Object Detection Algorithm for Improving Non-Maximum Suppression Using GIoU

Zhiqiang Hou¹,², Xiaoyi Liu¹,²∗ and Lilin Chen¹,²
¹College of Computer Science and Technology, Xian University of Posts and Telecommunications, Xi’an, China
²Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi’an University of Posts and Telecommunications, Xi’an, China

∗Corresponding author

Abstract. To address the problem of object miss detection and object false detection in single threshold-Non-Maximum Suppression algorithm, this paper proposed a GDT-NMS (Generalized Intersection over Union Dual Threshold NMS, GDT-NMS) algorithm which using GIoU (Generalized Intersection over Union). Using the GIoU indicator computing the similarity between objects, can better describe the relative position and overlap between the objects. And we proposed the dual-threshold NMS algorithm, which can balance the relationship between the object missed detection problem and the object false detection problem, reduce "false positive example" problem. By nonlinearly processing the weight function, the object is better distinguished. The algorithm uses Faster R-CNN as the detector. The experimental results show that the improved algorithm has outstanding performance.

1. Introduction

Object detection [1, 2] is an important research direction and research hotspot in the field of computer vision. It includes many directions, such as text detection, pedestrian detection, etc. [3-4]. Currently, the object detection algorithms based on Deep Learning have made great progress. The main methods are the one-stage detectors represented by YOLOv3[5], SSD[6] and the two-stage detectors represented the Faster R-CNN [7] algorithm.

In object detection, people usually use NMS (Non-Maximum Suppression) and its improved algorithm, The soft-NMS algorithm to suppress the detection boxes. These methods have a great influences in the object detection. However, traditional NMS algorithm is easy to leads to the object is missed largely. And there is a "false positive example" problem that the same object is repeatedly detected with the soft-NMS algorithm [11].

In this paper, we propose a GDT-NMS (Generalized Dual Threshold NMS, GDT-NMS) algorithm. Using the GIoU indicator to compute the similarity between objects, can better describe the relative position and overlap between the objects. And the dual-threshold NMS algorithm can balance the relationship between the object missed detection problem and the object false detection problem caused by the single threshold algorithm, and reduce the repeated detection problem. By nonlinearly processing the weight function, the object is better distinguished. Moreover, we propose object repeated detection rate and the false detection rate of multiple detections to measure "false positive example" problem. The comparison of results indicate that the improved algorithm achieves better detection results. The improved algorithm has a detection accuracy of 74.8% and 25.9% on PASCAL VOC2007 and MSCOCO, respectively. And performance improvements of 1.6% and 1.5% compared to using the NMS...
algorithm. At the same time, compared with the soft-NMS algorithm, the repeated detection rate of our improved algorithm on PASCAL VOC2007 is reduced by 2.6%, and the error rate of multiple detections is reduced by 2.2%.

2. Related Works

2.1. Object Detection Detector
In recent years, the object detectors based on Deep Learning can be divided into two types: one-stage detectors, such as YOLOv3[5] and SSD[6], two-stage detectors, for example, Faster R-CNN [7] algorithm. Specially, In 2015, Ren proposed the Faster R-CNN algorithm, which is a classic algorithm in the field of object detection. At present, many state-of-the-art algorithms have improved on Faster R-CNN. but, the Faster R-CNN algorithm is still the necessary algorithm to compare experimental results and performance. Therefore, we also apply the Faster R-CNN algorithm as detector. Figure 1 shows the flow of the Faster R-CNN algorithm.

2.2. Improvements to the NMS Algorithm
In object detection, the NMS algorithm is a general post-processing method, which adopts the idea of "greedy algorithm" when merging the detection boxes. In this way, largely leads to the object is missed. Therefore, many improvements based on the NMS algorithm have emerged. Reference [8] uses a convolution neural network instead of the NMS algorithm to generate a specific detection box for each object. Reference [9] trained an LSTM to replace the NMS. In reference [10], the NMS algorithm was included in the training phase, and an end-to-end training method was proposed. The soft-NMS algorithm [11] uses the "weight penalty" to solve the missed detection problem in the NMS algorithm, and has achieved good results to a certain extent. All in all, These methods have some contribute to the NMS algorithm, but often can not balance the detection performance and detection speed, some algorithms require much more computation than the NMS algorithm. Especially, The soft-NMS algorithm brings the "false positive example" problem.

3. Improved Non-Maximum Suppression Algorithm

3.1. GIoU as a Measure in the Non-Maximum Suppression Algorithm
IoU’s full name is Intersection over Union. People often use formula (1) to calculate similarity of two different boxes A and B. It is a general evaluation indicator in many fields.

\[
\text{IoU} = \frac{|A \cap B|}{|A \cup B|}
\]

However, using IoU as an evaluation has two problems:
(1) When the two objects don’t overlap, the IoU value is 0. IoU value can’t mirror the distance between the two objects.
(2) According to[12], The IoU value can not correctly reflect the different overlap mode of the two objects.

Figure 1. The flow of the Faster R-CNN algorithm.
Instead, GIoU uses equation (2) to characterize the similarity of two closely spaced detection boxes. Compared to IoU, its advantages are:

\[
GIoU = IoU - \frac{|C \setminus (A \cup B)|}{|C|}
\]  

(1) By introducing a minimum circumscribed rectangle C, the information D in the rectangle C is added, as labeled D in figure 2. And the distance between the two boxes is jointly drawn by A, B, and D. When A and B do not overlap, the GIoU value, that is not be zero, can better “reflect” the distance between the two boxes than IoU.

(2) Different information D can represent the degree of alignment and similarity between two objects. When two objects have the same IoU value, apply GIoU value can distinguish the different overlap modes. As shown in figure 2, obviously, when the two objects have the same IoU value, the value of GIoU can describe overlap modes between the two detection boxes.

Figure 2. Comparison of the overlap ways of two sets of detection boxes with the same IoU values but different GIoU values. A1, A2 are two candidate objects, the B represents the object ground-truth, and rectangle enclosed by dotted lines is the rectangle C containing both objects.

3.2. Double threshold Non-Maximum Suppression Algorithm using GIoU as a Measure

In this work, we propose a GDT-NMS (Generalized Intersection over Union Dual Threshold NMS, GDT-NMS) algorithm using GIoU. The algorithm uses formula (3) to suppress the detection boxes:

In the formula, \( b_i \) is the to-be-detected box, and \( s_i \) and \( s_f \) are the original score and the final score of \( b_i \). \( M \) is the highest-scoring detection box currently, and the thresholds are \( N_i \) and \( N_f \). \( GIoU(M, b_i) \) is GIoU value between \( M \) and \( b_i \).

First, we use the GIoU value to measure the similarity between the two object boxes. Its use weakens the shortcomings of IoU, not only can describe the distance between the two boxes by added blank information, but also better explain the overlap ways of the two boxes, which helps a lot for the positioning of the object boxes. Second, The non-linear log()weight ratio is used as the penalty term of the detection boxes. In this way, the difference between the detection boxes is increased, the object is better distinguished, and the remaining detection boxes is more accurate. Finally, we use the double threshold to improve the NMS algorithm, which balances the relationship between the object missed detection problem and the object false detection problem, reduces "false positive example" problem.

\[
s_f = \begin{cases} 
  s_i & GIoU(M, b_i) \leq N_i \\
  s_i \times \left(1 - \log \left(GIoU(M, b_i) + 1 \right) \right) & N_i < GIoU(M, b_i) < N_f \\
  0 & GIoU(M, b_i) \geq N_f 
\end{cases}
\]  

4. Experiments

The experiments in this paper runs on a computer with 8400CPU and GTX1080GPU, and tests the algorithm in the PASCAL VOC and MSCOCO datasets. And we apply the mean Accuracy Precision (mAP) and speed as performance indicator.
Figure 3. Comparison of experimental results of the algorithm and other algorithms on PASCAL VOC and MSCOCO (a) NMS algorithm (b) soft-NMS algorithm (c) our algorithm.

In this paper, we choose some representative results are selected as Figure 3. It can be seen that our algorithm can balance the relationship between the object missed detection and the object false detection problem and improve "false positive example" problem.

Through many experiments, the algorithm chooses $N_t=0.4$, $N_i=0.9$ as the final threshold. At this point, the algorithm has the highest mean Accuracy Precision and speed, 74.8% and 15.6FPS.

**Table 1.** Comparison of the results on the repeated detection problem.

| Algorithm | Rd_total | Rd2  | Rd3  | C_error | Rd_rate | C_error_rate |
|-----------|----------|------|------|---------|---------|--------------|
| soft-NMS  | 592      | 543  | 49   | 64      | 11.9%   | 10.8%        |
| OURS      | 461      | 433  | 28   | 40      | 9.3%    | 8.6%         |

Moreover, compared with the soft-NMS algorithm, the Rd_rate of the algorithm is reduced by 2.6%, and the error rate of multiple detections is reduced by 2.2% on PASCAL VOC2007 dataset. Table 1 shows comparison result. The total number of repeated detections is indicated by $Rd_{total}$, and the number of repeated detections of 2 and 3 times for the same object are indicated by $Rd_2$ and $Rd_3$, respectively. $C_{error}$ indicates the number of repeated detection and each detection classification is different, and $Rd_{rate}$ is the repeated detection rate and $C_{error_rate}$ is object misclassification rate for repeated detection. We apply the formula (4) and (5) to calculate them.

\[
Rd_{rate} = \frac{Rd_{total}}{pic\ total} \times 100\% \\
C_{error\ rate} = \frac{C_{error}}{Rd_{total}} \times 100\%
\]

Table 2 shows partial results of our algorithm and other algorithms in the PASCAL VOC2007 dataset. It can be seen that our algorithm have higher mAP, such as bird, table, car. Table 3 shows the comparison of the results of our algorithm and other algorithm. It can be seen that, the mAP of our algorithm is 74.8% and 25.9%, respectively. Our algorithm both the PASCAL VOC2007 and MSCOCO datasets have achieved best results.

**Table 2.** Partial results of the algorithm and other algorithms in the PASCAL VOC2007 dataset.

| Algorithm  | aero  | bike | bird | table | motor | car   | person | cat   | plant | tv   |
|------------|-------|------|------|-------|-------|-------|--------|-------|-------|------|
| FasterR-CNN| 76.5  | 79.0 | 70.9 | 65.7  | 77.5  | 84.7  | 76.7   | 86.4  | 38.8  | 72.6 |
| soft-NMS   | **76.9** | **81.3** | 74.8 | 67.5  | **79.6** | 86.2  | 81.0   | 85.6  | 43.4  | 73.7 |
| OURS       | 77.3  | 79.8 | **75.1** | **67.8** | 78.8  | **87.8** | **81.9** | 84.3  | **45.3** | **74.4** |
Table 3. Test results of the algorithm and other algorithms.

| Algorithm | Network | PASCAL VOC | MSCOCO |
|-----------|---------|------------|--------|
|           |         | Trainingdata | mAP | speed | trainingdata | AP0.5:0.95 | AP@0.5 |
| SSD       | VGG16   | 07+12       | 74.3 | 46    | trainval     | 23.2      | 41.2   |
| FasterR-CNN | VGG16 | 07+12       | 73.2 | 15.1  | trainval     | 24.4      | 45.7   |
| RON320    | VGG16   | 07+12       | 74.2 | 15.0  | train        | 22.7      | 44.7   |
| soft-NMS  | VGG16   | 07+12       | 74.3 | 15.3  | trainval     | 25.5      | 46.7   |
| OURS      | VGG16   | 07+12       | **74.8** | 15.6 | trainval     | **25.9** | **47.1** |

5. Conclusion
In this article, a improve post-processing algorithm is proposed, and many experiments have been tested on PASCAL VOC and MSCOCO datasets, and good results have been obtained. In the future work, we will consider introducing adaptive threshold Non-Maximum Suppression algorithms into new basic networks, such as Res-Net [13] and Dense-Net [14] for better object detection performance.

Acknowledgments
I will express my thanks to all people that helped me during this work. First, I acknowledge the help of my teacher, Hou Zhiqiang. Because he has gave me valuable suggestions in this work. Second, I would like to thank Chen LiLin, who provided me help when I was writing.

References
[1] Liu L, Ouyang W, Wang X, et al. Deep Learning for Generic Object Detection: A Survey[J]. 2018, arxiv: 1809.02165.
[2] Huang K, Reng W, Tieniu T. Survey of image object classification and detection algorithms[J]. J. Comput. Sci, 2014, 37(6): 1225-1240.
[3] Jiejie Zhu, Omar Javed, Jingen Liu, Qian Yu, Hui Cheng, Harpreet Sawhney. Pedestrian detection in low-resolution imagery by learning multi-scale intrinsic motion structures [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3510-3517.
[4] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, Hwalsuk Lee. Character Region Awareness for Text Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 9365-9374.
[5] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[OL]. 2018, arXiv:1804.02767.
[6] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Cheng-Y F, Berg A C. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer VisionSpringer International Publishing, 2016:21-37.
[7] Ren S, He K M, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks[C]//International Conference on Neural Information Processing Systems. MIT Press, 2015:91-99.
[8] Hosang J, Benenson R, Schiele B. Learning non-maximum suppression [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6469-6477.
[9] Stewart R, Andriluka M, Ng A Y. End-to-end people detection in crowded scenes[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition,2016: 2325-2333.
[10] Henderson P, Ferrari V. End-to-end training of object class detectors for mean average precision[C]//Proceedings of the 13th Asian Conference on Computer Vision, 2016: 198-213.
[11] Bodla N, Singh B, Chellappa R, Davis L S. Soft-NMS – Improving Object Detection With One Line of Code [C]// IEEE International Conference on Computer Vision, 2017:5562 - 5570.
[12] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, Silvio Savarese. Generalized Intersection over Union: A Metric and A Loss for Bounding Box, 2019, arXiv:1902.09630

[13] He K M, Zhang XY, Ren SQ, Sun J. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.

[14] Huang G, Liu Z, Maaten L V D, Weinberger K Q. Densely Connected Convolutional Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017:2261-2269.