# Improvement of Pedestrian Detection Algorithm Based on YOLO

**Xuan Li, Jing Li[a]**

School of Electronic Information Engineering, Shenyang Aerospace University, Shenyang 110136, China

[a]844511219@qq.com

**Keywords:** deep learning, object detection, detection boxes, penalty factor, YOLO V2.

**Abstract:** The non-maximum suppression algorithm is usually used in the post-position of deep learning object detection algorithm. It suppresses the detection boxes with high overlap rate while using the algorithm. In order to avoid the missed and false detection caused by non-maximum suppression algorithm, an improved maximum value suppression algorithm is proposed. When the IOU of the suppression window and the suppressed window is greater than the given threshold, the confidence multiply by the penalty factor instead of discarding it directly. After multiple iterations, we need to remove the lower scores detection boxes. Experiments show that the YOLO V2 deep learning model with improved algorithm has improved accuracy on different data sets as well as strong versatility and robustness.

## 1. Introduction

Object detection is a hot research field in computer vision. The main task of object detection is to accurately identify and accurately locate the object from a complex scene. It is a very importance process for subsequent tasks, such as object tracking and scene understanding. It has been applied in many fields such as medical, transportation, aerospace and so on.

With the development of deep learning, great progress has been made in the field of object detection. In 2013, Ross Girshick proposed a candidate region-based convolutional neural network (R-CNN) [1]. Comparing with sliding window, manual feature selection and classifier detection method, selective search [2] was proposed instead of sliding window. Although R-CNN reduces the range of object search to a certain extent, it still requires a lot of calculation. Then the Kaiming He team and the Ross Girshick team successively proposes series of optimization algorithms, which has greatly improved the accuracy rate and calculate quantity of SPP-Net [3], Fast R-CNN [4], Faster R-CNN [5] and R-FCN [6].

Although the series algorithms of R-CNN have been greatly improved in speed, it still cannot meet the real-time requirements. The emergence of regression algorithms provides a new solution for the accelerating of detection speed. YOLO (You Only Look Once) [7] and SSD (Single Shot MultiBox Detector) [8] which are excellent methods among these algorithms YOLO directly outputs the coordinates and class of the detection boxes, which improves the detection rate and can reach 45 fps, but its detection accuracy is reduced, and it is prone to miss or false detection of small objects. SSD inherits the advantages of Faster R-CNN and YOLO, which combines anchor and regression ideas, so that does prediction from different scales on feature maps, then can greatly enhances its performance. Since then, Redmon et al. proposed the YOLO v2 boxes work. Through the use of batch normalization, dimensional clustering, multi-scale training and other improved methods, the detection speed and accuracy have achieved a better effect, it also has some defects. Non-maximum suppression algorithm (NMS) is often used in object detection algorithms. The traditional NMS algorithm is a greedy algorithm that suppresses all IOU larger than the threshold in the detection of boxes repetition rate. The traditional one leads to missed detection in the objects with higher overlapping domains and reduces the average detection rate of the algorithm (Average Precision, AP). Moreover, the whole process of the algorithm is greatly affected by the threshold. When the threshold is low, the adjacent object detection boxes is mistaken the different subject as the same one and different detection boxes are suppressed. If the threshold is larger, multiple

detection boxes of the same object will not be suppressed. The occurrence of redundancy creates a false positive. According to the situation, the threshold needs to be artificially set which will limit the independence and accuracy of the algorithm.

## 2. YOLO V2 object detection model

In 2016, Redmon et al. proposed YOLO (You Only Look Once) which is an algorithm that regards the object detection problem as a regression problem. As the name says, YOLO does not need additional operations and simply pass the images to the neural network to predict the Bounding Box and class probability of the object. It can not only do predictions based on the global information of the image, but also has a simple structure and speedily detects.

YOLO V2 [9] is improved based on the YOLO network boxeswork, and the network structure is shown in Figure 1. Yolov2 draws on the anchor mechanism of Faster R-CNN and proposes to use K-Means clustering method to find the appropriate number of anchors. In addition, Batch Normalization is added between the convolutional layer and the pooling layer. The input data of each layer is a normal distribution with a mean of 0 and a variance of 1, which accelerates the convergence speed of the model and achieves a certain regularization effect to prevent over-fitting of the model. In the network structure, YOLO V2 adopts the new feature to extract model Darknet-19, which includes 19 convolutional layers and 5 maximum pooling layers. It uses $1 \times 1$ convolutional compression feature maps between $3 \times 3$ convolutions to reduce the model calculations and parameters. The use of global average pooling replaces the full connection layer for network prediction, which greatly improves the ability of the network to extract deep features of the image. And the recognition accuracy of the network has been greatly improved.

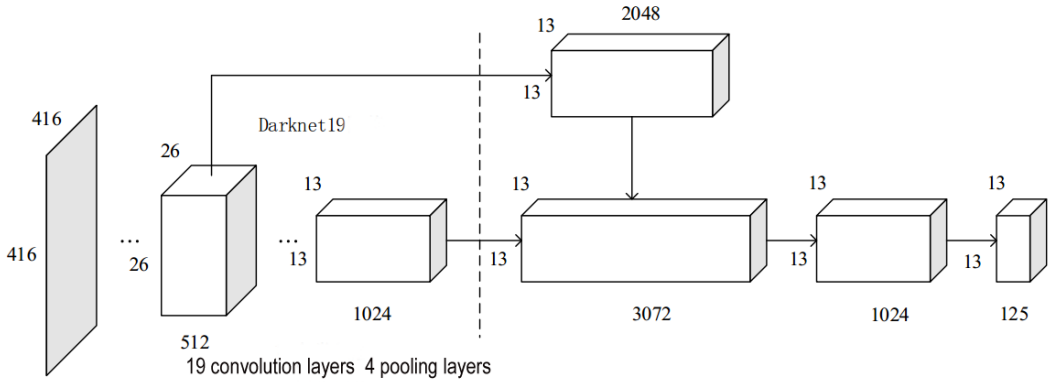


Figure 1. Yolo v2 structure

## 3. Improvement of NMS algorithm for object detection model

### 3.1. Traditional NMS Algorithm

The detection boxes of the neural network output need to be suppressed twice; According to the class confidence, the detection boxes with lower scores will be suppressed firstly. For the second time, the detection boxes with higher scoring rate but high repetition rate will be suppressed until one object corresponds to one detection box. Most of the current object detection algorithms are suppressed by the NMS. The algorithm steps are as follows:

(1) Arranging the detection boxes of all neural network outputs in the confidence descending order from high to low;

(2) Using the window with the highest score as the initial suppression detection box;

(3) The remaining detection boxes performs the IOU calculation by the detection box with the highest score in turn;

(4) Set the threshold and remove all the detection boxes whose cross ratio are bigger than the threshold.

(5) Take the next detection boxes with the highest score again as the initial suppression box;
Repeat steps 2-5 until the IOU among all detection boxes are less than the threshold.
The NMS algorithm can be represented by the following formula:

$$S_i = \begin{cases} S_i & iou(a_i, b_i) < N_t \\ 0 & iou(a_i, b_i) \geq N_t \end{cases} \tag{1}$$

It can be seen that the NMS algorithm is affected by the threshold. And it is difficult to adapt the accuracy rate and the recall rate in this moment. In addition, the NMS algorithm sets the scores of the prediction boxes with high repetition rate to zero, which easily makes the similar and densely distributes objects lose.

### 3.2. Improved Pie-NMS algorithm

Due to the problem of NMS algorithm, the paper proposes a new Pie-NMS algorithm. When the IOU value of the suppression window and the suppressed window is greater than the set threshold, it illustrates that the class of the current two detection boxes is the same as each other. If the class is same, we multiply the class confidence of the box by the new weight instead of directly returning to zero. The prediction box which multiplied by the new weight will be re-executed as a new box, until the optimal detection frame is obtained. If it is different, the current detection boxes will be retained.

$$S_i = \begin{cases} S_i & iou(a_i, b_i) < N_t \\ S_i * (e^{-iou(a_i, b_i)} - e^{-1}) & iou(a_i, b_i) \geq N_t \end{cases} \tag{2}$$

The weight function of Pie-NMS is $y = e^{-x} - e^{-1}$, which is a monotonically decreasing function, and the offset $e^{-1}$ is responsible for mediating the output range of the weight function.

Steps of the Pie-NMS algorithm:

(1) Input detection boxes matrix $B = \{b_1, b_2, b_3 \ldots b_n\}$, confidence matrix $P = \{p_1, p_2, p_3 \ldots p_n\}$, class matrix $C = \{c_1, c_2, c_3 \ldots c_n\}$, horizontal splicing to obtain a new output matrix;

$$B' = \begin{Bmatrix} b_1, p_1, c_1 \\ \ldots\ldots \\ b_n, p_n, c_n \end{Bmatrix}; \tag{3}$$

(2) Select the detection boxes with the highest confidence as the suppression box, and calculate the IOU value of the detection boxes with other detection boxes. If the IOU is greater than the given threshold, we should determine whether the class $C_i$ of the two detection boxes are consistent;

(3) If the class is consistently $P_i$ confidence multiplied by the penalty factor $w_i\left(e^{-IOU} - e^{-1}\right)$. If the class is inconsistent or the IOU value is less than the given threshold ( $N_t = 0.5$ ), the original weight is retained;

(4) Determine whether the confidence of the penalty is less than the given penalty threshold $(sigma = 0.1)$. If it is smaller than the threshold, it will be discarded directly. If it is greater than the threshold, it will return to the second step and re-detect.

The improved Pie-NMS algorithm can effectively suppress the phenomenon of missed detection of objects with dense objects and overlapping objects. When the overlap rate of the same type of objects is higher, we will punish it and then make a second judgment. When different classes of objects overlap, the detection boxes are directly retained.

## 4. Experimental analysis

### 4.1. Experimental data and parameters

The data set used in this experiment is the PASCAL VOC 2007 data sets and the self-made data sets. The PASCAL VOC 2007 data sets have 20 classes of objects, divided into training sets, validation sets, and test sets. The self-made data set is a joint data set which is composed of INRIA pedestrian detection data sets and Beijing pedestrians. The INRIA datasets is the most commonly used static pedestrian detection data sets. The training set has 614 positive samples, the test set has 288 positive samples, the amplified data sets training set has 1200 positive samples, and the test set has 823 positive samples.

### 4.2. Pie-NMS algorithm experimental results

The improved Pie-NMS algorithm on the PASCAL VOC person class and the self-made pedestrian datasets have a significant improvement on accuracy and can suppress the false positive cases by the traditional NMS algorithm is prone to some extent, Figure 2 and Figure 3 respectively show the effect of the improved Pie-NMS algorithm on target missed detection. The threshold T is set to 0.5, and the parameter sigma is 0.1 in the Pie-NMS algorithm.


Figure 2. (a) Original NMS algorithm Figure 2. (b) Improved Pie-NMS algorithm

Figure 2. Pedestrian test data set results


Figure 3. (a) Original NMS algorithm Figure 3. (b) Improved Pie-NMS algorithm

Figure 3. Pascal VOC 2007 dataset results

### 4.3. MAP test

Table 1 shows the MAP comparison of the traditional algorithms on the above two improved algorithms. It can be seen in the table that the algorithm improves the algorithm by 1.07% on the pedestrian detection data set and 1.02% in the VOC2007 Person class.

Table.1. Three Scheme comparing

| Algorithm | Self-made data set | VOC-Person |
|---|---|---|
| This paper algorithm MAP/% | 91.33 | 81.71 |
| Traditional algorithm MAP/% | 90.26 | 82.73 |

The PR curve is one of the important indicators using to evaluate the performance of the model. As shown in Figure 4, it can be seen that the area under the improved b-picture PR curve is larger than the area before the improvement, and it can be seen that when the abscissa is the same as well as the recall rate. The accuracy of the improved algorithm is higher than before, and the better detection result is achieved.

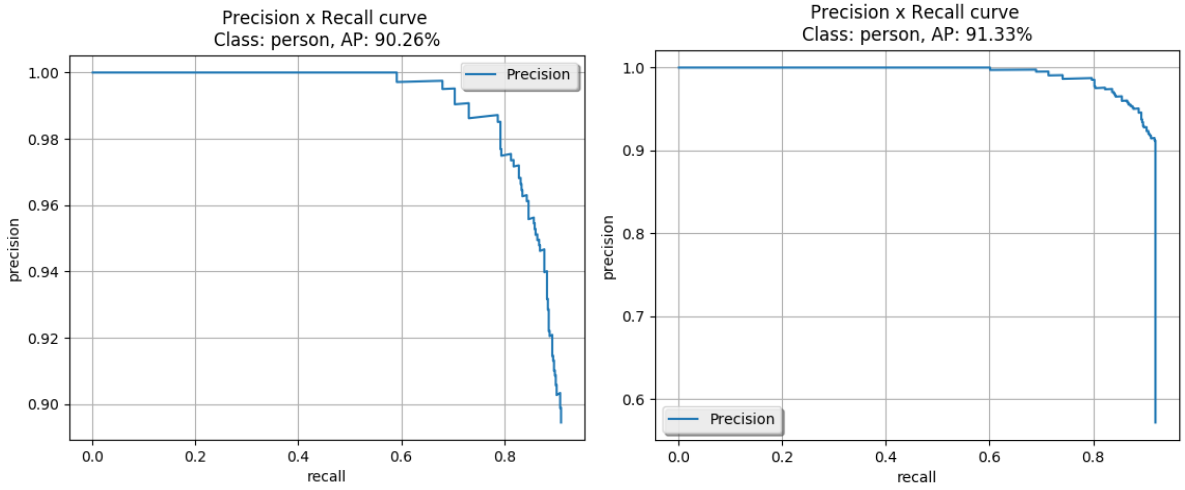


Figure 4. (a) improves the previous PR curve Figure 4. (b) Improved PR curve

Figure 4. Pedestrian test data set PR curve

## 5. Conclusion

Compared with the traditional NMS algorithm, the improved Pie-NMS algorithm which is proposed in this paper can effectively suppress false detection and missed detection ,when the objects are dense. The penalty factor is used to reduce the confidence scores of the detection boxes with higher overlap rate instead of directly zeroing it, so that the missed detection and false detection of the objects that are caused by the improper setting of the threshold T are avoided to some extent. Moreover, the improved NMS algorithm does not add extra computational complexity and algorithm complexity. The accuracy of the YOLO V2 model uses the pie-nms algorithm in PASCAL VOC 2007 and the self-made pedestrian detection INRIA dataset is significantly improved. The algorithm proposed in the paper has strong practicability and robustness.


## Acknowledgments

This work was supported by Liaoning education department science and technology research project (L201715). The authors deeply appreciate the supports.

## References

[1] Girshick R, Donahue J, Darrell T.et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of IEEE conference on Computer Vision and Pattern Recognition. 2014.

[2] J. R. R. Uijlings, K. E. A. Sande, T. Gevers, A. W. M. Smeulders. Selective Search for Object Recognition [J]. International Journal of Computer Vision. 2013 ,104 (2):154-171.

[3] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [C]. European Conference on Computer Vision. 2014.

[4] R. Girshick. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. 2015.

[5] S. Ren, K. He R. Girshick, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Neural Information Processing Systems. 2015.

[6] J. Dai, Y. Li, K. He, et al. R-FCN: Object detection via region-based fully convolutional network [C]. Proceedings of Conference and Workshop on Neural Information Processing Systems,2016.

[7] J. Redmon, S. Divvala, R. Girshick, et al. You only look once: unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[8] Liu Wei, et al. SSD: Single Shot Multi Box Detector [C] PrOCEEDINGS of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[9] J. Redmon, A. Farhadi. YOLO9000: better faster stronger [J] arXiv preprint arXiv:1612.08242,2016.