IoU의 최적화에 관한 연구

서신*
*한양대학교 컴퓨터소프트웨어학과
xiner6899@gmail.com

A Study on the Optimization of IoU

Xu Xin*
*Dept. of Computer Software, Hanyang University

요 약

IoU (Intersection over Union) is the most commonly used index in target detection. The core requirement of target detection is what is in the image and where. Based on these two problems, classification training and positional regression training are needed. However, in the process of position regression, the most commonly used method is to obtain the IoU of the predicted bounding box and ground-truth bounding box. Calculating bounding box regression losses should take into account three important geometric measures, namely the overlap area, the distance, and the aspect ratio. Although GIoU (Generalized Intersection over Union) improves the calculation function of image overlap degree, it still can't represent the distance and aspect ratio of the graph well. As a result of technological progress, Bounding-Box is no longer represented by coordinates x,y,w and h of four positions. Therefore, the IoU can be further optimized with the center point and aspect ratio of Bounding-Box.

1..서론

IoU (Intersection over Union) is a common evaluation standard in target detection. It mainly measures the degree of overlap between the bounding box and ground-truth bounding box generated by the model, and is often used to evaluate the advantages and disadvantages of the bounding box. It is often used in target detection or semantic segmentation tasks in the field of deep learning.

In the target detection task, we often make the model generate a large number of candidate bounds at one time, and then sort the frames according to the confidence of each frame, and then calculate the IoU between the frames in turn. The method of NMS (Non-maximum suppression) is used to judge which one is the object we are really looking for and which ones should be deleted. After we get the final output, we can also take the IoU between the output box and the ground-truth bounding box and use 1-IoU as the Loss (interval [0,1] to find the minimum value), and In this way, iterative optimization of the model is achieved.

Advantage：

[1] It can reflect the detection effect of prediction detection box and real detection box.

[2] It also has a good feature of being scale invariant. In the regression task,the most direct indicator of the distance between the predict box and the ground-truth bounding box is IoU.

Disadvantage:

[1] If the two boxes do not intersect,by definition, IoU=0, it cannot reflect the distance between them (coincidence degree). At the same time, because Loss=0, there is no gradient return, so the learning training cannot be carried out.

[2] For two objects with the same IoU, their alignment IoU is not sensitive.

GIoU(Generalized Intersection over Union) has one more "Generalized" than the IoU. This also means that it can calculate IoU on a more general level, which can solve the problem of "when tw

o images do not intersect, the distance between two images cannot be compared".

Features of GIoU:

[1] Similar to the IoU, GIoU is also a distance measure. As a loss function, it can meet the basic requirements of the loss function.

[2] GIoU is still insensitive to scale.

[3] GIoU is the lower bound of IoU. If the two boxes coincide, IoU=GIoU.

[4] The value of IoU is [0,1], but GIoU has a symmetric interval, the value range is [-1,1]. The maximum value is 1 when they coincide, and the minimum value is -1 when they have no intersection and are infinitely far apart. So GIoU is a very good measure of distance.

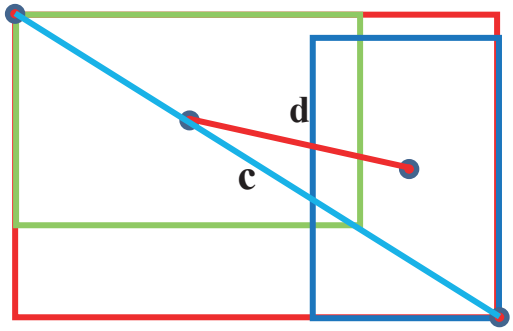
[5] Unlike an IoU, which only focuses on overlap. GIoU focuses not only on overlapping regions, but also on other non-overlapping regions. Therefore, it can better reflect the degree of coincidence between the two.

Although GIoU has improved the calculation function of image overlap, it still cannot express the distance and similarity of the image well.

2..IoU의 최적화의 요구사항

The calculation of the IoU can be considered as comparing whether two boxes or two image regions are in the same position. Therefore, the Euclidean distance can be used to calculate the normalized distance between the center points of the two bounding boxes. On the basis of distance, we can increase the idea of fitting the ratio of width to height of prediction box and the ratio of width to height of target box.

$$IoU - \frac{\rho^2_{(O,O^{\text{Ground-truth}})}}{C^2} - \boxed{X}$$



(그림 1) 공식 사진.

$O, O^{\text{Ground-truth}}$: represents the center point of the two boxes.

ρ : represents the Euclidean distance between the two center points.

C : represents the diagonal of the smallest enclosing rectangle.

X is the penalty term for the shape difference calculated from the aspect ratio. Used to reflect the difference between the two boxes.

$$X = \frac{4}{\pi^2}(\arctan\frac{w^{Ground-truth}}{h^{Ground-truth}} - \arctan\frac{w}{h})^2$$

The range of arctan is $[0, \pi/2)$.

Finally, gradient descent is needed to optimize the loss function.

The gradient of w and h needs to be specified.

$$\frac{\partial X}{\partial w} = 2*\frac{4}{\pi^2}*(\arctan\frac{w^{\text{Ground-truth}}}{h^{\text{Ground-truth}}} - \arctan\frac{w}{h})*(-1)*\frac{1}{1+(\frac{w}{h})^2}*h^{-1}$$

$$= \frac{8}{\pi^2}*(\arctan\frac{w^{\text{Ground-truth}}}{h^{\text{Ground-truth}}} - \arctan\frac{w}{h})*(-1)*\frac{h^2}{w^2+h^2}*h^{-1}$$

$$= -\frac{8}{\pi^2}(\arctan\frac{w^{\text{Ground-truth}}}{h^{\text{Ground-truth}}} - \arctan\frac{w}{h})*\frac{h}{w^2+h^2}$$

$$\frac{\partial X}{\partial h} = 2*\frac{4}{\pi^2}*(\arctan\frac{w^{\text{Ground-truth}}}{h^{\text{Ground-truth}}} - \arctan\frac{w}{h})*(-1)*\frac{1}{1+(\frac{w}{h})^2}*w*(-1)*h^{-2}$$

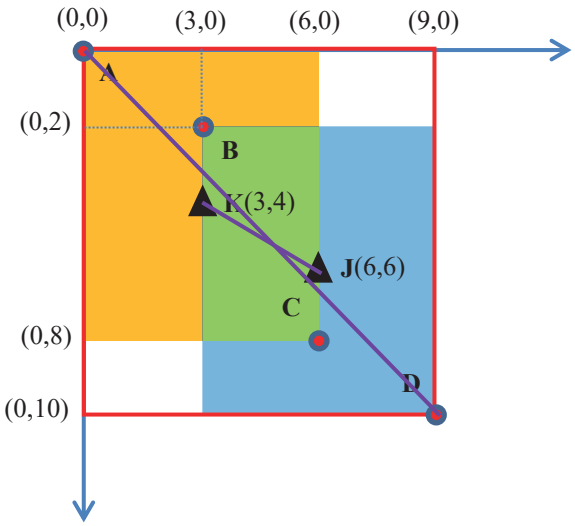
$$= \frac{8}{\pi^2}*(\arctan\frac{w^{\text{Ground-truth}}}{h^{\text{Ground-truth}}} - \arctan\frac{w}{h})*\frac{h^2}{w^2+h^2}*w*h^{-2}$$

$$= \frac{8}{\pi^2}*(\arctan\frac{w^{\text{Ground-truth}}}{h^{\text{Ground-truth}}} - \arctan\frac{w}{h})*\frac{w}{w^2+h^2}$$

Overlap region factor has higher priority in regression, especially in non-overlap case. Therefore, a weight can be added, and the overlapping area can control the weight.

$$\frac{X}{Loss_{IoU} * X}$$

As shown in figure:



(그림 2) 실제의 예 사진.

By calculation, the coordinates of center point K in the yellow box and center point J in the blue box :

K: (3,4) J: (6,6)

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

=(Cx-Bx) * (Cy-By) / (Cx-Ax) * (Cy-Ay) + (Dx-Bx) * (Dy-By) - (Cx-Bx) * (Cy-By)

=(6-3)*(8-2)/(6-0)*(8-0)+(9-3)*(10-2)-(6-3)*(8-2)

=(3*6) / ((6*8)+(6*8)-(3*6))

= 18/78

= 0.23

$$IoU - \frac{\rho^2_{(o,o^{Ground\text{-}truth})}}{C^2} - \frac{X}{Loss_{IoU} * X} * X$$

$$= 0.23 - \frac{\sqrt{3^2+2^2}^2}{\sqrt{9^2+10^2}^2} - \frac{X}{Loss_{IoU} * X} * \frac{4}{\pi^2}(\arctan\frac{6^{Ground\text{-}truth}}{8^{Ground\text{-}truth}} - \arctan\frac{6}{8})^2$$

= 0.158 - 0

= 0.158

In the example, the first two matrices have the same shape. So the penalty term for the calculated shape difference X is 0.

Initially, IoU was only used as a simple evaluation standard, mainly used to measure the degree of overlap between the bounding box generated by the model and the ground-truth box that is the correct result of the label. In target detection, in order to make the positioning more accurate. Make region-proposal closer to ground-truth. You need to fine-tune region-proposal with bounding-box regression. Therefore, IoU can be optimized by making full use of the characteristics of IoU in the original technology. Not only are you limited to measuring the overlap between the two boxes, you can also measure the distance and shape differences between the two boxes. More flexible than before, can be further rapid convergence and performance improvement.

참고문헌

[1] Ross Girshick, “Fast R-CNN”, The IEEE International Conference on Computer Vision (ICCV), Santiago Chile, 2015, pp.1440-1448

[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas America, 2016, pp.779-788

[3] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick, “Mask R-CNN”, The IEEE International Conference on Computer Vision (ICCV), Venice Italy, 2017, pp. 2961-2969

[4] Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis, “Soft-NMS -- Improving Object Detection With One Line of Code”, The IEEE International Conference on Computer Vision (ICCV), Venice Italy, 2017, pp.5561-5569

[5] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar, “Focal Loss for Dense Object Detection”, The IEEE International Conference on Computer Vision (ICCV), Venice Italy, 2017, pp.2980-2988

[6] Zhaowei Cai, Nuno Vasconcelos, “Cascade R-CNN: Delving Into High Quality Object Detection”, The IEEE Conference on Computer Vision and Pattern Recogni

tion (CVPR), Salt Lake City America, 2018, pp.6154-6162

[7] Hei Law, Jia Deng, "CornerNet: Detecting Objects as Paired Keypoints", The European Conference on Computer Vision (ECCV), Munich Germany, 2018, pp.734-750

[8] Chenchen Zhu, Yihui He, Marios Savvides, " Feature Selective Anchor-Free Module for Single-Shot Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach California America, 2019, pp.840-849

[9] Yihui He, Chenchen Zhu, Jianren Wang, Marios Savvides, Xiangyu Zhang, "Bounding Box Regression With Uncertainty for Accurate Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach California America, 2019, pp.2888-2897

[10] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, Silvio Savarese, "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach California America, 2019, pp.658-666