

基于改进边界框回归损失的YOLOv3检测算法

沈记全, 陈相均, 翟海霞

(河南理工大学 计算机科学与技术学院, 河南 焦作 454000)

摘要: YOLOv3检测算法中的边界框回归损失函数对边界框尺度敏感, 且与算法检测效果评价标准交并比(IoU)之间的优化不具有强相关性, 无法准确反映真值框与预测框之间的重叠情况, 造成收敛效果不佳。针对上述问题, 提出基于IoU的改进边界框回归损失算法BR-IoU。将IoU作为边界框回归损失函数的损失项, 使不同尺度的边界框在回归过程中获得更均衡的损失优化权重。在此基础上, 通过添加惩罚项最小化预测框与真值框中心点间围成的矩形面积, 并提高预测框与真值框之间宽高比的一致性, 从而优化边界框的回归收敛效果。在PASCAL VOC和COCO数据集上的实验结果表明, 在不影响实时性的前提下, BR-IoU能够有效提高检测精度, 采用BR-IoU的YOLOv3算法在PASCAL VOC 2007测试集上mAP较原YOLOv3算法和G-YOLO算法分别提高2.5和1.51个百分点, 在COCO测试集上分别提高2.07和0.66个百分点。

关键词: YOLOv3检测算法; 边界框回归; 交并比; BR-IoU损失算法; 宽高比

开放科学(资源服务)标志码(OSID):



中文引用格式: 沈记全, 陈相均, 翟海霞. 基于改进边界框回归损失的YOLOv3检测算法[J]. 计算机工程, 2022, 48(3): 236-243.

英文引用格式: SHEN J Q, CHEN X J, ZHAI H X. YOLOv3 detection algorithm based on the improved bounding box regression loss[J]. Computer Engineering, 2022, 48(3): 236-243.

YOLOv3 Detection Algorithm Based on the Improved Bounding Box Regression Loss

SHEN Jiquan, CHEN Xiangjun, ZHAI Haixia

(College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, China)

[Abstract] The bounding box regression loss function in the YOLOv3 detection algorithm is sensitive to the bounding box scale but does not have a strong correlation with the optimization of the algorithm detection effect evaluation standard Intersection over Union(IoU). Furthermore, and the loss function cannot accurately reflect the overlap between the ground truth and prediction boxes, resulting in poor convergence effect. In response to these problems, in this study, an improved bounding box regression loss algorithm based on IoU is proposed, namely, BR-IoU. This algorithm adopts IoU as the loss term of the bounding box regression loss function, so that the boundary boxes with different scales can obtain more balanced loss optimization weight in the regression process. On this basis, it adds a penalty item to minimize the area of the rectangle enclosed by the center point of the ground truth and prediction boxes. The added penalty item improves the consistency of the aspect ratio between the ground truth and prediction boxes, which improves regression convergence effect of the bounding box. The experimental results on PASCAL VOC and COCO datasets demonstrate that YOLOv3 using BR-IoU can effectively improve the detection accuracy without affecting the real-time performance. Compared to YOLOV3 and G-YOLO algorithms, mAP value of the proposed algorithm increases by 2.5 and 1.5 percentage points respectively on PASCAL VOC test set and 2.07 and 0.66 percentage points respectively on COCO test set.

[Key words] YOLOv3 detection algorithm; bounding box regression; Intersection over Union (IoU); BR-IoU loss algorithm; aspect ratio

DOI: 10.19678/j.issn.1000-3428.0060479

基金项目: 国家自然科学基金(61972134); 河南省科技攻关项目(182102310946)。

作者简介: 沈记全(1969—), 男, 教授、博士, 主研方向为计算机视觉、智能信息系统; 陈相均, 硕士研究生; 翟海霞, 副教授、硕士。

收稿日期: 2021-01-04 **修回日期:** 2021-03-26 **E-mail:** 1552006123@qq.com

0 概述

目标检测技术是很多计算机视觉任务的基础,如实例分割^[1-3]、图像描述^[4-5]、目标跟踪^[6]等,并且在工业、安防、视频监控、人脸识别^[7]、机器人视觉^[8]、自动驾驶^[9]等诸多领域有极大的研究价值和应用前景,受到了学者们的广泛关注。当前,目标检测的研究已取得较大进展,检测精度及速度在不断提高。但是,由于尺度变化、旋转、遮挡、光照以及物体的稠密度、角度等不同因素的影响,目标检测的精度仍有较大的提升空间。近年来,基于深度学习的目标检测成为研究热点,卷积神经网络可以快速、准确地从大量样本中学习更通用的特征,并且无需对样品进行预处理,避免了复杂的手工制作和设计。

基于深度学习的目标检测算法大致分为两阶段检测算法和单阶段检测算法两种:一种是包括R-CNN^[10]、Fast R-CNN^[11]、Faster R-CNN^[12]和Pyramid Networks^[13]的两阶段检测算法,被称为基于候选区域的目标检测算法,其将目标检测过程分解为候选区域提取、候选区域分类和候选区域坐标修正3个步骤;另一种是包括SSD^[14]、RetinaNet^[15]和YOLO^[16-18]系列的单阶段检测算法,被称为基于回归分析的目标检测算法,其将目标检测问题视为对目标位置和类别信息的回归分析问题,通过一个神经网络模型直接输出检测结果。尽管存在这些不同的检测框架,但无论对于单阶段算法还是两阶段算法,边界框回归都是预测矩形框以对目标对象进行定位的关键步骤。

现有的目标检测算法常用平均绝对误差(Mean Absolute Error, MAE)损失(又称L1范数损失)、均方误差(Mean Square Error, MSE)损失(又称L2范数损失)函数计算边界框位置坐标的回归损失,但是L1、L2范数损失对边界框的尺度具有敏感性,尺度越小的边界框预测偏差对其影响越大,且卷积神经网络在评价边界框的回归效果时使用交并比(Intersection over Union, IoU)作为评价标准,而L1、L2范数损失与IoU之间的优化并非等价。文献[19]提出将IoU作为边界框回归损失函数对检测模型进行优化,但是存在真值框与预测框无交集时优化效果不佳的情况。文献[20]提出一种通用的优化边界框的方式破折号GIoU,解决了在真值框和预测框无交集的情况下将IoU作为边界框损失函数不能反映边界框之间的距离、以及函数梯度为零导致模型无法被优化等问题。文献[21]采用GIoU损失对YOLOv2模型损失函数进行改进,提高了模型对车身焊点的检测精度。文献[22]在YOLOv3网络中引入空间金字塔池化单元,并采用GIoU损失计算坐标损失,在不降低速度的情况下,提高了无人机对罂粟的检测效率。上述损失函数虽然总体上比传统边界框损失函数效果更好,但在某些情况下仍存在收敛效果不佳的问题。

本文提出边界框回归损失算法BR-IoU。将IoU作为边界框回归的损失项,加强损失函数的优化与IoU之间的联系,并添加惩罚项加快真值框与预测框中心点之间的重叠速度,通过使预测框和真值框宽高比保持一致,提升边界框的回归收敛效果。

1 相关工作

1.1 IoU与GIoU

交并比(IoU)也被称为Jaccard系数,用于衡量两个边界框重叠面积的相对大小,是2D和3D目标检测中常用的评价标准。IoU具有尺度不变性、非负性和对称性,不仅可以用来确定正负样本,而且还可以反映模型中预测框的检测效果。如图1所示,A和B为2个重叠的边界框,其Jaccard系数公式如下:

$$I_{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

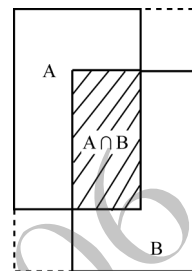


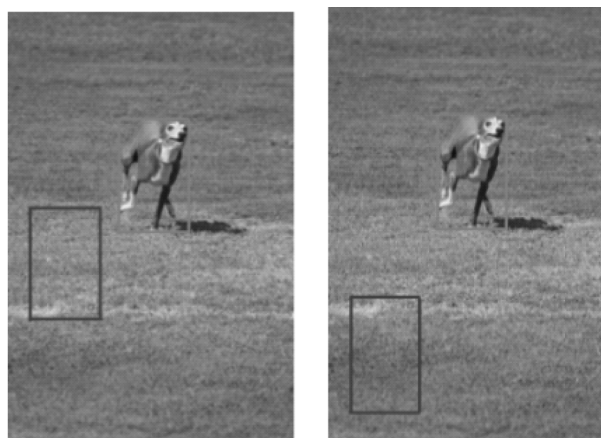
图1 矩形A、B与最小封闭框C

Fig.1 Rectangle A, B and the smallest enclosed frame C

将IoU作为边界框回归损失,其损失函数如下:

$$I_{IoU_loss} = 1 - I_{IoU} \quad (2)$$

虽然IoU作为评价标准具有优势,但是当边框无交集时将IoU作为损失函数会导致无法反映两个边界框彼此之间的距离、损失函数梯度为零不能被优化等问题。如图2所示,图2(a)场景预测框的位置相较于图2(b)场景离真实框更近,但是IoU的值却都为0。



(a)预测框与真实框距离较近的情况 (b)预测框与真实框距离较远的情况

图2 边界框无交集时IoU的比较

Fig.2 IoU comparison when bounding boxes have no intersection

基于此,2019年REZATOF GHI等^[20]提出的GIoU继承了IoU具有的尺度不变性、非负性和对称性,同时克服了IoU在边界框无交集情况下的不足。令图1中包含A和B的最小封闭框为C,则GIoU公式如下:

$$G_{\text{IoU}}(A, B) = I_{\text{IoU}}(A, B) - \frac{|C \setminus (A \cup B)|}{|C|} \quad (3)$$

当A、B两个边界框完全重叠时, $I_{\text{IoU}}(A, B) = 1$; 当A、B两个边界框无交集时, $I_{\text{IoU}}(A, B) = 0$ 。因此, IoU的取值区间为 $[0, 1]$ 。而GIoU的取值区间具有对称性, 即 $-1 \leq G_{\text{IoU}}(A, B) \leq 1$, 与IoU相似: 当两个边界框完全重叠, 即 $|A \cup B| = |A \cap B|$ 时, $G_{\text{IoU}}(A, B) = 1$; 当两个边界框无交集且距离无限大时, 最小包围框C的面积趋近于无穷大, 此时 $G_{\text{IoU}}(A, B) = -1$ 。

$$\lim_{\frac{|A \cup B|}{|C|} \rightarrow 0} G_{\text{IoU}}(A, B) = -1 \quad (4)$$

GIoU取值区间的对称性使其比IoU更能反映边界框的重叠和非重叠情况。本文将GIoU作为边界框回归损失, 其损失函数如下:

$$G_{\text{IoU_loss}} = 1 - G_{\text{IoU}} \quad (5)$$

可以看出, 边界框之间的GIoU值越大, $G_{\text{IoU_loss}}$ 值越小, 边界框的回归收敛效果越好。

1.2 YOLOv3模型

YOLOv3模型是具有代表性的单阶段目标检测算法, 其通过缩放、填充将不同尺寸输入图像的大小调整为32的倍数, 再分成 $S \times S$ 个非重叠的网格, 如图3所示 ($S=7$)。

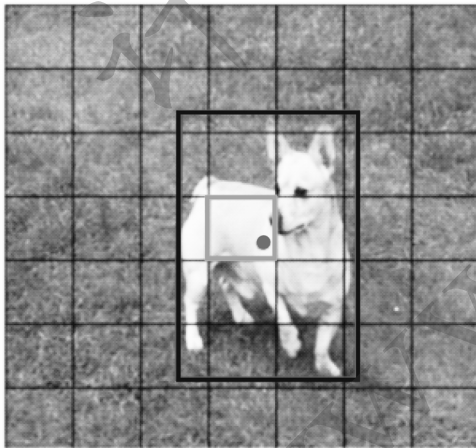


图3 输入图像网格示意图

Fig.3 Grid diagram of input image

每个网格负责检测中心点位置坐标落入该格的物体, 网格中预设了3个边界框和C个类别概率。在

边界框预测的5个分量 x, y, w, h 和置信度中, (x, y) 坐标表示边界框的中心点相对于网格单元的位置, (w, h) 坐标表示边界框的尺寸, 置信度用来判断边界框中存在物体的概率。YOLOv3通过对置信度设定阈值过滤掉低分的边界框, 再对剩下的边界框使用非极大值抑制 (Non-Maximum Supression, NMS) 算法去除冗余边界框, 得到预测结果, 因此, YOLOv3每个网格有 $3 \times (5 + C)$ 个属性值。为了加强对小物体的检测性能, YOLOv3借鉴了特征金字塔网络 (Feature Pyramid Network, FPN) 的思想, 采用多尺度特征融合的方式, 在不同特征层对不同大小的物体进行预测。同时, YOLOv3通过上采样, 自顶向下地融合底层的高分辨率信息和高层的高语义信息, 提高了对不同尺度物体的检测精度, 尤其增强了对小物体的检测精度。

YOLOv3采用L2范数计算边界框坐标的回归损失, L2范数损失中不同尺寸边界框对其预测框坐标产生偏差的敏感度也不同。相对于大尺寸边界框, 尺寸越小的边界框预测偏差对其IoU的影响越大, 因此, YOLOv3在边界框位置损失部分乘以系数 $(2 - t_w \times t_h)$ 来缓解这一情况。为了实现多标签分类, YOLOv3采用二元交叉熵函数取代softmax函数计算边界框的类别损失。softmax函数假设每个对象只属于一个类别, 但是在一些复杂场景下, 一个对象可能属于多个类别 (例如女人和人), 因此, 在目标类别较复杂的数据集上, 多标签分类的二元交叉熵函数能取得更好的预测结果。YOLOv3网络模型的损失函数分为边界框坐标损失、置信度损失和分类损失。损失函数如式(6)所示:

$$\begin{aligned} L_{\text{Loss}} = & b_{\text{bbox_loss}} + c_{\text{conf_loss}} + p_{\text{prob_loss}} = \\ & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (2 - w_i^j \times h_i^j) \times \\ & \left[(x_i^j - \hat{x}_i^j)^2 + (y_i^j - \hat{y}_i^j)^2 + (w_i^j - \hat{w}_i^j)^2 + (h_i^j - \hat{h}_i^j)^2 \right] - \\ & \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[\hat{C}_i^j \ln(C_i^j) + (1 - \hat{C}_i^j) \ln(1 - C_i^j) \right] - \\ & \lambda_{\text{nobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{nobj}} \left[\hat{C}_i^j \ln(C_i^j) + (1 - \hat{C}_i^j) \ln(1 - C_i^j) \right] - \\ & \sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \left[\hat{P}_i^j \ln(P_i^j) + (1 - \hat{P}_i^j) \ln(1 - P_i^j) \right] \quad (6) \end{aligned}$$

其中: I_{ij}^{obj} 、 I_{ij}^{nobj} 表示第 i 个网格中的第 j 个先验框是否存在待检测物体, 若存在, 则 I_{ij}^{obj} 值为1, I_{ij}^{nobj} 值为0, 否则 I_{ij}^{obj} 值为0, I_{ij}^{nobj} 值为1。YOLOv3算法检测过程如图4所示。

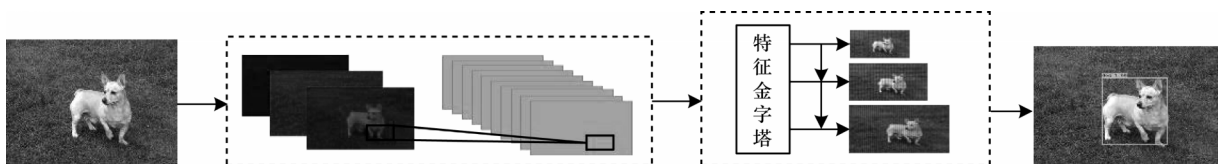


图4 YOLOv3算法检测过程

Fig.4 Detection process of YOLOv3 algorithm

2 BR-IoU 边界框回归损失算法

采用L2范数边界框回归损失函数的YOLOv3检测算法通常会出现以下2个问题:

1)如图5所示,4组边界框角点坐标距离的L2范数损失值相等,但IoU值却不同,这表明L2范数损失不能准确地反映IoU值的变化,与IoU之间不具有强相关性。

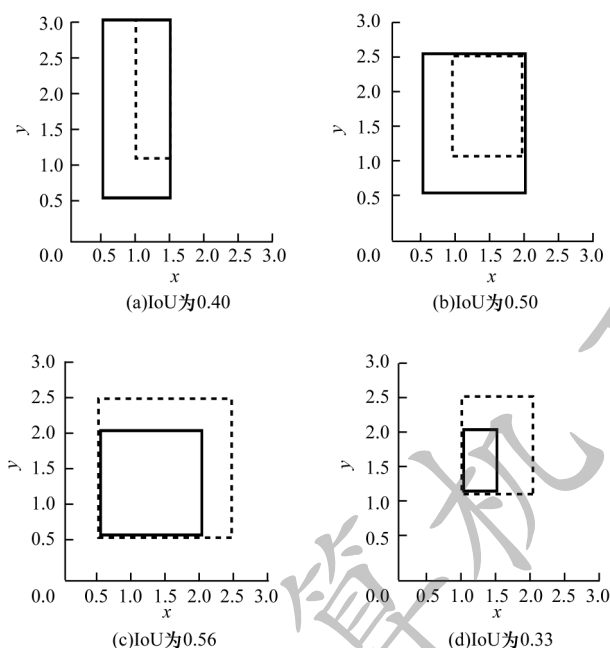


图5 L2范数损失值相等时IoU的变化

Fig.5 Change of IoU when L2 norm loss values are equal

2)L2范数损失具有尺度敏感性,在算法训练过程中,当两个边界框的IoU值相同时,大尺寸边界框会产生更多的损失值使小尺寸边界框难以被优化,导致算法对小尺寸目标的检测效果不佳。

针对上述问题,本文提出BR-IoU算法,为方便不同边界框损失算法的性能对比,将BR-IoU算法分为BR-IoU-A算法和BR-IoU-B算法。

BR-IoU-A算法将IoU作为边界框回归的损失项,强化与IoU之间的联系,并继承了IoU的尺度不变性,使得不同尺度的边界框在回归过程中获得更均衡的损失优化权重。然后算法在IoU损失项的基础上添加一个惩罚项,通过最小化预测框与真值框中心点围成矩形的面积,加快预测框中心点与真值框中心点的重叠速度。算法的惩罚项使边界框在任何情况下都存在梯度,在边界框无交集的情况下仍然能够反映边界框间的相对距离,为边界框的优化提供移动方向,如图6所示,D是真值框 B^g 的中心点 b^g 和预测框B的中心点b围成的矩形框, B^c 是真值框 B^g 与预测框B的最小封闭框, w 和 h 为D的宽和高, w^c 和 h^c 为 B^c 的宽和高,则BR-IoU-A计算公式如下:

$$B_{BR-IoU-A} = I_{IoU}(B^g, B) - \frac{w \times h}{w^c \times h^c} \quad (7)$$

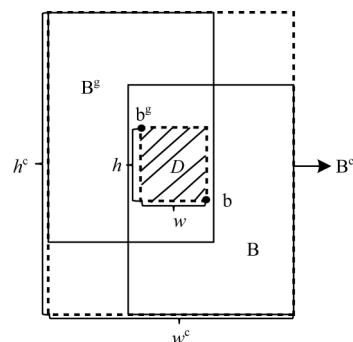


图6 BR-IoU-A算法示意图

Fig.6 Schematic diagram of BR-IoU-A algorithm

利用BR-IoU-A算法计算边界框回归损失,损失函数如下:

$$B_{BR-IoU-A_loss} = 1 - I_{IoU}(B^g, B) + \frac{w \times h}{w^c \times h^c} \quad (8)$$

BR-IoU-A损失继承了IoU和GIoU损失的优良属性:1)BR-IoU-A损失对边界框的尺度不敏感,具有尺度不变性;2)BR-IoU-A损失在边界框无交集情况下仍然存在损失梯度,可以为边界框的优化提供移动方向;3)当两个边框完全重叠时, $G_{GIoU} = B_{BR-IoU-A} = 1$, $G_{GIoU_loss} = B_{BR-IoU-A_loss} = 0$;当两个边框距离无限远时, $G_{GIoU} = B_{BR-IoU-A} = -1$, $G_{GIoU_loss} = B_{BR-IoU-A_loss} = 2$ 。但在某些情景下,如图7所示(其中,实线框为真值框,虚线框为平面测框),真值框宽和高的值都为1,预测框宽和高的值为0.5。此时无论预测框与真值框的中心点间的距离如何变化,GIoU与IoU损失值都不变,由此可见,BR-IoU-A损失算法能很好地区分边界框之间的重叠情况。

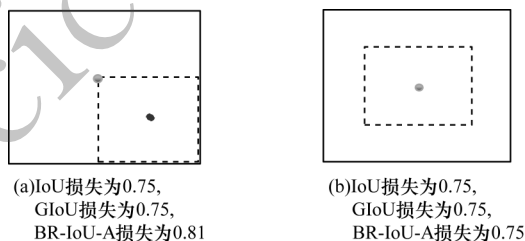


图7 中心点间相对位置比较

Fig.7 Comparison of relative positions between center points

图7(a)图像的检测效果优于图7(b)图像, BR-IoU-A损失算法通过值的变化来反映这种情况。

算法 BR-IoU-A边界框回归损失算法

输入 预测框B和真实框 B^g ,坐标分别为 (x_1, y_1, x_2, y_2) 和 $(x_1^g, y_1^g, x_2^g, y_2^g)$

输出 $B_{BR-IoU-A_loss}$

1.确保预测框B中 $x_2 > x_1, y_2 > y_1$:

$\hat{x}_1 = \min(x_1, x_2), \hat{x}_2 = \max(x_1, x_2)$

$\hat{y}_1 = \min(y_1, y_2), \hat{y}_2 = \max(y_1, y_2)$

$A^g = (x_2^g - x_1^g) \times (y_2^g - y_1^g)$

3.计算预测框B的面积:

$A = (\hat{x}_2 - \hat{x}_1) \times (\hat{y}_2 - \hat{y}_1)$

4.计算预测框B与真值框 B^g 的重叠面积I:

$$x_1^l = \max(\hat{x}_1, x_1^g), x_2^l = \min(\hat{x}_2, x_2^g)$$

$$y_1^l = \max(\hat{y}_1, y_1^g), y_2^l = \min(\hat{y}_2, y_2^g)$$

$$I = \begin{cases} (x_2^l - x_1^l) \times (y_2^l - y_1^l), & x_2^l > x_1^l, y_2^l > y_1^l \\ 0, & \text{otherwise} \end{cases}$$

$$5. I_{IoU} = \frac{I}{u}, \text{其中 } u = A + A^g - I$$

6. 求最小封闭框 B^c 的坐标:

$$x_1^c = \min(\hat{x}_1, x_1^g), x_2^c = \max(\hat{x}_2, x_2^g)$$

$$y_1^c = \min(\hat{y}_1, y_1^g), y_2^c = \max(\hat{y}_2, y_2^g)$$

7. 计算最小封闭框 B^c 的宽和高:

$$w^c = x_2^c - x_1^c, h^c = y_2^c - y_1^c$$

8. 计算预测框 B 和真值框 B^g 的中心点坐标 b 和 b^g :

$$b = \left(\frac{\hat{x}_1 + \hat{x}_2}{2}, \frac{\hat{y}_1 + \hat{y}_2}{2} \right), b^g = \left(\frac{x_1^g + x_2^g}{2}, \frac{y_1^g + y_2^g}{2} \right)$$

9. 计算预测框 B 与真值框 B^g 中心点围成的矩形框 D 的坐标:

$$x_1^D = \min\left(\frac{\hat{x}_1 + \hat{x}_2}{2}, \frac{x_1^g + x_2^g}{2}\right)$$

$$x_2^D = \max\left(\frac{\hat{x}_1 + \hat{x}_2}{2}, \frac{x_1^g + x_2^g}{2}\right)$$

$$y_1^D = \min\left(\frac{\hat{y}_1 + \hat{y}_2}{2}, \frac{y_1^g + y_2^g}{2}\right)$$

$$y_2^D = \max\left(\frac{\hat{y}_1 + \hat{y}_2}{2}, \frac{y_1^g + y_2^g}{2}\right)$$

10. 计算矩形框 D 的宽和高:

$$w = x_2^D - x_1^D, h = y_2^D - y_1^D$$

$$11. B_{BR-IoU-A} = I_{IoU} - \frac{w \times h}{w^c \times h^c}, B_{BR-IoU-A_loss} = 1 - B_{BR-IoU-A}$$

在 BR-IoU-A 算法中, $\frac{w \times h}{w^c \times h^c}$ 是 BR-IoU-A 损失

在 IoU 损失的基础上添加的惩罚项, 由于最小封闭框 B^c 的面积大于真值框 B^g 的面积, 因此 $\frac{w \times h}{w^c \times h^c}$ 的分

母 $w^c \times h^c$ 始终为正非零数且 $0 \leq \frac{w \times h}{w^c \times h^c} \leq 1$ 。IoU 和

GIoU 损失仅考虑到边界框的重叠区域, 而 BR-IoU-A 损失同时考虑到边界框重叠区域和中心点间的相对位置。然而, 预测框与真值框宽高比的一致性也是评价检测效果的重要几何因素, 如图 8 所示(其中, 实线框为真值框, 虚框为预测框), 真值框的宽和高的值都为 4, 左侧预测框的宽和高的值分别为 4 和 0.25, 右侧预测框宽和高的值都为 1。

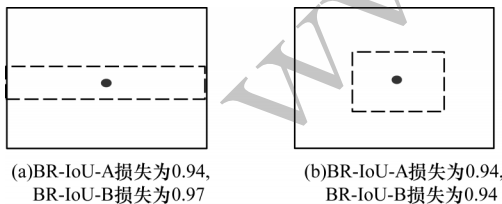


图 8 预测框不同宽高比下的检测效果对比

Fig.8 Comparison of detection effects under different aspect ratios of predicted bounding boxes

图 8(a) 预测框的宽高比值相较于图 8(b) 预测框更接近真实框, 预测效果更好, 但 BR-IoU-A 值却相同。因此, 本文在 BR-IoU-A 的基础上, 结合预测框

与真值框之间宽高比的一致性提出 BR-IoU-B 损失算法, 计算公式如下:

$$B_{BR-IoU-B} = B_{BR-IoU-A} - V \times C \quad (9)$$

其中, 参数 V 使预测框宽高比与真值框的保持一致, C 是 V 的平衡参数, 当两个边界框无交集或 IoU 值较低时, 使边界框之间的重叠因子在回归过程中获得较高优先级, 从而加快边界框回归的收敛速度。

$$V = \left(\frac{w^g}{h^g + w^g} - \frac{w^p}{h^p + w^p} \right)^2$$

$$C = \frac{V}{(1 - I_{IoU}) + V} \quad (10)$$

其中: w^g, h^g 分别为真值框的宽和高; w^p, h^p 分别为预测框的宽和高。综上, BR-IoU-B 损失函数如下:

$$B_{BR-IoU-B_loss} = 1 - B_{BR-IoU-A_loss} + V \times C \quad (11)$$

使用 BR-IoU 边界框回归损失算法对 YOLOv3 模型进行改进, 改进后的损失函数如下:

$$L_{Loss} = b_{bbox_loss} + c_{conf_loss} + p_{prob_loss} =$$

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (2 - w_i^j \times h_i^j) \times (1 - B_{BR-IoU}) -$$

$$\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[\hat{C}_i^j \ln(C_i^j) + (1 - \hat{C}_i^j) \ln(1 - C_i^j) \right] -$$

$$\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} \left[\hat{C}_i^j \ln(C_i^j) + (1 - \hat{C}_i^j) \ln(1 - C_i^j) \right] -$$

$$\sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} \left[\hat{P}_i^c \ln(P_i^c) + (1 - \hat{P}_i^c) \ln(1 - P_i^c) \right] \quad (12)$$

3 实验结果与分析

本文实验模型基于 Tensorflow1.13.1 框架, 编程语言为 python3.7, 实验操作系统为 Ubuntu16.04, GPU 型号为 NVIDIA Tesla P40, CUDA 版本为 10.0。为了验证 BR-IoU 算法的实际性能, 将原始 YOLOv3 的边界框回归损失部分分别替换成 GIoU 损失算法(简称 G-YOLO)、BR-IoU-A 损失算法(简称 BR-YOLO-A)、BR-IoU-B 损失算法(简称 BR-YOLO-B)。在 PASCAL VOC 2007+2012 数据集和 COCO 2014 数据集上进行模型评测和对比实验。

在算法训练过程中, 为了增强模型性能, 对数据集中的原始图像进行数据增强操作, 通过图像随机裁剪、旋转、平移、颜色变化等方法, 增加图像的多样性, 使神经网络具有更强的泛化效果, 提高模型鲁棒性。此外, 为检测不同尺度图像中包含的不同尺度和形状的目标, 在每个检测层中设定不同尺度的先验框与真实标签进行匹配。网络输入大小为 416×416 , 模型参数更新方式为 Adam, 初始学习率为 0.000 1, 权重衰减设置为 0.000 5。对于 VOC 和 COCO 数据集, 模型加载预训练权重来完成初始化, 令其他没有预训练权重的部分在训练过程中自适应微调参数, 然后训练整个模型。为了在训练过程中使模型稳定, 算法为模型训练设置两个轮次迭代的热身阶段。实验采用平均精度(Average Precision, AP)反映每一类目标的检测效果, 平均精度是从准

确率和召回率两个方面来衡量检测算法的准确性,可以直观地表现模型对单个类别的检测效果;采用平均精度均值(mean Average Precision,mAP)来衡量多类目标的平均检测精度,mAP值越高,模型在全部类别中综合性能越高。

3.1 在PASCAL VOC数据集上的对比实验

PASCAL VOC 2007+2012训练集共包含 16 551 张

图像、40 058 个样本框,分为 4 个大类和 20 个小类。VOC 2007 测试集包含 4 952 张图像、12 032 个样本框。PASCAL VOC 通过计算数据集 IoU=0.5 时的 mAP 对模型进行评价。该组实验在 PASCAL VOC 数据集上的训练迭代次数为 100 轮,测试集上的 mAP 及各类别目标的 AP 检测结果如图 9 和表 1 所示。

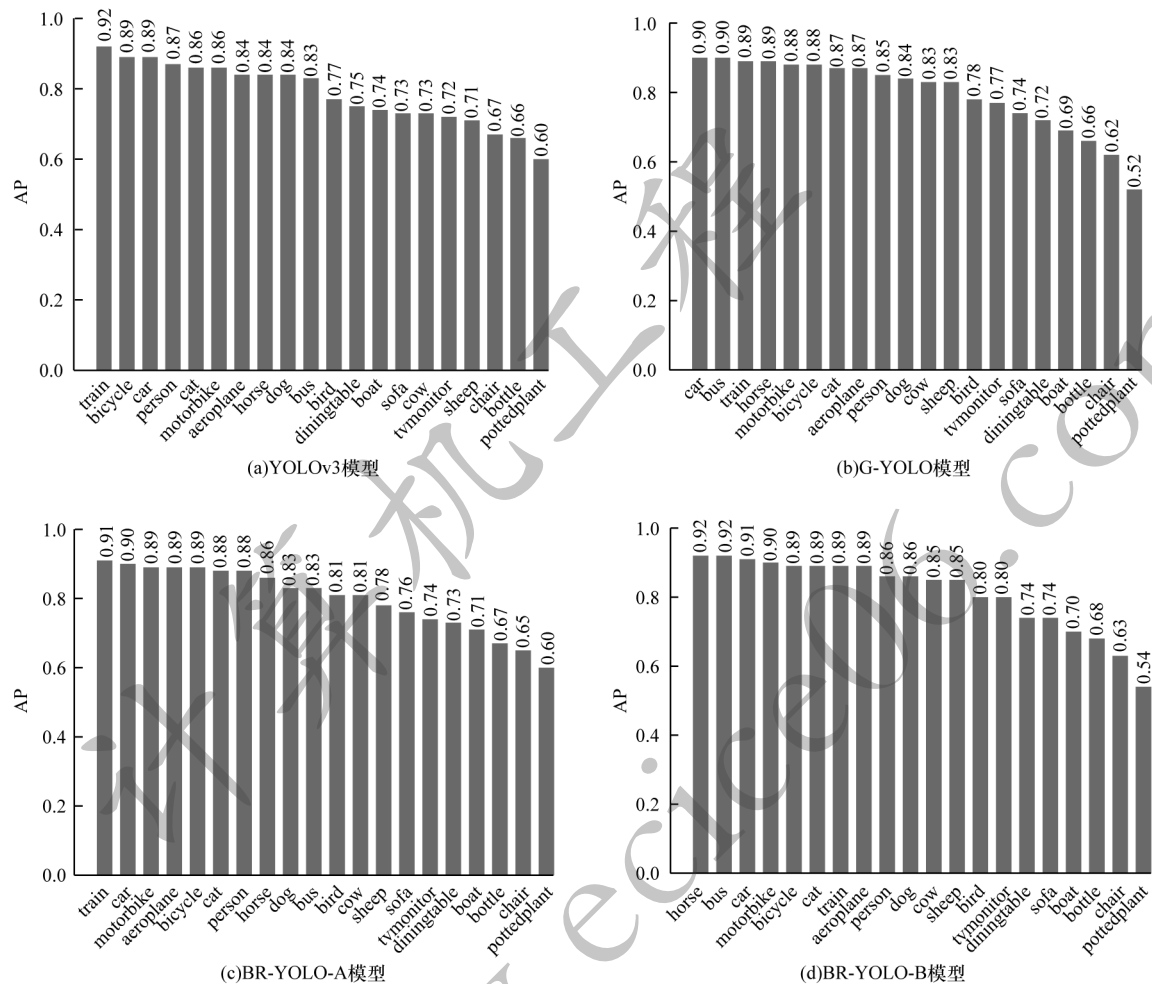


图9 PASCAL VOC数据集上不同算法的AP对比
Fig.9 AP comparison of different algorithms on PASCAL VOC data set

表1 PASCAL VOC数据集上不同算法的性能对比
Table 1 Performance comparison of different algorithms on PASCAL VOC data set

检测算法	训练集	mAP/%	检测速度/(frame·s ⁻¹)
Fast R-CNN	2007+2012	70.0	0.5
Faster R-CNN (VGG-16)	2007+2012	73.2	7.0
Faster R-CNN (ResNet)	2007+2012	76.4	5.0
SSD300	2007+2012	74.3	46.0
SSD500	2007+2012	76.8	19.0
YOLOV1	2007+2012	63.4	45.0
YOLOv2	2007+2012	76.8	67.0
YOLOv3	2007+2012	78.7	26.8
G-YOLO	2007+2012	79.7	28.3
BR-YOLO-A	2007+2012	80.0	27.9
BR-YOLO-B	2007+2012	81.2	27.2

实验结果表明,在不影响模型检测速度的情况下,BR-YOLO-B算法的 mAP 高出原 YOLOv3 算法 2.54 个百分点,高出 G-YOLO 算法 1.51 个百分点。

3.2 在COCO数据集上的对比实验

COCO 2014 训练集共包含 82 783 张图像,分为 80 个类别。从验证集中选取 5 000 张图像作为测试集对模型进行评价。COCO 数据集将 IoU 在 [0.5:0.95] 区间内,步长为 0.05 的 10 个取值分别计算 mAP 并取平均值作为检测模型的评价结果,相比 VOC 数据集,COCO 数据集的多 IoU 评价标准对检测算法的要求更高,更能反映检测算法的综合性能。该组实验在 COCO 2014 数据集上的迭代次数为 120 轮,实验结果如图 10 和表 2 所示。可以看出,后 3 种模型相对 YOLOv3 的检测效果均

有所提高,这得益于BR-IoU-A损失算法相比L2损失算法对不同尺度边界框的回归优化更均衡,并且与评价指标IoU之间的优化更加密切,模型BR-YOLO-A算法的mAP对比YOLOv3提高了

1.63个百分点。此外,由于BR-IoU-B算法进一步对预测框与真值框之间宽高比的一致性进行优化,因此BR-YOLO-B算法的mAP比YOLOv3提高了2.07个百分点。



图10 PASCAL VOC数据集上的检测效果对比
Fig.10 Comparison of detection effects on PASCAL VOC data set

表2 COCO 2014数据集上不同算法的mAP对比

算法	Table 2 mAP comparison of different algorithms on COCO 2014 data set										%
	IoU为0.50	IoU为0.55	IoU为0.60	IoU为0.65	IoU为0.70	IoU为0.75	IoU为0.80	IoU为0.85	IoU为0.90	IoU为0.95	
YOLOv3	55.13	53.32	48.61	44.02	39.16	32.28	21.62	10.76	5.09	0.57	31.06
G-YOLO	56.31	54.47	50.24	45.83	41.03	34.79	23.42	12.02	5.84	0.71	32.47
BR-YOLO-A	55.97	54.22	50.06	46.15	41.83	35.13	24.53	12.63	5.67	0.68	32.69
BR-YOLO-B	56.34	55.19	50.63	46.79	42.52	35.57	25.06	12.70	5.77	0.73	33.13

表3所示为4种检测算法的空间与时间复杂度比较结果,其中时间复杂度为模型检测每张图像所需要的时间。由于模型的骨干网络都为Darknet53,因此4种算法的网络参数数量基本相同,约为 236×10^6 。由表3可以看出,相比YOLOv3算法,BR-YOLO-A和BR-YOLO-B算法检测速度更快。

表3 不同算法的时空复杂度对比

Table 3 Time and space complexity comparison of different algorithms		
算法	空间复杂度/ 10^6	时间复杂度/ms
YOLOv3	236	67.6
G-YOLO	236	56.3
BR-YOLO-A	236	59.1
BR-YOLO-B	236	63.8

4 结束语

针对YOLOv3的L2范数边界框回归损失函数对边界框的尺度敏感,以及与算法检测效果的评价标准IoU之间相关性不强等问题,本文提出BR-IoU算法,通过将IoU作为损失项强化边界框回归与算法评估标准的相关性,最小化真值框和预测框中心点形成矩形的面积,同时提高预测框与真值框之间宽高比的一致性,以加强边界框的回归收敛性能。实验结果表明,BR-IoU能够有效提高YOLOv3算法的检测精度和检测速度。下一步将结合本文算法设计思想对其他目标检测算法的边界框回归损失函数进行改进,在不影响检测速度的情况下达到更高的检测精度。

参考文献

[1] HARIHARAN B, ARBELAEZ P, GIRSHICK R, et al. Simultaneous detection and segmentation[C]//Proceedings of European Conference on Computer Vision. Berlin,

- Germany; Springer, 2014; 297-312.
- [2] DAI J F, HE K M, SUN J. Instance-aware semantic segmentation via multi-task network cascades [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016; 3150-3158.
- [3] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2017; 2980-2988.
- [4] KARPATY A, FEI-FEI L. Deep visual-semantic alignments for generating image descriptions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 664-676.
- [5] WU Q, SHEN C, WANG P, et al. Image captioning and visual question answering based on attributes and external knowledge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(6): 1367-1381.
- [6] KANG K, LI H S, YAN J J, et al. T-CNN; tubelets with convolutional neural networks for object detection from videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2896-2907.
- [7] DENG J, GUO J, XUE N, et al. ArcFace: additive angular margin loss for deep face recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2019; 4685-4694.
- [8] 崔坤坤,樊绍胜. 基于动态双窗口的机器人视觉导航与特征识别方法[J]. 计算机工程, 2020, 46(9): 313-320. CUI K K, FAN S S. Visual navigation and feature recognition method of robot based on dynamic double windows[J]. Computer Engineering, 2020, 46(9): 313-320. (in Chinese)
- [9] CHEN X Z, MA H M, WAN J, et al. Multi-view 3d object detection network for autonomous driving[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2017; 6526-6534.
- [10] 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. Washington D. C. , USA; IEEE Press, 2014; 580-587.
- [11] GIRSHICK R. Fast R-CNN [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2015; 1440-1448.
- [12] REN S, HE K, 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.
- [13] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2017; 936-944.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of European Conference on Computer Vision. [S. l.]: Eurographics Association Press, 2016; 21-37.
- [15] LIN T, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA; IEEE Press, 2017; 2999-3007.
- [16] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2016; 779-788.
- [17] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2017; 6517-6525.
- [18] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2019-11-02]. <https://arxiv.org/abs/1804.02767>.
- [19] 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. New York, USA; ACM Press, 2016; 516-520.
- [20] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2019; 658-666.
- [21] 何智成,王振兴. 基于改进YOLOv2的白车身焊点检测方法[J]. 计算机工程, 2020, 46(11): 246-254. HE Z C, WANG Z X. Welding spot detection method for body in white based on improved YOLOv2[J]. Computer Engineering, 2020, 46(11): 246-254. (in Chinese)
- [22] ZHOU J, TIAN Y C, YUAN C, et al. Improved UAV opium poppy detection using an updated YOLOv3 model[J]. Sensors, 2019, 19(22): 4851.