EasyChair Preprint

№ 8186

An Advanced IoU Loss Function for Accurate Bounding Box Regression

Ho-Si-Hung Nguyen, Thi-Hoang-Giang Tran, Dinh-Khoa Tran and Duc-Duong Pham

June 4, 2022

An Advanced IoU Loss Function for Accurate Bounding Box Regression

Ho-Si-Hung Nguyen[1], Thi-Hoang-Giang Tran[2*], Dinh-Khoa Tran[3], and Duc-Duong Pham[4]

[1] University of Science and Technology, The University of Danang, Vietnam
nhshung@dut.udn.vn
[2] University of Science and Technology, The University of Danang, Vietnam
tthgiang@dut.udn.vn
[3] University of Science and Technology, The University of Danang, Vietnam
tdkhoa@dut.udn.vn
[4] FTECH CO., LTD, Vietnam
duongpd@ftech.ai
*Corresponding author: tthgiang@dut.udn.vn

Abstract. Bounding box regression (BBR) plays a key role in object detection. To improve accuracy of recognition ability between true object and prediction object, many researches have developed loss functions for BBR. In existing researches, some main drawbacks can be shown: (*i*) IoU loss functions is inefficient enough to detect the object; (*ii*) the loss functions ignore the imbalance issues in BBR when the huge amount of anchor boxes and the target boxes overlap; (*iii*) the loss functions own redundant parameters which lead to extend training process. To solve these issues, this paper is proposed a new approach by using an Advanced IoU (AIoU) loss function. Three geometric factors including overlap area, distances and side length are considered in the proposed function. The proposal focuses on the overlap area to improve accuracy for object detection. By this way, the proposal can relocate anchor box for covering the ground truth in the training process and optimize anchor boxes for object detection. The proposal is tested on MS COCO and VOC Pascal dataset. The experimental results are compared to existing IoU models and show that the proposal can improve accuracy level for Bounding box regression.

Keywords: Advanced IoU, bounding box regression, object detection, loss function, geometric factors

1 Introduction

In recent years, using deep learning for object detection have got many extraordinary achievements. Some applications can be mentioned as visual tracking [7], face recognition[2], face detection[8], etc. Thanks to the quickly development of deep learning, many modern object detection models had been successfully studied. Depending on the number of detector module to create candidate bound-

ing boxes, the basic solutions include one-stage[14], two-stage[4] detector. Two-stage detector is restricted by a small number of bounding box proposal using a category-independent method such as R-CNN[5], and etc. On the other hand, one-stage detector takes advantage of the densely pre-defined candidate boxes (anchors) to achieve high inference speed. From pieces of information in AP-loss article[1] point that one-stage detector generally computes generally faster than two-stage detectors. So that, they could be easily scrutinized about notable holes in the accuracy.

One of the problems that one-stage detector can solve is the imbalance between the foreground and background regions. To decrease this imbalance, Zhang et al. [15] built a novel single-shot based detector (RefineDet). This model bases on a feed-forward convolutional neural network that generates a certain number of bounding boxes and the scores proving the presence presence difference in classes of those bounding boxes. As a result, a non-maximum suppression produced can reduce imbalance between the foreground and background regions. Another modern object detection is YOLOv5[11]. The structure of the model is composed of two parts: the backbone part containing the bottleneckCSP, SPP and other modules. Based on this structure, the characteristic information is well maintained. In addition, this structure can prevents background noise inside the foreground in the training process

Besides that, the process of Intersection over Union (IoU) mainly contributes to stabilizing the foreground-background issue. To enhance the ability of bounding box regression, researchers have been proposed the definition of Intersection over Union (IoU) [13]. IoU is defined as the proportion of the intersection and union of two bounding boxes. A lot of researches have proposed IoU models to enhance the accuracy for object detection such as Complete IoU (CIoU) [16], Distance IoU (DIoU) [16], Generalized IoU (GIoU) [13], IoU+ [12] and etc. In our study, we also propose a new IoU, named Advanced IoU(AIoU), to concentrate on the space in the true bounding box and enhance the geometric factors of bounding box regression as well as the better inference of deep models for object detection. The AIoU is a promising solution to improve accuracy level for BBR.

The rest of this paper is constructed as follows. The section 2 is assigned to describe the related works and proposes an AIoU loss function by taking into account complete geometric factors. The next section gives experimental results tested on MS COCO and PASCAL VOC dataset. After that, the evaluation is proposed based on the analysis of the experimental section. Finally, the conclusion deduced from this work is presented in the last section.

2 Related work and AIoU loss function

2.1 The overlap area

In recent years, the overlap area in BBR is usually thought of as IoU loss function[12].

$$L_{iou} = 1 - IoU \tag{1}$$

This function faces much trouble in training regression such as regression speed, stability optimization problems and so on. Lu, Z. et al. [10] had figured out a new highlight formula to fix these issues as follows:

$$Co = \frac{B^{pd} \cap B^{gt}}{B^{gt}} \tag{2}$$

$$L_{Co} = \gamma * IoU + (1 - \gamma) * Co \tag{3}$$

Where γ represents a positive trade-off parameter. It demonstrates the weight of intersection over union and intersection over ground true. From the study of Lu, Z. et al. [10], γ should be set in [0.75, 0.8, 0.85] to get a good result of validation period in the training duration. In our study, γ has been defaulted as 0.8.

2.2 The distances

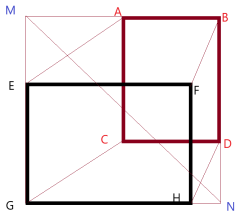


Fig. 1: Schematic diagram of corner distance

The Lo-IoU[10] and CIoU[17] were proposed to optimize the route between the true box and the prediction box. The problem of these proposals is how decrease the distances between two vertices of two boxes and change the length of edges of predicted boxes in order to close to the length of true boxes gradually. To solve this problem, a distance loss function is proposed as below:

$$L_{distances} = \frac{EA^2 + BF^2 + DH^2 + CG^2}{4MN^2} \tag{4}$$

Where AE, BF, CG, DH are the distance between each corner of two boxes. MN is the diagonal distance lengthwise of the smallest enclosing box enveloping two boxes as Fig 1. It should be noted that black rectangle is defined the true box and the red rectangle represents the prediction box.

2.3 The side length

The CIoU given by Zheng, Z. et al.[17] is a great technique to regress scale of w^{gt} and w toward 1 as well as h^{gt} and h toward 1. The formula was defined as:

$$V = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{5}$$

The trade-off parameter helped normalize **V** to [0, 1], which was defined as:

$$\alpha = \begin{cases} 0, & \text{if } IoU < 0.5; \\ \frac{V}{(1-IoU)+V} & \text{if } IoU \geqslant 0.5; \end{cases} \tag{6}$$

The aspect ratio is a partial improvement of CIoU[17] over DioU[16]. When $IoU \geqslant 0.5$, αV will contribute to enhance the accuracy level of BBR.

2.4 The proposed AIoU loss function

In order to enhance the accuracy level of BBR, we proposed an AIoU loss function. Three geometric factors are considered in the proposed function including the side length, the distances and the overlap area. As a result, the formula was defined as:

$$L_{AIoU} = 1 - L_{co} + L_{distance} + V\alpha \tag{7}$$

$$= 1 - (\gamma IOU + (1-\gamma)Co) + \frac{EA^2 + BF^2 + DH^2 + CG^2}{4MN^2} + V\alpha \tag{8}$$

According to theory provided by Kosub, S [6], $1 - L_{co}$ is defined as a part of AIoU. This part plays the key role in the classification task. In that way, the calculation process is going to increase the percentage of union between the ground true area and predicted area. The proposed $L_{distance}$ presented in subsection 2.2 is used to obtain more well performance. $L_{distance}$ not only adjusts the rate between each edge length of the true box and predicted box, but also the calculation obtains a considerably improved localization accuracy without a complex operation. Finally, the side length $V\alpha$ taken from subsection 2.3 helps to improve the accuracy level for BBR.

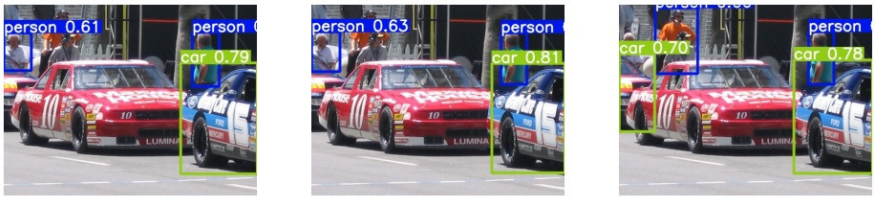
The stability level of L_{AIoU} can be checked by technique requirements . Each of parts must follow properties: (1) $0 \leqslant 1 - L_{co} \leqslant 1$, (2) $0 \leqslant L_{distance} < 1$, (3) $0 \leqslant V\alpha \leqslant 1$, $V\alpha$ is normalized to [0, 1] according to the proving in CIoU[17]. Hence, it confidently ensures that the value of L_{AIoU} always limits in range of [0, 3). So that, this evidence points out the good stability of the formula. L_{AIoU} significantly reduces risks of overwhelmed computation.

3 Experimental results

In this part, the experimental datasets are two popular benchmarks, they are MS COCO[9] and PASCAL VOC dataset[3]. PASCAL VOC dataset has 20 classes. the VOC2012 containing 11,530 for train/val is used for this experience. MS COCO is large image recognition/classification, object detection, segmentation, and captioning dataset. The COCO train-2017 split (115k images) is used for training and report the ablation studies on the 2017 split(5k images). The dataset has 80 different object categories with 5 captions per image.

3.1 Base model

YOLOv5: YOLOv5 is selected for the based model due to its highlighting quality. Model is trained with default parameters for three stages Train > Val > Test in turn. We choose version S of YOLOv5 due to the simple structure of the model, with little BottleneckCSP lead to decrease time-consuming in the training process. the results are collected after 12 iterations, 0.6 confidence score for VOC PASCAL dataset. The comparison results of three demos are demonstrated in Fig. 2



(a) IoU (b) CIoU (c) AIoU

Fig. 2: Comparison examples between AIoU and other models

Table 1 presents the obtained results when the proposed AIoU is compared to the existing models for object detection. The primary baseline of these determinants had been applied on the two validation datasets of MS COCO and PAS-CAL VOC. The experimental results implemented prove that the proposed AIoU loss is better accuracy than the others as shown in Table 1. The performance is significantly improved for object categories with a large ratio of background to object pixel as Fig. 2. It should be noted that although accuracy is improved thanks to the proposed AIoU, its training time is not higher than other models as provided by Table 2.

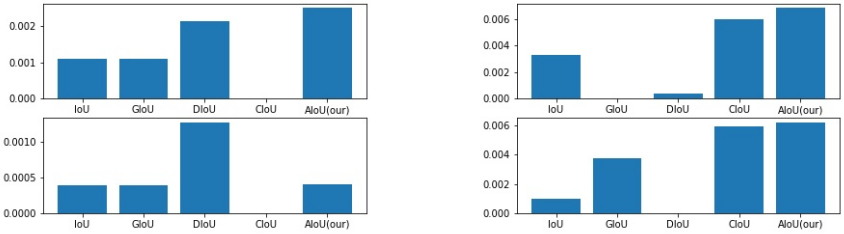
Table 1: Accuracy ability of different IoU models

Method	$mAP_{0.5}(VOC)$	$mAP_{0.5:0.95}(VOC)$	$mAP_{0.5}(COCO)$	$mAP_{0.5:0.95}(COCO)$
IoU	0.78031	0.50398	0.53934	0.34755
GIoU	0.77705	0.50671	0.53934	0.34755
DIoU	0.77744	0.50297	0.54037	0.34844
CIoU	0.78301	0.50892	0.53823	0.34716
AIoU(our)	**0.78388**	**0.50916**	**0.54074**	**0.34757**

The Fig. 3 has shown that the mAP score of AIoU models got extraordinary ability at the final iteration. Because the AIoU is more outstanding than other IoU models at mAP score after a permanent training time.

Table 2: Training time of different IoU models during 12 iterations

$Method$	$time_{minute}(VOC)$	$time_{minute}(COCO)$
IoU	74.5	535.8
GIoU	74.3	530.2
DIoU	74	529
CIoU	75.4	545
AIoU(our)	73.4	540.4



(a) MS COCO (b) PASCAL VOC

Fig. 3: The change in meam average precision (mAP) of each method in tests with the MS COCO and the PASCAL VOC datasets.

4 Conclusions

In this paper, the AIoU loss function is proposed to enhances geometric factors. The proposed loss function deeply focus on overlap area to get the most outstanding foreground. The AIoU loss function simultaneously considers three geometric factors including the side length, the distances and the overlap area. Based on the proposal, the ratio of length and width side are changed at the same time in training process of deep learning model. It should be noted that the purpose of training process is to find and classify a huge number of objects on an image. By this way, the prediction box reaches to the true box fastly. The experimental results confirm that the proposed AIoU loss significantly contribute toward object detection in term of accuracy. The disadvantage of the proposed model is running time. In the near future, we try to shorten the running time by combining the distances and the side length to create a more simple IoU model.

Acknowledgment

This work was supported by The University of Danang, University of Science and Technology, code number of Project: T2020-02-38.

References

1. Chen, K., Lin, W., See, J., Wang, J., Zou, J., et al.: Ap-loss for accurate one-stage object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
2. Deng, J., Guo, J., An, X., Zhu, Z., Zafeiriou, S.: Masked face recognition challenge: The insightface track report. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1437–1444 (2021)
3. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html
4. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 580–587 (2014)
6. Kosub, S.: A note on the triangle inequality for the jaccard distance. Pattern Recognition Letters **120**, 36–38 (2019)
7. Li, W., Xiong, Y., Yang, S., Deng, S., Xia, W.: Smot: Single-shot multi object tracking. arXiv preprint arXiv:2010.16031 (2020)
8. Li, X., Lai, S., Qian, X.: Dbcface: Towards pure convolutional neural network face detection. IEEE Transactions on Circuits and Systems for Video Technology (2021)
9. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
10. Lu, Z., Liao, J., Lv, J., Chen, F.: Relocation with coverage and intersection over union loss for target matching. In: VISIGRAPP (4: VISAPP). pp. 253–260 (2021)
11. Wang, L., Zhang, H., Yang, T., Zhang, J., Cui, Z., Zhu, N., Liu, Y., Zuo, Y.: Optimized detection method for siberian crane (grus leucogeranus) based on yolov5. Tech. rep., EasyChair (2021)
12. Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: Unitbox: An advanced object detection network. In: Proceedings of the 24th ACM international conference on Multimedia. pp. 516–520 (2016)
13. Zhai, H., Cheng, J., Wang, M.: Rethink the iou-based loss functions for bounding box regression. In: 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). vol. 9, pp. 1522–1528. IEEE (2020)
14. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9759–9768 (2020)
15. Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z.: Single-shot refinement neural network for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4203–4212 (2018)
16. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: Distance-iou loss: Faster and better learning for bounding box regression. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12993–13000 (2020)
17. Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., Zuo, W.: Enhancing geometric factors in model learning and inference for object detection and instance segmentation. arXiv preprint arXiv:2005.03572 (2020)