

文章编号:1007-2780(2023)11-1468-13

基于自适应聚焦 CRIoU 损失的目标检测算法

肖振久¹, 赵昊泽², 张莉莉², 夏羽³, 郭杰龙^{4*}, 俞辉⁴, 李成龙², 王俐文²

(1. 辽宁工程技术大学 软件学院, 辽宁 葫芦岛 125000;

2. 中国兵器工业集团 航空弹药研究院有限公司, 黑龙江 哈尔滨, 150000;

3. 上海宇航系统工程研究所, 上海 201100;

4. 中国科学院 海西研究院 泉州装备制造研究中心, 福建 泉州 362000)

摘要: 在目标检测任务中, 传统的边界框回归损失函数所回归的内容与评价标准 IoU (Intersection over Union) 之间存在不相关性, 并且对于边界框的回归属性存在一定不合理性, 使得回归属性不完整, 降低了检测精度和收敛速度, 甚至还会造成回归阻碍的情况。并且在回归任务中也存在样本不均衡的情况, 大量的低质量样本影响了损失收敛。为了提高检测精度和回归收敛速度提出了一种新的边界框回归损失函数。首先确定设计思想并设计 IoU 系列损失函数的范式; 其次在 IoU 损失的基础上引入两中心点形成矩形的周长和两框形成的最小闭包矩形周长的比值作为边界框中心点距离惩罚项, 并且将改进的 IoU 损失应用到非极大值抑制 (Non-Maximum Suppression, NMS) 处理中。接着引入两框的宽高误差和最小外包框的宽高平方作为宽高惩罚项, 确定 CRIoU (Complete Relativity IoU, CRIoU) 损失函数。最后在 CRIoU 的基础上加入自适应加权因子, 对高质量样本的回归损失加权, 定义了自适应聚焦 CRIoU (Adaptive focal CRIoU, AF-CRIoU)。实验结果表明, 使用 AF-CRIoU 损失函数对比传统非 IoU 系列损失的检测精度最高相对提升了 8.52%, 对比 CIoU 系列损失的检测精度最高相对提升了 2.69%, 使用 A-CRIoU-NMS (Around CRIoU NMS) 方法对比原 NMS 方法的检测精度提升 0.14%。将 AF-CRIoU 损失应用到安全帽检测中, 也达到了很好的检测效果。

关键词: 目标检测; 边界框回归; IoU 损失函数; 非极大值抑制; 自适应聚焦损失

中图分类号: TP389.1 **文献标识码:** A **doi:** 10.37188/CJLCD.2023-0005

Object detection algorithm based on adaptive focal CRIoU loss

XIAO Zhen-jiu¹, ZHAO Hao-ze², ZHANG Li-li², XIA Yu³, GUO Jie-long^{4*},

YU Hui⁴, LI Cheng-long², WANG Li-wen²

(1. College of Software Engineering, Liaoning Technical University, Huludao 125000, China;

2. Air Ammunition Research Institute Co. Ltd., NORINCO Group, Haerbin 150000, China;

3. Shanghai Institute of Aerospace System Engineer, Shanghai 201100, China;

4. Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Quanzhou 362000, China)

Abstract: In the object detection task, there is no correlation between the regression content of the traditional

收稿日期: 2023-01-06; **修订日期:** 2023-01-26.

基金项目: 福建省科技计划项目 (No. 2021T3003); 泉州市科技计划项目 (No. 2021C065L)

Supported by Fujian Science and Technology Plan (No. 2021T3003); Quanzhou Science and Technology Plan (No. 2021C065L)

*通信联系人, E-mail: gjl@fjirsm.ac.cn

bounding box regression loss function and the evaluation standard IoU (Intersection over Union), and there is some irrationality for the regression attribute of the bounding box, which reduces the detection accuracy and convergence speed. In addition, the sample imbalance also exists in the regression task, and a large number of low-quality samples affect the loss function convergence. In this paper, a novel loss function, termed as CRIoU (Complete Relativity Intersection over Union), is proposed to improve the detection accuracy and convergence speed. Firstly, this work determines the design idea and determines the improved IoU loss function normal form. Secondly, on the basis of IoU loss, the ratio of the perimeter of the rectangle formed by the two center points and the minimum closure area formed by the two frames is introduced as the penalty term for the distance between the center points of the bounding box, and the improved IoU loss is applied to the non-maximum suppression. Then, the side error of the two frames and the side square of the minimum bounding box are introduced as the side penalty term, a novel loss function, termed as CRIoU (Complete Relativity Intersection over Union), is proposed. Finally, on the basis of CRIoU, an adaptive weighting factor is added to weight the regression loss of high-quality samples, and an AF-CRIoU (Adaptive focal CRIoU) is defined. The experimental results show that the detection accuracy of the AF-CRIoU loss function compared with the traditional non IoU series loss is up to 8.52%, the detection accuracy of the CIoU series loss is up to 2.69%, and the A-CRIoU NMS (Around CRIoU Non Maximum Suppression) method compared with the original NMS method is up to 0.14%. In addition, AF-CRIoU loss is applied to the detection of safety helmet, which also achieves good detection results.

Key words: object detection; bounding box regression; IoU loss function; non-maximum suppression; adaptive focal loss

1 引 言

几年来深度学习在交通检测、汽车自动驾驶技术^[1-2]、目标跟踪^[3]、面部识别^[4]、光学计量^[5]、光学重建^[6-7]等方面都有着十分广泛的应用,深度学习在这些场景广泛的应用都归根于深度神经网络的快速发展。

使用卷积神经网络(Convolutional Neural Networks, CNN)检测物体的想法在 1998 年就产生了。Lecun^[8]等人最先将 CNN 分类方法应用在检测图像中的目标。后来,这种网络变换方式被称为全卷积网络(Fully Convolutional Network, FCN)^[9]。因为 AlexNet^[10]的提出,使得以图像分类网络为主干图像特征提取网络的下游任务如:3D 目标检测^[11]、姿态检测、图像分割、面部识别等得到了快速发展。Girshick 等人将 AlexNet 的目标识别功能应用到目标检测中,提出了一个目标检测算法(Regions with features, R-CNN^[12])。R-CNN 基于卷积神经网络(CNN)、线性回归和支持向量机(Support Vector Machines, SVM^[13])等算法的目标检测框架,通过大量的计算开销实现了目标检测技术。Girshick 等人在 2015 年先后提出了 Fast

R-CNN^[14]、Faster R-CNN^[15],算法将目标检测任务分为两个阶段,并且分类和检测共同训练,这种方式增加了检测速度和检测精度,但仍存在计算量较大的问题。Joseph 等人在 2015 年提出了 YOLOv1^[16]算法。但 YOLOv1 算法缺少对锚框的预先设定,虽然检测速度很快,但是检测精度却很低。Joseph 在 YOLOv1 算法基础上引入了不同尺度的预设锚框,提出了 YOLOv2^[17]算法和 YOLOv3^[18]算法。这些改进使得 YOLO^[19-20]系列算法在保障检测速度的同时也在一定程度上提高了算法的精度。

在深度学习的目标检测算法中,为了使预测框更快、更好、更准地回归到目标的位置,对边界框回归损失函数的研究也经历了一系列的发展。在 Fast R-CNN 中边界框的回归使用了 Smooth L1 损失函数。而在早期 YOLO 系列算法中使用的是误差平方和函数作为损失函数。以上两种损失函数存在回归数值和评价指标不相关的问题,从而回归效率不高。Yu^[21]等人提出 IoU 损失函数,直接对真实框和预测框相交的部分进行优化,实现了到了优化对象和评价指标的相关性,但 IoU 作为损失优化存在无法梯度回传的特殊情况。由

此 Rezatofig H^[22]等人提出广义交并比(Generalized Intersection over Union, GIoU)作为回归损失部分的损失函数, GIoU通过回归缩小不相交区域来优化,但在两框包含情况下会退化为IoU损失。Zheng^[23]等人认为一个优秀的IoU损失应该考虑回归3个几何度量,即重叠面积、中心点距离和宽高比,并且在IoU损失后加入中心点距离惩罚项和宽高比惩罚项,提出了Distance-IoU(DIoU)和Complete IoU(CIoU)。但CIoU中回归的是宽高的比值,会出现回归阻碍的问题。基于以上问题,本文提出了一种自适应聚集CRIoU(Adaptive focal Complete Relativity, AF-CRIoU),完善回归属性,解决IoU、GIoU、CIoU在特殊情况下的回归阻碍问题。本文主要贡献总结如下:

(1)基于IoU系列损失函数范式,提出A-CRIoU(Around CRIoU)损失函数,完善了回归任务和评价指标的相关性和边界框中心点回归属性的完整性,提高了检测精度和收敛速度,同时弥补了IoU与GIoU在特殊情况下的缺点。将A-CRIoU损失应用到传统NMS方法中,改善漏检的情况。

(2)基于A-CRIoU完善边界框宽高属性回归属性,不但考虑重叠面积和边界框距离,进一步添加关于宽高回归的惩罚项,而且惩罚项避免了CIoU回归宽高比所造成的阻碍回归问题,保证了回归边界框回归属性的完整性,提出了CRIoU损失。结合Focal Loss解决分类问题的样本不均衡问题的方法,解决在回归过程中高质量样本和低质量样本不均衡的问题,减少低质量样本过多而影响模型的学习。选用自适应的加权因子,避免了Focal Loss中对超参数调整的外开销,由此定义AF-CRIoU损失。

(3)将创新的A-CRIoU损失应用到目标检测网络的非极大值抑制(Non-Maximum Suppression, NMS)处理中,通过调整不同得分阈值,提升了检测精度。

(4)将最终创新的AF-CRIoU损失函数分别在YOLOv3和Faster R-CNN网络中进行实验,将AF-CIoU应用到安全帽检测中并分析结果。

2 相关工作

2.1 基于CNN的目标检测

基于CNN的目标检测算法包括单阶段检测

算法和双阶段目标检测算法。相比于传统目标检测,基于分类的卷积神经网络目标检测有:通过滑动窗口提取感兴趣区域、筛选感兴趣区域、后处理的主要3个步骤。滑动窗口和后处理部分都是常用的方法。所以这种方法通过对网络中的特征提取模块和分类模块的改进,来提升网络的特征提取能力和分类能力,从而获得更为准确的感兴趣区域。

单阶段目标检测算法没有感兴趣区域筛选部分,直接通过卷积神经网络完成特征提取、边界框定位以及目标分类等任务。最早被提出的单阶段目标算法是OverFeat^[24]。YOLO系列算法和SSD^[25]通过对网络的锚框定位部分和特征提取结构的改进达到了更好的检测效果,其中SSD算法在多尺度目标检测中取得了显著效果。RetinaNet^[26]中的Focal Loss损失函数解决了数据集中难易样本和正负样本的影响,其优秀的结构和检测效果受到了广泛关注。

2.2 IoU损失函数

IoU作为一个评价指标,真实反映了检测任务中预测和真实物体两个形状的重叠程度。IoU通过面积比值的方式来表示两个形状之间的相交程度。IoU具有包括尺度不变性在内的众多特质,这些特质保证了IoU的衡量效果。IoU计算公式如式(1):

$$\text{IoU} = \frac{\text{Overlap}}{\text{Union}}, \quad (1)$$

其中:Overlap代表两个形状相交部分的面积,Union代表两个形状面积的总和。由于IoU特性和计算结果满足了目标检测任务中目标的定位,它也被应用于回归损失函数中,并且IoU损失在目标检测任务中取得了较好的结果。但当两个形状没有相交区域时,IoU损失函数的计算值为1,因没有梯度回传而无法进行优化。

2.3 GIoU、DIoU、CIoU问题分析

GIoU损失通过在损失函数中引入惩罚项的方式来避免损失值为0,梯度无法回传的情况。如公式(2),式中 $1 - \text{IoU}$ 部分表示真实框和预测框之间不相交的面积,通过对这一部分进行回归来缩小两框不相交的面积,从而提高两框的重叠程度,公式的后半部分表示两框在最小外接矩形C中A与B不相交的空白区域,所以当两框不相交时,依然可以通过最小化空白区域来进行边界框

回归。图 1 仿真了 GIoU 损失的回归过程,红色为目标框,绿色和黑色为预测框。当两框关系为互相包含关系时,GIoU 会退化成 IoU 损失,导致收敛速度变慢。

$$L_{GIoU} = 1 - IoU + \frac{|C - A \cup B|}{|C|}. \quad (2)$$

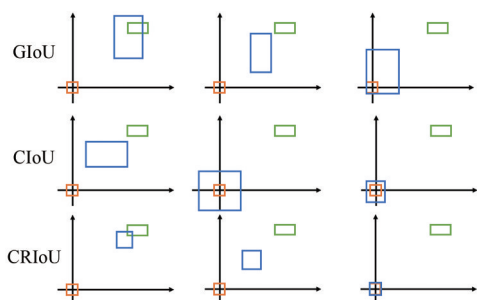


图 1 不同边界框损失函数回归过程的区别

Fig. 1 Difference of regression process of different boundary box loss functions

Zheng^[20]等人表示一个优秀的边界框回归应该考虑重叠面积、中心点距离和宽高比这 3 种几何度量。DIoU 和 CIoU 则包含这 3 种几何度量的回归。DIoU 是在 IoU 的基础上引入了两框中心点距离的惩罚项,CIoU 损失则是在 DIoU 的基础上添加宽高比惩罚项,因此 CIoU 损失包含了重叠面积、中心点距离和宽高比 3 种几何度量的回归。但是在 CIoU 添加的宽高比惩罚项中回归的具体数值是宽高比,CIoU 中的宽高比惩罚项和具体的宽高值没有真实的对应关系。在回归比值过程中,如果预测的宽和高的真实值增大如图 1 中的 CIoU 回归过程所示,在第二次采样回归状态时,由于回归比值,使得预测框的高误差增大,因此 CIoU 在回归比值的情况下会造成真实值和预测值的误差变大,这种情况下会对边界框的回归造成阻碍。

2.4 Focal Loss 函数

在目标检测中单阶段目标检测运行速度较快,但是受数据集中难易样本和正负样本的影响较大,其精度低于两阶段目标检测。为解决上述问题,Lin 等^[26]提出了 Focal Loss 函数。当数据集中存在大量简单的样本时,这些损失值可以淹没稀有的困难类,使困难类别不能被学习影响目标检测的精度。而 Focal Loss 引入超参数的调制因子就是为了解决损失值被大量简单样本淹没的情

况,平衡难易样本的数量,使困难样本在训练期间得以表现,避免了只有简单样本对模型的贡献。Focal Loss 公式如图 2 中公式所示, γ 是控制曲线形状的超参数, γ 的值越大,易分类的样本的损失就越小,这样就可以把模型的注意力投向那些难分类的样本,一个大的 γ 让获得小损失的样本范围变大。同时,当 $\gamma=0$ 时,Focal Loss 会退化为交叉熵损失函数,不同参数值对应的损失状态如图 2 所示。同样的问题不仅在分类损失上,对于边界框回归损失也存在着大量低质量样本影响回归收敛。

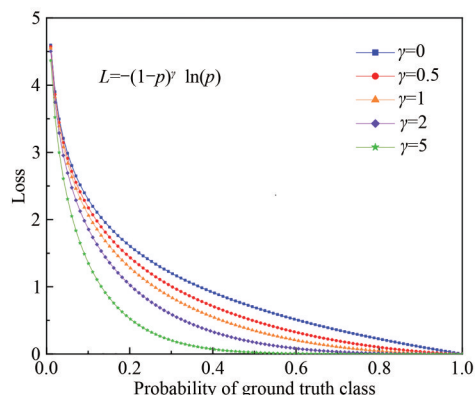


图 2 Focal Loss 不同 γ 的损失变化情况

Fig. 2 Loss changes of different γ of Focal Loss

3 自适应聚焦 CRIoU 损失

3.1 设计思想

IoU 具有尺度不变性的优点。根据 IoU 损失函数进行改进通常是在 IoU 损失函数后加入适当的惩罚项来对 IoU 损失函数进行修正,本文也根据同样的思想进行改进。使用相关性较强的 IoU 损失作为损失函数的基础范式,加入两框中心点距离信息作为 IoU 损失的惩罚项,提高了模型的收敛速度和回归精度。改进设计的公式范式如公式(3)所示:

$$L = 1 - IoU + R(G, B). \quad (3)$$

其中: L 为损失函数, $R(G, B)$ 是改进的 IoU 损失函数中加入的惩罚项, G 表示预测框的参数, B 表示目标框的参数。本文提出的改进 IoU 损失函数根据公式(3),加入带有特殊信息的惩罚项来提升检测效果,并对 GIoU 损失的缺点加以弥补。其中引入的惩罚项之中的特殊信息要包含合适的边界

框的误差信息,如距离、重叠面积等,因此引入的惩罚项要满足如下条件:

(a)损失函数的收敛梯度方向要与两框重叠程度的评价指标相同。若优化的误差信息和重叠评价指标的梯度方向不相同,会造成模型的收敛方向出现偏差,影响回归的性能。

(b)惩罚项数值稳定性。原始 IoU 损失在引入惩罚项后,收敛速度会加快。如果惩罚项的数值稳定性低,会造成惩罚项衰减速度过快,从而损失函数会退化成 IoU 损失,造成后期回归性能下降。

(c)尺度不变性。对于大数据集中的目标进行检测时,需要考虑图像中各个目标的尺度信息。当损失函数只对特定尺度敏感时,会降低模型的整体性能,降低了模型的泛化能力。

3.2 A-CRIoU 损失

根据 2.1 节中的 3 个设计思想,本文提出一种新型 IoU 损失函数,在 IoU 函数基础上引入惩罚项,惩罚项中包含完整的回归属性,使得回归公式满足完整相关的条件。本文选取两边界框的中心点所围成矩形的周长和包围两框的最小外包框的周长的比值作为惩罚项,这个周长比值中包含两框中心点的距离信息,如图 3 中的黑色虚线框长度为 P_1 ,蓝色虚线框长度为 P_2 。设计以广义周长比值为距离信息的损失函数 A-CRIoU (Around-Complete Relativity IoU)如公式(4)所示:

$$L_{A-CRIoU} = 1 - R^{A-CRIoU}(G, B) = 1 - IoU + \frac{P_1}{P_2}, \quad (4)$$

其中: P_1 是表示两框中点所围成矩形的周长, P_2 是最小外包框的周长。当预测框和回归框完全重叠时 $GIoU=A-CRIoU=1$, $GIoU\ LOSS=A-CRIoU\ LOSS=0$;当两框距离无限远时。 $GIoU=A-CRIoU=-1$, $GIoU\ LOSS=A-CRIoU\ LOSS=2$ 。但是在某些情况下,如图 4 所示,真实框的宽高都为 1,预测框宽高的值为 0.5。右侧回归状态明显优于左侧回归状态。此时无论预测框与真实值

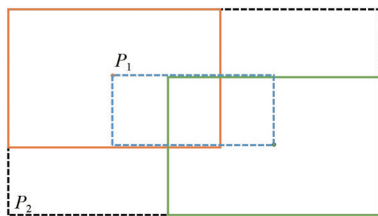


图 3 A-CRIoU 惩罚项示意图

Fig. 3 Schematic diagram of A-CRIoU penalty items

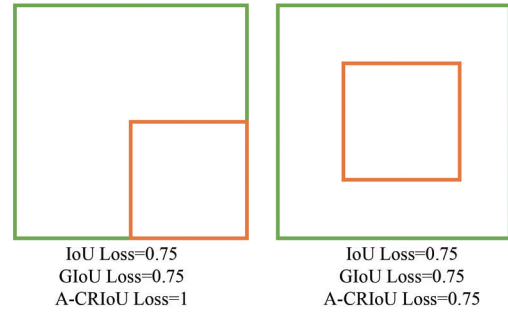


图 4 包含情况下,不同回归状态时损失值的区别。

Fig. 4 Difference of loss in different regression states under inclusion

的中心点距离如何变化, $GIoU$ 与 IoU 的损失值都不变,但是 $A-CRIoU$ 的损失却可以在数值上很好地反映出两框的重叠情况,证明了 $A-CRIoU$ 的数值稳定性。

3.3 A-CRIoU 在 NMS 中的应用

本文将 A-CRIoU 应用到非极大值抑制中,这是因为 A-CRIoU 中的惩罚项中包含中心点距离信息,在 NMS 过程中可以作为更好的信息引入到其中。在图 5 中,A 和 B 分别对应两个正确预测的边界框,且 A 的置信度大于 B。但是在两个不同物体距离很近时,由于 IoU 值比较大,经过 NMS 处理后,只剩下一个检测框,这会导致漏检的错误发生。所以引入中心点距离信息,即使两框 IoU 比较大,但是两个框的中心点距离信息值比较大时,可能会认为这是两个物体的检测框,这种漏选情况时常出现在密集物体检测场景中。使用 A-IoU-NMS 在除 IoU 信息外,引入两框的中心点距离信息,防止了正确预测的框被剔除的情况,减小了漏检概率。A-CRIoU-NMS 的公式如公式(5)所示:

$$S_i = \begin{cases} S_i, IoU - R_{A-CRIoU}(M, B_i) < \epsilon \\ 0, IoU - R_{A-CRIoU}(M, B_i) \geq \epsilon \end{cases}, \quad (5)$$

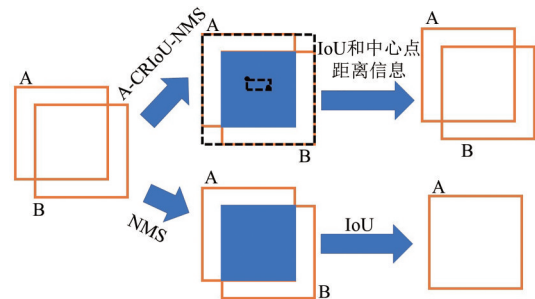


图 5 密集目标检测下 NMS 和 A-CRIoU-NMS 效果对比

Fig. 5 Comparison of NMS and a-criou-nms in dense target

其中: M 为最得分预测框, B_i 是其他预测框, ϵ 是得分的阈值, S_i 为分类得分。

3.4 基于回归属性完整性改进

参考文献[20]中的CIoU损失函数考虑3个几何度量,即重叠面积、中心点距离和宽高比。对于给定的预测框 b 和真实框 b^{gt} ,CIoU损失定义如公式(6)~(8)所示:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{C^2} + \alpha v, \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \quad (7)$$

$$\alpha = \frac{v}{(1 - IoU) + v}, \quad (8)$$

其中:公式(6)中的第三项的分母是边界框中心点距离的平方,分子则是最小外包框的对角线平方。最后一项是宽高比惩罚项, α 是距离因子,可以根据两框的包含程度来自适应调节 v 的大小,对于公式(7)宽高比惩罚项的宽和高的梯度计算如式(9):

$$\begin{cases} \frac{\partial v}{\partial w} = \frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) * \frac{h}{w^2 + h^2} \\ \frac{\partial v}{\partial h} = -\frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) * \frac{w}{w^2 + h^2} \end{cases}, \quad (9)$$

$$\frac{\partial v}{\partial w} = -\frac{h}{w} \frac{\partial v}{\partial h}. \quad (10)$$

在公式(7)中, v 只回归了宽高比值,并没有对真实的宽高值做回归,单纯回归宽高比值会造成真实误差变大的情况,阻碍回归。公式(9)中对于梯度方向可以得到公式(10)。在式(10)中两个梯度方具有相反的符号,所以在任何情况下宽和高其中一个增加,另一个将会减少,这在回归过程中是不合理的。如图1中的CIoU回归过程所示,在最后迭代的预测框的宽高按照比例回归,受尺度不变性影响,宽高真实值在回归中变大,这使得真实的误差值也变大,造成了回归的阻碍。由此本文将实际的宽高平方误差与包围两框最小矩形宽高比值作为惩罚项引入A-CRIoU中,公式(11)是公式(12)后两项的计算方法,最后得到CRIoU损失函数如公式(12)所示:

$$\rho^2 = \frac{x_1^2 - x_2^2}{C^2}, \quad (11)$$

$$L_{CRIoU} = 1 - IoU + \frac{P_1}{P_2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2}. \quad (12)$$

3.5 CRIoU 回归仿真

为验证CRIoU的数值稳定性,本文在以(10, 10)中心点、3为半径的圆中均匀选取5000个点。以(10, 10)为目标框的中点,预测框为以5000个点为中点的锚框。目标框和预测框都具有7个面积大小(0.5、0.67、0.75、1、1.33、1.5和2)和7个尺度(1:4、1:3、1:2、1:1、2:1、3:1和4:1),共有1715000=7×7×7×5000个回归案例,其中回归案例中包含重叠框和非重叠框。对于这些回归点,本文使用梯度下降算法模拟均匀分布的预测框的回归过程。当前预测框 B_i 的坐标通过公式(13)计算:

$$B_i^t = B_i^{t-1} + lr(2 - IoU_i^{t-1}) \nabla B_i^{t-1}, \quad (13)$$

式中: B_i^t 是迭代到 t 轮时的预测框, B_i^{t-1} 是 $t-1$ 轮时的预测框, ∇B_i^{t-1} 是迭代到 t 轮时计算的梯度, lr 是步长。公式中的梯度乘(2-IoU $_i^{t-1}$)用于函数收敛,使用1范数来评估,当到达 $t=200$ 时,终止模拟实验,绘制迭代次数和误差的曲线图如图6所示。由图中曲线可知,CRIoU和CIoU由于包含中心点距离惩罚项,其收敛速度远大于IoU和GIoU,并且CRIoU改善了CIoU在回归中出现的误差增大情况,在后期收敛情况训练中优于CIoU。

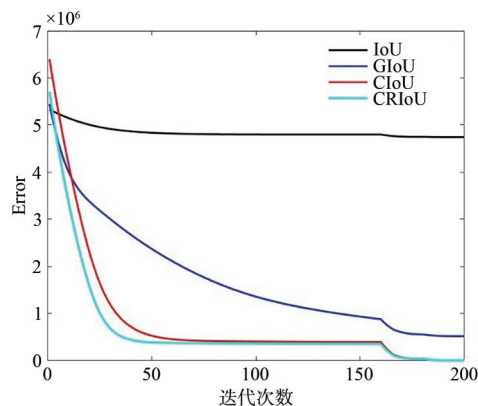


图6 不同损失函数的回归仿真收敛对比

Fig. 6 Comparison of regression simulation convergence for different loss function

图7是仿真实验在最后一轮迭代次数时的误差可视化。由图(a)和(b)可知,对于每一个回归案例 (x, y) ,误差越小对应越良好的回归状态,由此形成了凹陷部分。但在图7(a)未形成明显凹陷部分中可以看到IoU在非重叠情况下有很大的误差,在图7(b)中可以看到GIoU在水平和垂直位

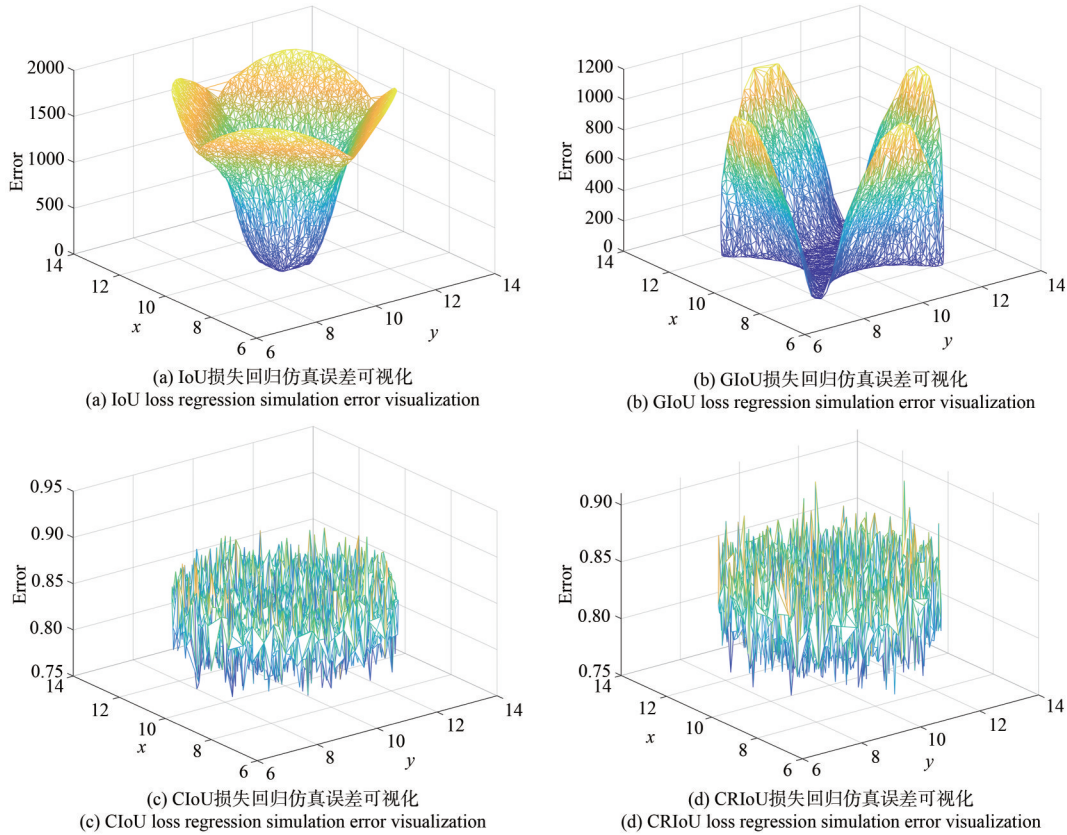


图 7 不同损失函数最后一轮回归误差值分布情况

Fig. 7 Distribution of the last round regression error values of different loss function

置下有很大误差,而在(c)和(d)中CIoU和CRIoU的误差分布比较均匀,在任何地方都有很小的误差,保持较好的回归状态,CRIoU回归误差总体略小于CIoU,但其误差分布没有CIoU稳定。

3.6 自适应聚焦加权因子

目标检测算法在进行回归任务时,存在高质量回归框和低质量回归框不平衡的情况,存在大量的低质量回归框。图8中的anchor机制生成了大量的先验框样本,但是能够检测到目标的高质量



图 8 先验框样本分布情况

Fig. 8 Distribution of anchors

样本数量远小于远离目标的低质量预测框。在模型训练过程中低质量回归框会造成更大的损失值而造成损失的波动,影响模型的收敛。本节基于Focal Loss函数的思想,引入加权因子来平衡低质量样本对损失造成的波动。本文定义加权因子为IoU,因为IoU可以直观反映出在回归任务中样本的质量。IoU值越小证明样本和真实框的重合度越小,证明该样本是一个低质量样本。通过IoU加权来减少大量低质量样本带来的损失波动,使得损失更加稳定,有助于损失收敛。使用IoU作为平衡样本质量的加权因子巧妙地避免了Focal Loss中的加权因子通过每次对比实验的结果进行调制加权因子的弊端,赋予了加权因子自适应聚焦与高质量边界框特性,没有额外的调参开销。最后将自适应加权因子定义为公式(14):

$$\mu = \frac{B^{bb} \cap B^{gt}}{B^{bb} \cup B^{gt}}, \quad (14)$$

其中: B^{bb} 为真实框, B^{gt} 为预测框,它们的IoU为自适应的加权因子 μ 。在训练过程中预测框和真实

框的相交部分越小,预测框的质量越低, μ 值就越小,同时回归损失的权重就越小,这样就会降低大量低质量样本对模型训练的影响。这种加权方法借鉴了 Focal Loss 的聚焦思想,在训练中为高质量的预测框分配较大的权重,使得模型训练聚焦于更有意义的高质量样本,并且加权因子 μ 随着样本质量可以自适应调节,避免了大量的调参开销。由此在 CRIoU 损失的基础上定义自适应聚焦 CRIoU 损失函数(Adaptive Focal CRIoU Focal Loss, AF-CRIoU)如公式(15)所示:

$$L_{AF-CRIoU} = \mu L_{CRIoU}. \quad (15)$$

4 实验结果与分析

4.1 实验平台、数据集和评价指标

为了验证 AF-CRIoU 损失函数的实际有效性,本文实验分别使用单阶段目标检测框架 YOLOv3 和双阶段目标检测模型 Faster R-CNN,并且替换原回归损失函数为 IoU 系列损失函数与 AF-CRIoU 损失函数进行对比。实验均为统一的实验环境,实验硬件环境如表 1 所示,实验软件环境如表 2 所示。实验数据集使用 PASCAL VOC 2012,数据集提供了包括分类、检测和分割 3 个任务的数据集。其中目标检测任务数据集里包含 20 类物体,共 17 125 张图片。

为了验证 AF-CRIoU 损失函数的有效性,实验使用 YOLOv3 和 Faster R-CNN 分别在 PASCAL

VOC2012 上进行训练和测试。通过对比不同模型在不同数据集上的检测精度和收敛速度来验证 AF-CRIoU 的有效性。本文使用平均精度 mAP、AP50 和 AP75 作为评价指标,将 IoU 的阈值设定为 $\{0.5, 0.55, \dots, 0.95\}$,在此阈值范围的结果下除以类别得到 mAP。AP50 是 IoU 阈值为 0.5 时检测各个类别的平均检测精度(AP75 同理),并且沿用参考文献[19]和参考文献[20]的实验评估标准,对比 AF-CRIoU 和其他损失函数的相对提升精度(Relative improv. %, RI)。相对精度计算方法如公式(16)所示:

$$RI = \frac{AP_{AF-CRIoU} - AP_{other}}{AP_{other}}. \quad (16)$$

4.2 PASCAL VOC2012 实验结果

使用 PASCAL VOC 2012 数据集分别计算在 YOLOv3、Faster RCNN 下使用不同损失函数的 AP50 和 AP75,并且对比相对提升精度。如表 3 和表 4 所示。将两个目标检测算法的原来计算边界框回归的损失分别更改为 IoU、GIoU、CIoU、AF-CRIoU,加上模型原有的回归损失函数(Sum of Square Error, SSE),分别计算了 PASCAL VOC 2012 数据集上的 AP75 和 AP50。由表 3 中数据可知, CRIoU 对比原来模型中的损失函数 SSE 在 AP75 和 AP50 条件下精度分别提升 8.52% 和 8.04%;在 AP50 条件下, AF-CRIoU 对比 GIoU 精度检测相对提升 2.46%,在 AP75 条件下的精度相对提升 3.18%;在 AP50 条件下, AF-CRIoU 对比 CIoU 检测精度相对提升 1.67%,在 AP75 条件下的精度相对提升 2.31%。表 4 是在 Faster RCNN 上的实验数据。由表 4 数据可知,在 AP75

表 1 实验硬件环境

Tab. 1 Experimental hardware environment

实验硬件环境	环境配置
核心处理器	Intel Xeon Gold 5220@2.2 GHz
内存容量	256 GB
显卡型号	NVIDIA GeForce RTX 2080Ti

表 2 实验软件环境

Tab. 2 Experimental software environment

实验软件环境	环境配置
服务器系统	Linux 7.6.1810
编程语言	Python 3.7.11
深度学习框架	Pytorch 1.7.1
开发工具	PyCharm 11.0.11; Matlab
CUDA 版本	10.1.243

表 3 不同阈值下 YOLOv3 检测精度对比实验数据

Tab. 3 Comparison of YOLOv3 accuracy under different threshold

Loss function	AP75/%	AP50/%
AF-CRIoU	65.45	59.1
SSE	60.31	54.7
Relative improv.	8.52	8.04
IoU	62.51	56.51
Relative improv.	4.7	4.58
GIoU	63.43	57.68
Relative improv.	3.18	2.46
CIoU	63.97	58.13
Relative improv.	2.31	1.67

表 4 不同阈值下 Faster RCNN 检测精度对比实验数据

Tab. 4 Comparison of Faster RCNN accuracy under different threshold

Loss function	AP75/%	AP50/%
AF-CRIoU	56.61	52.28
Smooth L1 Loss	53.94	48.43
Relative improv.	4.94	7.9
IoU	54.62	48.78
Relative improv.	3.64	7.18
GIoU	54.93	48.93
Relative improv.	3.06	2.29
CIoU	55.21	50.91
Relative improv.	2.53	2.69

和 AP50 条件下, AF CRIoU 对比 GIoU 精度分别提升 3.06% 和 2.29%, AF-CRIoU 对比 CIoU 精度分别提升 2.53% 和 2.69%, 避免了 CIoU 的回归阻碍, 使模型达到一个更好的收敛状态, 同时也解决了 IoU 和 GIoU 的局限。实验结果表明, AF-CRIoU 很好地改善了传统 IoU 系列函数中的回归退化和回归阻碍, 可以达到更好的检测效果。

图 9 是 IoU、GIoU、CIoU 损失函数和 AF-CRIoU 损失函数在 YOLO v3 上使用 PASCAL VOC 2012 数据集的回归损失曲线图。从曲线中可以看出, AF-CRIoU 损失在前期收敛速度上优于 GIoU 损失。这是因为 AF-CRIoU 引入了包含距离信息的周长比值惩罚项, 加快了收敛速度, 可在训练前期保持较为稳定的收敛趋势。通过自适应加权因子可以加强对高质量边界框的学习, 稳定损

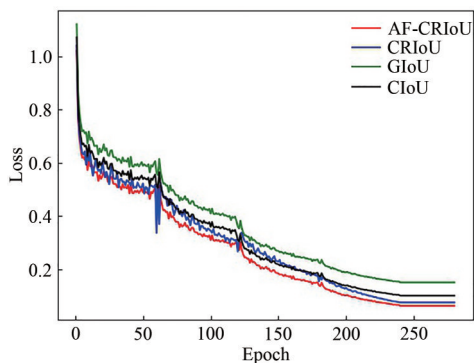


图 9 不同损失函数在 PASCAL VOC2012 上使用 YOLOv3 收敛情况

Fig. 9 YOLOv3 convergence of different loss functions on PASCAL VOC2012

失, 可以使损失收敛到更小, 体现了本文方法的有效性。

4.3 AF-CRIoU 与 A-CRIoU-NMS 消融实验

在 PASCAL VOC 2012 上使用 YOLOv3 的实验中, 带入了 A-CRIoU 作为非极大值抑制处理。表 5 和表 6 是使用 A-CRIoU-NMS 方法和 AF-CRIoU 损失函数对精度提升效果的消融实验结果。表 5 表明, 在 AP75 条件下不加入 AF-CRIoU, 原始的 NMS 方法在选定的阈值范围内的检测精度达到 65.09%, A-CRIoU-NMS 的检测精度最高达到 65.23%, 对比 NMS 提高了 0.14%。使用 AF-CRIoU 后, 原始 NMS 方法检测精度达到 65.45%, 对比没有使用 AF-CRIoU 损失的原始 NMS 方法检测精度提高 0.36%。同理对比只使用 A-CRIoU-NMS 方法的检测精度提高了 0.24%。使用 A-CRIoU-NMS 方法和 AF-CRIoU 损失的情况下, 检测精度达到了最高值。同理, 表 6 中在 AP50 条件下对于不同损失函数和 NMS 方法的消融实验结果中, 也有明显的提升效果。

图 10 和图 11 是在 [0.42:0.47] 范围内在 PASCAL VOC 2012 的检测精度变化趋势。4 种线条分别对应原始 NMS+AF-CRIoU、原始 NMS+原损失、AF-CRIoU+A-CRIoU-NMS 和 A-CRIoU-

表 5 AP75 条件下采用不同非极大值抑制方法和损失函数的实验结果对比

Tab. 5 Comparison of experimental results using different non maximum suppression methods and loss functions under AP75 conditions

NMS	A-CRIoU-NMS	AF-CRIoU	AP75/%
✓			65.09
	✓		65.23
✓		✓	65.45
	✓	✓	65.47

表 6 AP50 条件下采用不同非极大值抑制方法和损失函数的实验结果对比

Tab. 6 Comparison of experimental results using different non maximum suppression methods and loss functions under AP50 conditions

NMS	A-CRIoU-NMS	AF-CRIoU	AP50/%
✓			58.52
	✓		58.90
✓		✓	59.10
	✓	✓	59.13

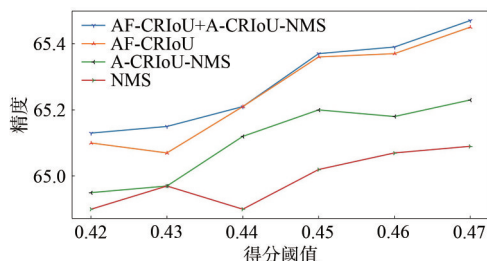


图 10 AP75条件下不同阈值的AF-CRIoU和NMS方法对比

Fig. 10 Comparison of AF-CRIoU and NMS methods with different thresholds under AP75 condition

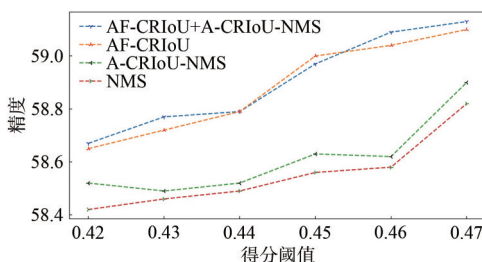


图 11 AP50条件下不同阈值的AF-CRIoU和NMS方法对比

Fig. 11 Comparison of AF-CRIoU and NMS methods with different thresholds under AP50 condition

NMS+原损失。由图中变化趋势可知,在调参过程中,各个阈值下使用A-CRIoU-NMS方法或者AF-CRIoU损失乃至两种方法都使用的情况下,对比原始NMS和回归损失函数,精度都有明显提升。

4.4 基于AF-CRIoU损失的安全帽检测

除了使用大型公共数据集,本文还选取了具有应用背景的安全帽检测数据集进行验证。此数据集为实际应用项目中的私有自建数据集。场景包含工地、矿场,数据集包含9802张安全帽佩戴图像,使用了开源工具Labelimg进行标注。数据集包含两个类别,佩戴安全帽样本标签为hat,未佩戴安全帽样本标签为nohat。数据集按照7:2:1的比例划分成训练集、验证集、测试集。实验环境如表1和表2所示。

实验网络选取yolov3为检测网络,分别替换IoU、GIoU、CIoU、AF-CRIoU为网络的回归损失。训练输入图像尺寸为618×618,使用预训练模型,学习率设置为0.001,batch size设置为64,学习率衰减设置为0.7,每50轮衰减一次,共训练200轮。

检测结果如表7所示。AF-CRIoU在安全帽检测中mAP达到73.43%,比CIoU、GIoU和IoU分别提升1.28%、3.26%和6.42%。

表 7 检测精度对比

Tab. 7 Comparison of detection accuracy

Loss	hat/%	nohat/%	mAP/%
IoU	68.31	65.72	67.01
GIoU	70.47	69.87	70.17
CIoU	72.92	71.38	72.15
AF-CRIoU	74.28	72.57	73.43

图12为后100轮训练的损失收敛示意图。可以看出,当训练进行到第101轮时,AF-CRIoU的收敛速度和收敛状态最好,其次分别为IoU、GIoU、CIoU。但是随着训练轮数的迭代,CIoU和AF-CRIoU的收敛状况要优于IoU和GIoU,这是因为二者都包含边界框中心点距离信息,加快了回归速度,并且AF-CRIoU取消了对宽高比的回归,直接回归宽高的误差值,避免了CIoU回归宽高比时出现回归阻碍的情况,从而可以收敛到一个更好的状态。自适应加权因子的引入,也使AF-CRIoU在训练中受到低质量样本影响最小,从而收敛状态达到最好。

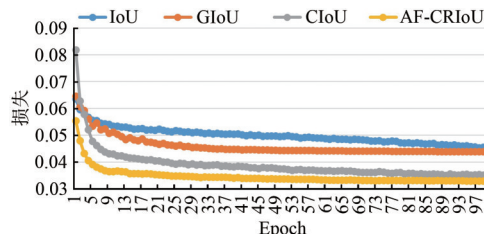


图 12 后 100 轮训练损失收敛情况

Fig. 12 Convergence of loss in the last 100 rounds of training

在安全帽检测数据集实验中,本文同时保存了后100轮训练中的预测框与真实框的IoU来评估AF-CRIoU损失的性能。平均交并比(AVG_IoU)^[27]是在语义分割中的评价指标,通过记录每轮的平均交并比,即每个类别真实框和预测框的平均IoU,根据平均IoU的变化趋势来评估回归质量。在图13中,经过200轮回后,AF-CRIoU损失的平均IoU总体保持在0.9以上,而其他方法的平均IoU均在小于0.9以下的范围波动。这是因

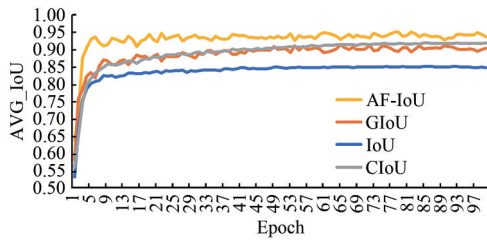


图 13 后 100 轮平均交并比

Fig. 13 Average intersection ratio of the last 100 rounds

为 AF-CRIoU 不但完善了中心点距离属性和及宽高属性,而且通过自适应加权的方式平衡了高低质量样本,降低了大量低质量样本对训练的影响,提高了边界框回归质量,使得预测框和真实框重叠面积更大,贴合程度更好。

图 14~图 16 是随机选取测试集图片的检测效果图。其中红色框为未佩戴安全帽,绿色框为佩戴安全帽,紫色椭圆标记为漏检情况,粉色椭圆标记为误检情况。如图 14 所示 IoU 和 GIoU 虽然正确检测到了目标,但是出现了误检情况,而且也并未完全将目标框选。而 CIoU 和 AF-CRIoU 都精准地检测到了目标。但是对于目标的框选,AF-CRIoU 要优于 CIoU。图 15 中每个损失都检测到了目标,但是 AF-CRIoU 的框选依然要比其他方法精准。图 16 中目标较多,其中包含小目标和密集目标,IoU、GIoU、CIoU 均出现了一到两个目标漏检,而且对于目标的框选质量也很差。AF-CRIoU 虽然框选存在误差,但对于小目标和密集目标都能检测出来。综上结果可知,CIoU 和 AF-CRIoU 二者都完善了回归属性,但是 AF-CRIoU 修改宽高比回归为误差回归,框选更为精准。AF-CRIoU 因为在训练中有自适应加权因

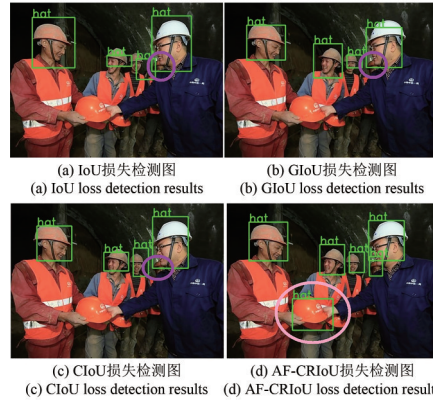


图 15 安全帽检测效果图 2

Fig. 15 Effect diagram 2 of helmet detection

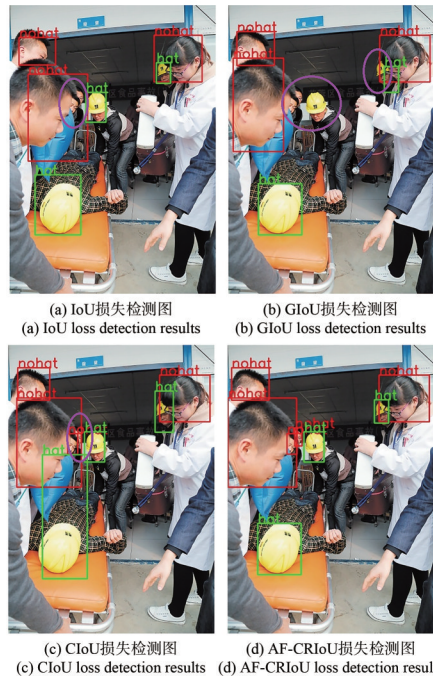


图 16 安全帽检测效果图 3

Fig. 16 Effect diagram 3 of helmet detection

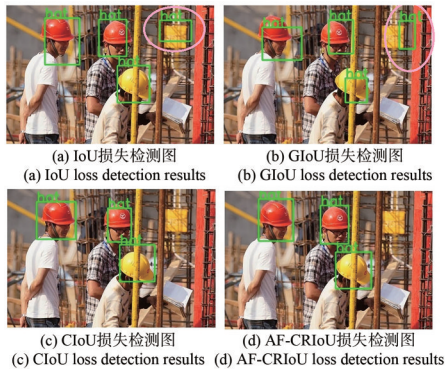


图 14 安全帽检测效果图 1

Fig. 14 Effect diagram 1 of helmet detection

子,使得其受负样本影响更小,可以收敛得更好,从而达到较好的检测精度。

5 结 论

本文提出一种新型交并比的边界框损失函数,通过对 IoU 损失函数的改进,引入带有距离信息和宽高比的惩罚项。以两框中点围成的矩形周长比值为距离惩罚项,避免了 IoU 损失函数在两框不相交情况下损失值无法优化的缺点,防止 GIoU 损失函数在两框包围的特殊情况下退化为 IoU

损失的情景,加快了函数的收敛速度。同时引入宽高误差惩罚项,解决了 CIoU 只回归宽高比时出现的回归阻碍情况,使损失值更加稳定,更加多样地反映边界框回归中的回归状态。结合 Focal Loss 的加权思想解决了回归问题中样本质量不均衡所带来的函数收敛问题。本文将改进的 A-

CRIoU 应用到非极大值抑制处理中,对比 NMS 提升了检测精度。实验结果表明,本文提出的 AF-CRIoU 损失函数在单阶段和双阶段的目标检测算法中起到了提升精度和加快收敛的作用。在 PASCAL VOC 2012 数据集和安全帽检测数据集上均有明显的精度提升,具有一定的实用价值。

参 考 文 献:

- [1] 张吉友,张荣芬,刘宇红,等. 基于注意力机制的多模态图像语义分割[J]. 液晶与显示,2023,38(7):975-984.
ZHANG J Y, ZHANG R F, LIU Y H, *et al.* Multimodal image semantic segmentation based on attention mechanism [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(7): 975-984. (in Chinese)
- [2] 张勇,郭杰龙,汪帆,等. 基于多级联递进卷积结构的图像去雨算法[J]. 液晶与显示,2023,38(10):1409-1422.
ZHANG Y, GUO J L, WANG F, *et al.* Image rain removal algorithm based on multi-cascade progressive convolution structure [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(10): 1409-1422. (in Chinese)
- [3] CHEN L, AI H Z, ZHUANG Z J, *et al.* Real-time multiple people tracking with deeply learned candidate selection and person Re-identification [C]. IEEE International Conference on Multimedia and Expo. San Diego: IEEE, 2018: 1-6.
- [4] DONG J Y, ZHANG L Y, ZHANG H W, *et al.* Occlusion-aware GAN for face de-occlusion in the wild [C]. 2020 IEEE International Conference on Multimedia and Expo (ICME). London: IEEE, 2020: 1-6.
- [5] ZUO C, QIAN J M, FENG S J, *et al.* Deep learning in optical metrology: a review [J]. *Light: Science & Applications*, 2022, 11(1): 39.
- [6] SITU G. Deep holography [J]. *Light: Advanced Manufacturing*, 2022, 3(2): 278-300.
- [7] LUO Y, ZHAO Y F, LI J X, *et al.* Computational imaging without a computer: seeing through random diffusers at the speed of light [J]. *eLight*, 2022, 2(1): 4.
- [8] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [9] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 3431-3440.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc., 2012: 1097-1105.
- [11] YANG Z T, SUN Y N, SHU L, *et al.* 3DSSD: Point-based 3D single stage object detector [C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 11037-11045.
- [12] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 580-587.
- [13] HEARST M A, DUMAIS S T, OSUNA E, *et al.* Support vector machines [J]. *IEEE Intelligent Systems and their applications*, 1998, 13(4): 18-28.
- [14] GIRSHICK R. Fast R-CNN [C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440-1448.
- [15] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [16] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016:

779-788.

- [17] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 65117-6525.
- [18] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J/OL]. *arXiv*, 2018: 1804.02767.
- [19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J/OL]. *arXiv*: 2004.10934, 2020.
- [20] GE Z, LIU S T, WANG F, *et al.* YOLOX: Exceeding YOLO series in 2021 [J/OL]. *arXiv*, 2021: 2107.08430.
- [21] YU J H, JIANG Y N, WANG Z Y, *et al.* UnitBox: An advanced object detection network [C]//*Proceedings of the 24th ACM International Conference on Multimedia*. Amsterdam: ACM, 2016: 516-520.
- [22] REZATOFIGHI H, TSOI N, GWAK J, *et al.* Generalized intersection over union: A metric and a loss for bounding box regression [C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 658-666.
- [23] ZHENG Z H, WANG P, LIU W, *et al.* Distance-IoU loss: Faster and better learning for bounding box regression [C]//*Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*. Washington: AAAI Press, 2020: 12993-13000.
- [24] SERMANET P, EIGEN D, ZHANG X, *et al.* Overfeat: Integrated recognition, localization and detection using convolutional networks [C]. 2nd International Conference on Learning Representations. Banff: ICIR, 2014.
- [25] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single shot multiBox detector [C]. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.
- [26] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 2999-3007.
- [27] TAN M X, CHEN B, PANG R M, *et al.* MnasNet: Platform-aware neural architecture search for mobile [C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 2815-2823.

作者简介:



肖振久(1968—),男,内蒙古宁城人,硕士,副教授,2004年于辽宁工程技术大学获得硕士学位,主要从事图像与视觉信息计算、机器学习、网络与信息安全方面的研究。E-mail: xiaozhenjiu@lntu.edu.cn



郭杰龙(1988—),男,福建泉州人,硕士,工程师,2015年于中南民族大学获得硕士学位,主要从事机器学习、深度学习方面的研究。E-mail: gjl@fjirsm.ac.cn