithmic model is required to be high. The use of deep learning for pedestrian detection, on the other hand, can effectively solve the problems of small target recognition, congestion generated by the occlusion, etc., to achieve the desired effect of pedestrian detection.

For road pedestrian detection, YOLO series of target detection algorithms have great advantages, YOLO only needs a single convolutional network computation to get the target's position and category information, fast detection speed, especially in the video and real-time monitoring of the scene, YOLO algorithms excel in the use of multi-scale fusion detection technology, the recognition of objects at different scales detection, because the pedestrian is a small target detection, compared to complex road environments, the pedestrian is a small target detection. YOLO series algorithms do not rely on specific domain knowledge in the target detection task, and the detection effect is better in all kinds of datasets, so the YOLO algorithm model is extremely versatile, has high performance in pedestrian detection, and has good real-time performance. real-time, presenting better detection speeds running on both CPU and GPU.

Nowadays, deep learning based algorithmic models have been applied to pedestrian detection many times. Cheng Wang, Yuansheng Wang, Feixiang Chang et al.[1] used the multimodal data YOLOv3 (MDY) algorithm for pedestrian detection on embedded devices, and the improved YOLOv3 improved 6.4 % AP and 11.3 FPS over the original algorithm. Lisang Liu, Chengyang Ke, He Lin et al.[2] proposed a pedestrian detection and recognition model MobileNet - YoLo to address the problem that large-scale pedestrian detection networks cannot be directly applied to small-scale device scenarios due to their

heavy weight and slow detection speed. Xue Zhang and Xue Zhang and Chang Li[3] in order to reduce the false detection rate of highway pedestrian detection, optimization is performed at various stages of YOLOv5s model and the mAP of YOLOv5s (IOU=0.5) is improved by 0.48. Panigrahi Sweta and Raju U.S.N.[4] proposed an improved YOLOv2, called Inception Depth- wiseYOLOv2, to detect the height of pedestrians. Xiao Ke, Xinru Lin and Liyun Qin[5] proposed an end-to-end pedestrian detection and re-recognition model in real scenarios to address the problems of lack of pedestrian expressiveness, occlusion, variety of pedestrian poses, and difficulty in detecting pedestrians at small scales.

The YOLO V5 model has good performance and high detection accuracy, but configuring YOLO V5 for road pedestrian detection belongs to a major challenge for low-performance devices, such as cell phones, car recorders, and various types of sports cameras. According to the analysis of the network structure of YOLO V5 model, to effectively solve the problem of YOLO V5 model adaptation, mainly from the Input part of the sample preprocessing, Backone part of the feature extraction, Neck part of the feature enhancement, Head part of the target prediction and loss function, training strategy and other aspects of the improvement and optimization of YOLO V5 according to the size of the model, there are four models: V5 model, V5 model, V5 model, V5 model and V5 model. YOLO V5 has four models: V5s, V5m, V5l, and V5x. Due to the limited hardware equipment, this paper mainly uses the smaller model of V5s for experiments.

Aiming at the problem of algorithm model adaptation in the process of road pedestrian detection, we mainly optimize and improve the YOLO V5s model as follows:

a. Improve the preprocessing of the Input part and adjust the input image size.

b.Improve the number of channels of the model, by reducing the number of channels, the number of parameters of the model and the amount of computation can be reduced, so as to speed up the training and inference of the model.

c.Improvement of Anchor, the re-clustering Anchor can adjust the size of the Anchor according to the actual size and proportion of the target in the dataset, thus better adapting to the characteristics of the target. This helps to improve the detection accuracy of the model and reduces target omission and misdetection.

2 Basic YOLO V5s Algorithm

The four parts of Input, Backone, Neck and Head constitute the YOLO V5s algorithm model. In order to avoid model overfitting caused by the small number of samples in the training set, a data enhancement strategy combining Mosaic data enhancement, adaptive anchor frame computation, and adaptive image scaling techniques is implemented in Input, which increases the diversity of pedestrian samples and improves the generalization ability of the model. CSP technology is introduced in Backone, which is combined with Darknet-53 to form a new structure of CSPDarkNet-53 for improving detection accuracy and speed, while BOS technology is incorporated to significantly improve the accuracy and efficiency of the model when performing pedestrian detection tasks. A new feature fusion module called FPN+FSA is incorporated into Neck, which enables the model to better deal with the complex scenes and changes in light and other factors in the road pedestrian detection task, and has better detection results while also reducing the occurrence of false alarms and missed detections, which improves the model's detection effectiveness and robustness. In order to effectively solve the detection inaccuracies caused by occlusion and polarization due to dense crowds, various techniques such as introducing GIoU_loss, Focal_loss, Anchor-Free and DropBlock are used in Head as a way to improve the ability to detect pedestrians in road environments.

Although the YOLO V5s model shows excellent detection and recognition results in all scenarios, and can effectively avoid the traditional target detection of false detection, omission, slow speed, poor accuracy, etc. as shown in Table 1, but in the pedestrian detection task in the road environment, when pedestrian detection is carried out on the device with a low-performance CPU or GPU, the model runtime is too long, and the requirements on the deployment of the device and environment are too high. environment requirements are too high.

Table 1. Comparison between traditional object detection algorithm and deep learning object detection algorithm

	Traditional object detection algorithm	Deep learning object detection algorithm
Operational speed	Relatively slow	Fast
Recognition accuracy	Low	High
precision	Lower	Higher
Stability performance	Poor	Good

Therefore, to address the above problems, this paper optimizes the YOLO V5s model and proposes an improved lightweight YOLO V5s pedestrian detection algorithm, YOLO V5s-LW, for low-maintenance and high-efficiency pedestrian detection. After research and testing, the effectiveness of YOLO V5s-LW model for pedestrian detection is verified.

3 Improved Lightweight YOLO V5s-LW Model

(1) Pre-processing improvements in the Input section

This experiment is planned to lighten the YOLO V5s model, and each part should be optimized to achieve the lightening goal during the model run. The original YOLO V5s model has an input image size of 640*640 in the Input section, reducing the size of the input samples of the YOLO V5 model can be achieved by adjusting the size of the input image. In the improved lightweight YOLO V5s-LW model, this study changes the image size to 320*320, and the modification is made by directly modifying the size through the code of the original frame of the YOLO V5s model, as shown in Figure 1.

```
def detect_img(self):
    model = self.model
    output_size = self.output_size
    source = self.img2predict  # file/dir/URL/glob, 0 for webcam
    imgsz = [320,320]  # inference size (pixels)
```

Figure 1. Sample size 320*320

Reducing the sample image size means that the model needs to process fewer pixels, which reduces the amount of computation. This can speed up the inference of the model, especially on resource-constrained devices. And smaller image sizes require less memory to store input data and model parameters, thus saving memory consumption, which makes the model easier to deploy and run on memory-limited devices, in line with the need for this experiment to be adapted to low-performance devices. Crucially the smaller image size speeds up training, as the model can complete a training round faster due to the shorter time to process each sample, thus the whole training process is accelerated.

(2) Reduced number of channels

The channels that can be modified in the YOLO V5s model mainly include the input channels and the number of channels in the convolutional layer. The input channels are not modified in this study

because a large number of sample images are color images, which corresponds to an RGB image input channel number of 3.

The number of convolutional layer channels in the YOLO V5s model is mainly modified by halving the number of convolutional layers, which are located in the Backone part of the YOLO V5s model, and the feature extractor part is mainly used for extracting the image features.By halving the number of SPP channels of the Focus, CSP network, and the Neck in the Backone, the number of parameters and the computational amount of the model can be reduced, thus speed up the training and inference of the model. It also reduces the memory footprint of the model, making the model easier to deploy and use in resource-constrained environments. Comparison of Input-size, Params, and GFLOPs of the model before and after modifying the number of channels is shown in Table 2.

Table 2. Comparison of model parameters

Model	Input-size	Params(M)	GFLOPs
YOLO V5s	640*640	7.2	16.5
YOLO V5s-LW	320*320	1.7	1.1

From the Table 2, it can be clearly seen that the lightweighted YOLO V5s-LW model has 16 times less computation and 7 times less parameters than the YOLO V5s model, and the reduction of computation and parameter quantities has great advantages for model lightweighting, improving detection efficiency and reducing the adaptation environment.

(3) Reclustering anchors

The Anchors of the YOLO V5s model are based on the COCO dataset for detection and recognition, the pedestrian detection in the road environment carried out in this experiment is the recognition and detection of the human body in the complex environment, the human body detection dataset mostly presents rectangular boxes, and the sample identification of the COCO database is not applicable to the purpose of this experiment, it is necessary to re-cluster the labeled boxes according to the existing dataset to obtain new Anchors.Figure 2 is the data before Anchors adjustment, the image sample of YOLO V5s-LW model is 320*320, the input resolution becomes smaller, its Anchor needs to be scaled down, Figure 3 is the data after Anchors adjustment based on YOLO V5s-LW model.

```
# Parameters
nc: 3  # number of classes
depth_multiple: 0.33  # model depth multiple
width_multiple: 0.50  # layer channel multiple
anchors:
  - [10,13, 16,30, 33,23]  # P3/8
  - [30,61, 62,45, 59,119]  # P4/16
  - [116,90, 156,198, 373,326]  # P5/32
```

Figure 2. Anchors parameter for 640*640

```
# Parameters
nc: 3  # number of classes
depth_multiple: 0.33  # model depth multiple
width_multiple: 0.50  # layer channel multiple
anchors:
  - [5,6, 8,15, 16,12]  # P3/8
  - [15,30, 31,22, 30,60]  # P4/16
  - [58,45, 78,99, 186,163]  # P5/32
```

Figure 3. Anchors parameter for 320*320

4 Analysis of Experimental Results

(1) Data set segmentation

There are 8530 pictures in this dataset, and the pictures are divided into training set train and test set val to be stored in the dataset. The robustness of the model depends on a large number of training samples, the lack of data will lead to overfitting problems when the model is trained, the model does not have good generalization, and also does not have practical application value, a large number of samples of data can make the model learn more information and improve the model's anti-interference ability. In order to make the training results more generalizable, the number of samples in the training set (7680 sheets) and the test set (850 sheets) are allocated by 9:1, in which the training set is only used for the evaluation of the model's performance, and is not involved in all the training and learning processes of the model.

(2) Model training strategy

In the training phase of the project, the image size of the input model is 320*320, the Batch Size size is set to 16, the initial learning rate is set to 0.01, the momentum is set to 0.9, and the Epoch is 300. this ensures the stability of the convergence, in addition, the model uses the Gradient Descent method (SGD) as the optimization function, the computational speed and convergence of the SGD is fast because the SGD method does not compute the full sample gradient, because of this, it improves the computational speed of the model.

(3) Model evaluation and comparison

This system measures the good performance of speed and accuracy in the target detection algorithm through the metrics of Accuracy, Recall, Precision, mAP (Mean Average Precision), AP (Averge precision), and F1 Score.

The results of the training and test sets in the dataset after model training are shown in Figure. 4, which shows that all six loss curves tend to 0, indicating that the model has a high prediction accuracy for location and target presence, and the experimental results of the test set are smoother, representing that the trained model can show better results on the test set.

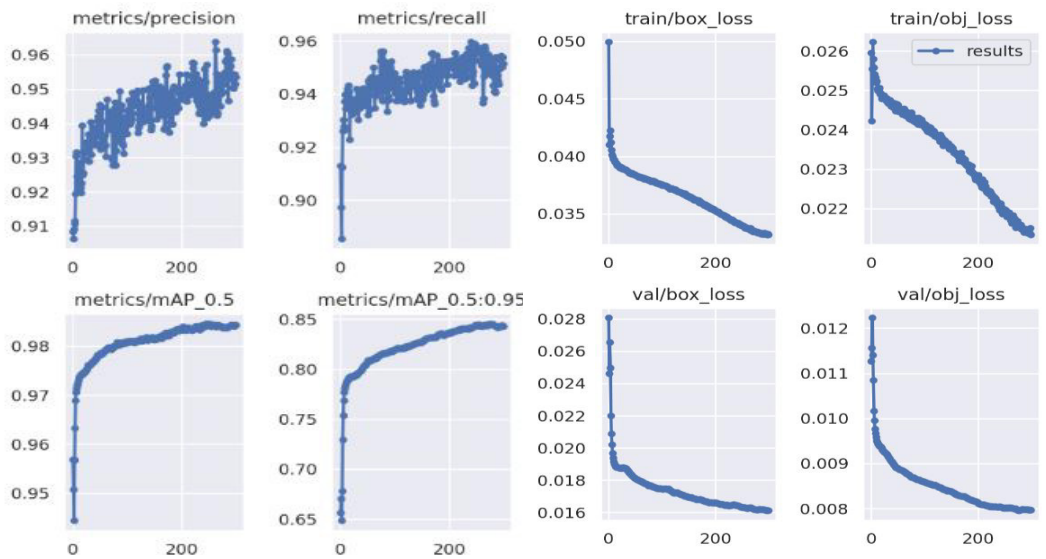

Figure 4.Loss curves

Comparison of various indexes between the improved lightweight YOLO V5s-LW model and the YOLO V5s model after training on the test set is shown in Table 3. The mAP of the YOLO V5s-LW model is reduced by 2.3% compared to that of the YOLO V5s model, and for low-performance devices, this mAP is able to show a better detection capability, which can indicate that the improved YOLO V5s-LW model can realize the low-performance adaptation environment.

Table 3. Comparison of model results

Model	mAP_0.5	mAP_0.5:0.95
YOLO V5s	0.974	0.835
YOLO V5s-LW	0.951	0.722

The final field test results are shown in Figure 5.

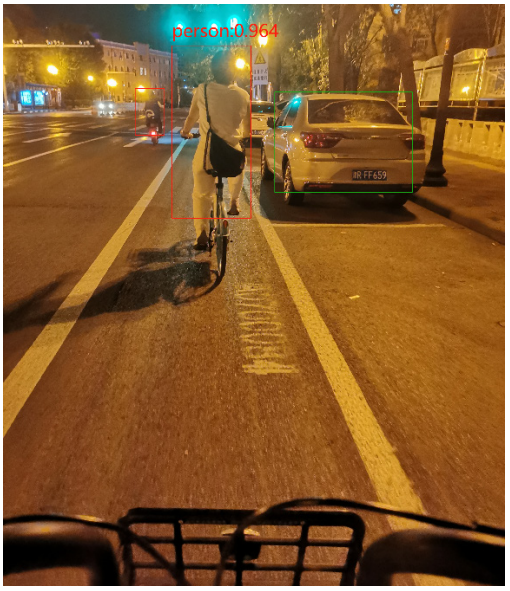

Figure 5. Field testing

5 Conclusion

The small target recognition and pedestrian occlusion problems in the pedestrian detection task are popular among the public, and the design of a set of algorithmic models adapted to low-performance devices is less involved by the predecessors, and at the same time it is required by some devices on the market at this stage. By changing the size of the input sample image, reducing the number of channels in the convolutional layer, and readjusting the Anchors, the author proposes an improved lightweight YOLO V5s-LW model. Compared with the YOLO V5s model, the mAP is reduced by 2.3%, but it can be adapted to low-performance CPU devices, and the slight loss of the mAP has a significant impact on the error rate of the pedestrian detection in road environments, and is within the error tolerance range. The slight loss of mAP is within the error rate of pedestrian detection in the road environment.

References

[1] Cheng Wang, Yuan-sheng Liu, Fei-xiang Chang & Ming Lu (2022) Pedestrian detection based on YOLOv3 multimodal data fusion, Systems Science & Control Engineering, 10, pp. 832-45.

[2] Liu Lisang,Ke Chengyang,Lin He & Xu Hui.(2022).Research on Pedestrian Detection Algorithm Based on MobileNet-

YoLo. Computational Intelligence and Neuroscience. Volume 2022 , pp. 12.
[3] Zhang Xue & Chang Li.(2022).Research on YOLOv5s Highway Pedestrian Detection Algorithm Integrating Attention Mechanism. American Journal of Electrical and Computer Engineering(1). Volume 6, Issue 1, pp. 47-53.
[4] Panigrahi Sweta & Raju U. S. N..(2022).InceptionDepth-wiseYOLOv2: improved implementation of YOLO framework for pedestrian detection. International Journal of Multimedia Information Retrieval(3). Int J Multimed Info Retr 11, pp. 409–30.
[5] Ke Xiao,Lin Xinru & Qin Liyun.(2021).Lightweight convolutional neural network-based pedestrian detection and re-identification in multiple scenarios. Machine Vision and Applications (2). Machine Vision and Applications 32, pp. 46.