

基于优化边界框回归的目标检测^①

赵文仓^② 徐长凯 王春鑫

(青岛科技大学自动化与电子工程学院 青岛 266061)

摘要 目标检测是计算机视觉任务中的关键问题之一,而边界框回归是目标检测的重要步骤。现有方法认为交并比(IoU)损失有利于度量 IoU,但仍然存在收敛速度慢和回归不准确的问题。此外,大规模的目标检测数据集尽可能地清晰定义真实边界框,但在标记边界框时仍然会引入偏差。本文基于 CIoU 损失和边界框回归 Kullback-Leibler(KL) 损失提出了优化 IoU(OIoU)边界框损失,主要通过学习边界变换和定位方差解决上述问题。OIoU 不仅用于评估本地化置信度还可以计算标准差和边界框位置,在解决边界框标注时带来的偏差、回归不准确问题的同时加快了收敛速度。在 PASCAL VOC 和 MS-COCO 两个数据集上证明了优化边界框损失在检测精度和收敛速度上要优于原有的方法。

关键词 目标检测; 优化边界框回归; Kullback-Leibler(KL) 损失; faster R-CNN

0 引言

由于边界框回归关乎物体定位的准确性,所以众多的计算机视觉任务例如目标定位^[1]、多目标检测^[2-4]、目标跟踪^[5]、语义分割和实例分割^[6-7]的精确度都要依赖于边界框回归。然而利用深度神经网络来提高精确度方法的主要趋势是提出更好的网络架构或者更好的方法来提取可靠的局部特征。交并比(intersection over union, IoU)^[8-10]是最常用的度量标准,用于比较任意形状的相似性,IoU 对所考虑问题的规模不变。由于这个特性,所有用于评估分割、对象检测和跟踪的性能都依赖这个度量。然而 IoU 作为度量存在两个主要问题:(1)如果两个对象不重叠,IoU 的值将为 0,并且不能反映两个形状之间的距离。在非重叠的情况下,如果使用 IoU 作为损失,则梯度为 0,无法优化;(2)IoU 不能正确区分两个对象的不同对齐。确切地说,对于两个具有相同交点水平的不同方向重叠的现象,IoU 将会是完全相等的,所以 IoU 的值并不能反映两个对象是如何

重叠的。为了解决这些问题,Hamid 等人^[11]将 IoU 的概念扩展到非重叠的情况下解决上述 IoU 存在的两个缺陷。IoU 的广义版本为 GIoU(generalized-IoU),其优点一方面当两个对象不重合时依然可以进行优化,另一方面 GIoU 的值更能反映两个形状之间是如何重叠的。虽然上述方法 IoU 和 GIoU 有利于评估 IoU 的指标,但是仍然存在一些问题,即收敛速度缓慢和回归不准确。为了解决这些问题,Zheng 等人^[12]提出了一种结合预测框和目标框之间距离的 DIoU(distance-IoU)损失,DIoU 考虑了重叠面积和中心点距离达到了不错的效果。然而,长宽比这个重要因素没有被考虑进来,因此研究人员又提出了 CIoU(complete-IoU)损失,在原来 DIoU 损失的基础上引入了长宽比的一致性。

边界框回归还存在另一个重要的问题,即像 MS-COCO^[13]、PASCALVOC^[14]等目标检测数据集需要清晰地定义真实边界框。然而在某些情况下真实边界框本质上是模糊的,模糊性使得边界框回归函数的标注困难、学习困难。而当下最先进的目标检

^① 国家自然科学基金(61171131),山东省重点研发计划(YD01033)和国家留学基金委项目(201608370049)资助。

^② 男,1973 年生,博士,教授;研究方向:模式识别与智能系统;联系人,E-mail: zhao_coinslab@outlook.com

(收稿日期:2020-07-16)

测框架依靠边框回归来定位目标。但是传统的边界框回归损失并没有考虑真实边界框存在的模糊性问题。为了解决这些问题,He 等人^[15]提出了 Kullback-Leibler(KL) 损失。本文综合考虑传统的 IoU 损失存在收敛速度慢、回归不准确的问题和真实的边界框模糊性,借鉴 CIoU 损失和 KL 损失提出了一种新的优化边界框回归框架,利用 KL 损失解决了当真实边界框存在模糊性时降低检测性能的问题;CIoU 损失考虑了边界框中心点距离和长宽比的信息使得边界框回归的速度和精度得到进一步提升。本文方法即优化交并比(optimize-IoU, OIoU)在 PASCAL VOC 数据集上将检测精度提升了 1.1%,在 MS-COCO 数据集上提升了 2.3%,进一步提升了检测性能。

1 IoU 损失

1.1 距离 IoU 损失

一般来说,基于 IoU 的损失可以定义为

$$L = 1 - IoU + R(B, B^{gt}) \quad (1)$$

其中, $R(B, B^{gt})$ 是预测框 B 和目标框 B^{gt} 的惩罚项。通过设计适当的惩罚条件来解决之前 IoU 存在的问题。为了解决传统 IoU 和 GIoU 存在收敛慢或者不能很好地收敛从而导致检测结果不准确的问题, DIoU 损失在 IoU 损失的基础上加入一个惩罚项, 以最小化两个边界框中心点之间的规范化距离, 从而比 GIoU 损失更快地收敛。DIoU 在只有 130 次迭代的情况下, 预测框与目标框可以完全匹配, 然而 GIoU 即使经过 400 次的迭代也不会很好地收敛。DIoU 损失最小化两个边界框中心点之间的规范化距离, 惩罚项定义为

$$R_{DIoU} = \frac{h^2(b, b^{gt})}{d^2} \quad (2)$$

其中 b 和 b^{gt} 表示 B 和 B^{gt} 的中心点, h 是欧几里得距离, d 是包含这两个边界框的最小边界框的对角线长度。而 DIoU 损失函数可以定义为

$$L_{DIoU} = 1 - IoU + \frac{h^2(b, b^{gt})}{d^2} \quad (3)$$

1.2 完整 IoU 损失

边界框回归的良好损失应该考虑 3 个重要的因

素, 即重叠面积、两个边界框的中心点距离和纵横比。通过统一坐标, IoU 损失考虑重叠区域。当预测边框与目标边框有重叠甚至存在包含关系时, DIoU 皆在考虑重叠面积和中心点距离达到了很好的效果, 然而边界框纵横比的一致性这一重要的几何因素还没有被考虑进来。因此在 DIoU 损失的基础上通过对纵横比的一致性施加影响, CIoU 损失的惩罚项为

$$R_{CIoU} = \frac{h^2(b, b^{gt})}{d^2} + fu \quad (4)$$

其中, f 是一个正的权衡参数, u 是测量长宽比的一致性。

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

所以 CIoU 损失函数定义为

$$L_{CIoU} = 1 - IoU + \frac{h^2(b, b^{gt})}{d^2} + fu \quad (6)$$

权衡参数 f 定义为

$$f = \frac{u}{(1 - IoU) + u} \quad (7)$$

其中重叠区域因子在回归中具有更高的优先权, 尤其是在非重叠情况下。CIoU 损失的优化与 DIoU 的优化基本相同, 只是需要指定 u 、 w 和 h 的梯度。

$$\begin{aligned} \frac{\partial u}{\partial w} &= \frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \times \frac{h}{w^2 + h^2} \\ \frac{\partial u}{\partial h} &= -\frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \times \frac{w}{w^2 + h^2} \end{aligned} \quad (8)$$

对于 h 和 w 在 $[0, 1]$ 范围内的情形, $w^2 + h^2$ 一般是一个比较小的值, 这样可能产生梯度爆炸。因此为了收敛的稳定性, 简单地去掉了控制因子 $w^2 + h^2$, 用 1 替代步长 $w^2 + h^2$, 这样梯度方向仍然会和式(8)一致。

2 优化边界框回归

2.1 OIoU 损失

为了解决边界框回归的第二个问题即当真实的边界框存在模糊性时会导致定位不准确, 引入边界框回归损失 KL 损失。特别地, 为了捕捉边界框预测的不确定性, 将预测边界框及真实边界框分别看

做高斯分布函数和狄克拉分布函数,则回归损失可以看作是预测分布和真实分布之间的 KL 散度。尽管分布可能非常复杂像混合高斯那样,但是可以假设坐标是独立分布的,并用一元高斯来简化:

$$P_\alpha = \frac{1}{\sqrt{2\pi t^2}} e^{-\frac{(x-x_e)^2}{2t^2}} \quad (9)$$

其中, α 是学习参数集合, x_e 是估计的边界框位置,标准差 t 表示估计的不确定性。当 t 趋于 0 时,说明框架的估计值非常接近真实值。真实边界框也可以表示为一个高斯分布,当 t 趋于 0 时代表狄克拉函数:

$$P_{D(x)} = \delta(x - x_g) \quad (10)$$

其中 x_g 是真实的边界框位置。物体定位的目标估计是 $\hat{\alpha}$,使 P_α 和 $P_{D(x)}$ 对于 N 个样本的 KL 散度最小化:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{N} \sum D_{KL}(P_{D(x)} \| P_\alpha(x)) \quad (11)$$

使用 KL 散度作为边界框回归的损失函数:

$$\begin{aligned} L_{KL} &= D_{KL}(P_D(x) \| P_\alpha(x)) \\ &= \int P_D(x) \log P_D(x) dx \\ &\quad - \int P_D(x) \log P_\alpha(x) dx \end{aligned} \quad (12)$$

因此为了解决传统的 IoU 损失存在收敛速度慢、回归不准确的问题和真实的边界框模糊性问题,本文提出优化交并比(OIoU)边界框损失,即:

$$\begin{aligned} L_{OIoU} &= 1 - IoU + \frac{h^2(b, b^{gt})}{d^2} + fu \\ &\quad + ED_{KL}(P_D(x) \| P_\alpha(x)) \end{aligned} \quad (13)$$

算法 1 IoU 和 OIoU 作为边界框回归损失

输入: 预测边界框 b 和真实的边界框 b^{gt} , 物体定位的目标估计 \hat{a} , 权衡参数 E 。并且:

$$b = (x_1, y_1, x_2, y_2), b^{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt})。$$

输出: L_{IoU}, L_{OIoU} 。

1) 对于预测边界框 b 确保 $x_2 > x_1, y_2 > y_1$, 主要是为了方便之后点的对应关系:

$$\begin{aligned} \hat{x}_1 &= \min(x_1, x_2), \hat{x}_2 = \max(x_1, x_2) \\ \hat{y}_1 &= \min(y_1, y_2), \hat{y}_2 = \max(y_1, y_2) \end{aligned}$$

2) 计算 b^{gt} 的面积: $A^{gt} = (x_2^{gt} - x_1^{gt}) \times (y_2^{gt} - y_1^{gt})$

3) 计算 b 的面积: $A = (x_2 - x_1) \times (y_2 - y_1)$

4) 计算预测框 b 和真实的边界框 b^{gt} 之间的重叠面积:

$$\begin{aligned} x_1^f &= \max(\hat{x}_1, x_1^{gt}), x_2^f = \min(\hat{x}_2, x_2^{gt}) \\ y_1^f &= \max(\hat{y}_1, y_1^{gt}), y_2^f = \min(\hat{y}_2, y_2^{gt}) \end{aligned}$$

其中 E 是权衡参数, E 的取值范围为 $[0, 1]$ 。当真实的边界框不存在上述的模糊性时, β 此时为 0。优化边界框回归有以下优点。(1) 数据集中的不确定性可以被很好地捕捉到, 边界框回归器从模糊的边界框中可以获得较小的损失;(2) 所学概率分布反映了边界框预测的不确定性程度;(3) OIoU 可以直接最小化两个边界框的差异, 所以收敛速度要比之前的方法快得多;(4) 当真实的边界框和预测边界框存在包含关系时 OIoU 损失可以使回归比较快。目标是估计定位置信度, 网络预测一个概率分布而不仅仅是边界框的位置。优化边界框回归框架如图 1 所示。本文方法的网络框架 OIoU 损失用于评估本地化置信度, 和两阶段检测网络的标准 Fast R-CNN 存在明显的不同, 如图 1 框内本文的方法 OIoU 网络框架用于计算标准偏差和边界框位置。

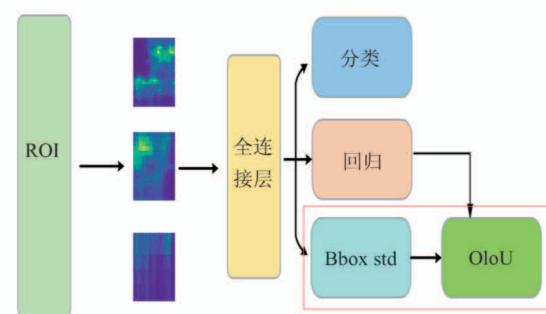


图 1 优化边界框回归框架

2.2 OIoU 算法流程

本文方法的参数更新流程如算法 1 所示。

$$f = \begin{cases} (x_2^f - x_1^f) \times (y_2^f - y_1^f) & \text{if } x_2^f > x_1^f, y_2^f > y_1^f \\ 0 & \text{其他} \end{cases}$$

5) 找到可以包含预测边界框 b 和真实的边界框 b^g 的最小边界框 b^c

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

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

6) 计算 b^c 的面积和 IoU:

$$A^c = (x_2^c - x_1^c) \times (y_2^c - y_1^c), IoU = \frac{f}{A + A^g - f}$$

7) 计算对于 N 个样本的 KL 散度最小化:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{N} \sum D_{KL}(P_{D(x)} \| P_{\alpha}(x))$$

8) 计算长宽比的一致性 u 和权衡参数 f :

$$u = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, f = \frac{u}{(1 - IoU) + u}$$

9) 计算 OIoU:

$$L_{OIoU} = 1 - IoU + \frac{h^2(b, b^{gt})}{d^2} + fu + ED_{KL}(P_D(x) \| P_{\alpha}(x))$$

3 实验

为了验证本文所提出的方法, 使用了 PASCAL VOC 和 MS-COCO 这两个数据集。所有实验均用 GPU 型号为 GTX 1080 Ti。训练计划和批量大小根据不同的网络架构进行调整, 对于 VGG-16^[16] 网络是基于 TensorFlow^[17] 实现的。对于 ResNet-50-FPN^[18] 是基于 PyTorch 实现。

3.1 在 PASCAL VOC 上验证

PASCAL VOC 是最流行的目标检测数据集之一。PASCAL VOC 2007 和 2012 数据集总共分为 4 个大类其中包括 20 个小类(加背景 21 类)。使用 Faster R-CNN 在 PASCAL VOC 上进行实验。使用 VOC2007 和 VOC2012 作为训练集, 包括来自 20 个类的 16 551 张图片。测试集用 VOC2007 由 4952 张图片组成。网络使用 VGG-16, 网络结构如图 2 所示。VGG-Net 是一种深层卷积网络结构, VGG 采用的是一种预训练的方式, 先训练浅层的简单网络 VGG-11, 再复用 VGG-11 的权重来初始化 VGG-13, 如此反复训练并初始化 VGG-16。VGG-16 包含了 16 个隐藏层, 分别为 13 个卷积层和 3 个全连接层, 整个网络都使用卷积核尺寸为 3×3 和最大尺寸 2×2 。简单来说, 在 VGG 中, 使用了 3 个 3×3 卷积

核来代替 7×7 卷积核, 使用了 2 个 3×3 卷积核来代替 5×5 卷积核, 这样做的主要目的是在保证具有相同感知野的条件下, 提升了网络的深度, 在一定程度上提升了神经网络的效果。VGG 的优点是结构非常简洁, 整个网络都使用了同样大小的卷积核尺寸(3×3)和最大池化尺寸(2×2); 几个小滤波器(3×3)卷积层的组合比一个大滤波器(5×5 或 7×7)卷积层好; 验证了通过不断加深网络结构可以提升性能。本次实验批量的大小设置为 256, 动量设置为 0.9, 在前面全连接层使用 Dropout 值设置为 0.5, 学习率初始化为 0.01 并随着迭代进行动态更新。

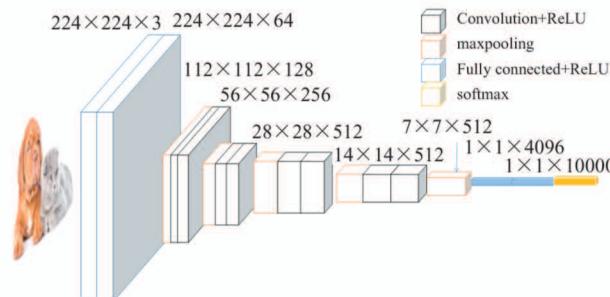


图 2 VGG-16 网络结构

实验结果呈现在表 1 和图 3 中。根据图表可以看出, 本文提出的方法在 PASCAL VOC 数据集上实现了比之前的方法更好的结果。由图 3 可以看出, OIoU 和其他方法相比在少量的训练情况下就可以

得到较高的精度。同时这也证明了此框架在收敛速度和准确率方面比之前的方法要好。

表 1 各种方法的精确度

方法	AP	AP70
IOU	0.511	0.548
GIOU	0.516	0.558
CIOU	0.532	0.568
KL	0.536	0.565
OIOU	0.539	0.571

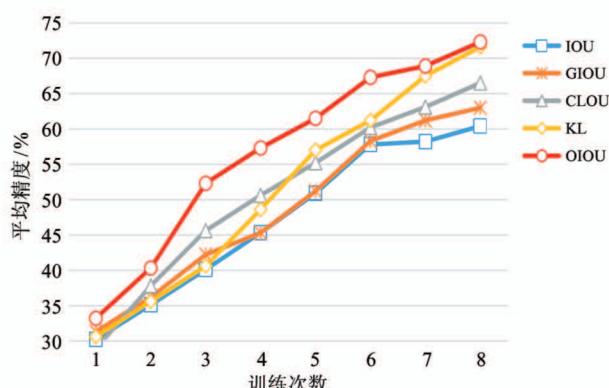


图 3 各方法的精确度随训练批次的变化

3.2 在 MS-COCO 上验证

为了更好地验证本文的方法,在 MS-COCO 数据集上进行实验。MS-COCO 数据集是微软开发维护的大型图像数据集,此数据集的任务包含识别、分割及检测。与 PASCAL VOC 数据集相比,MS-COCO 中的图片包含了自然图片及生活中常见的图片,背景比较复杂,目标数量比较多,目标尺寸更小。因此 MS-COCO 数据集上的任务更难,对于检测任务来说,现在衡量一个模型好坏的标准更加倾向于使用 MS-COCO 数据集的检测结果。网络使用 ResNet-50-FPN,它通过使用 Residual Unit 来训练 152 层深的神经网络,能够使超深神经网络的训练速度大大加快,在模型的准确率方面也有很大的提高。直观上,面对复杂问题越深的网络往往有更好的性能,然而实际发现,随着网络的加深,出现了训练集准确率下降的现象。原因是梯度消失,当网络达到一定深度后,网络的性能将恶化。为了解决退化问题,利用残差模块在输入和输出之间建立一个直接连接,这

样新增的层仅仅需要在原来的输入层基础上学习新的特征及深度残差网络。

从表 2 和图 4 可以看出本文的方法 OIoU 要优于之前的其他方法,图 5 展现了利用 Faster R-CNN 在 MS-COCO 上训练的检测示例。其中图 5(a)左图为之前的方法,其没有把车整个标注出来;图 5(b)左图虽然把猫和狗都标注出来了,但所用的边界框过大;图 5(c)左图为之前的方法,标注出现了错误,实际上应该是一列完整的火车。

表 2 各方法的精确度

方法	AP	AP ⁵⁰	AP ⁶⁰	AP ⁷⁰	AP ⁸⁰
IOU	0.392	0.579	0.536	0.474	0.365
GIOU	0.396	0.583	0.541	0.482	0.376
CIOU	0.402	0.593	0.542	0.485	0.379
KL	0.395	0.589	0.544	0.476	0.361
OIOU	0.415	0.599	0.546	0.492	0.385

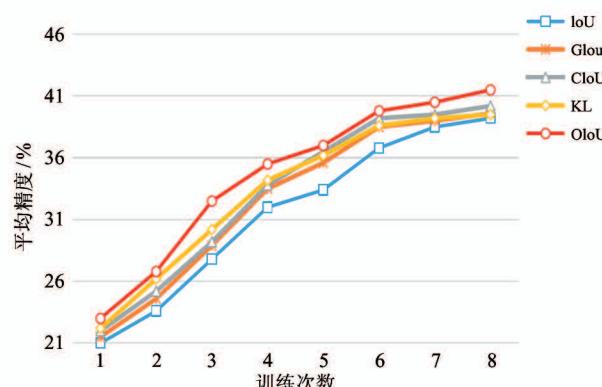


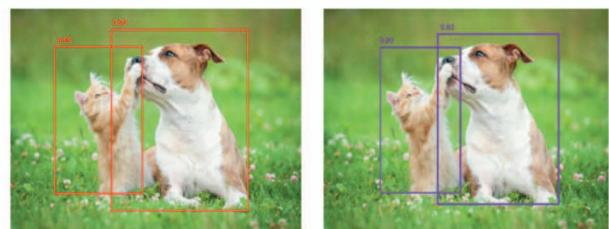
图 4 各方法的精确度随训练批次的变化

4 结 论

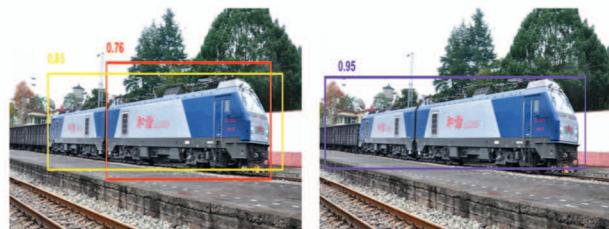
本文基于 CLoU 损失和 KL 损失提出了一种新的边界框回归 OIoU 框架,此框架综合考虑了传统的 IoU 损失存在收敛速度慢、回归不准确的问题和真实的边界框模糊性。本文方法的网络框架和两阶段标准的检测网络不同,OIoU 网络框架计算标准偏差和边界框位置。优化边界框回归损失在 PASCAL VOC 数据集上实现了比之前方法更佳的结果,并在背景比较复杂、目标数量比较多、目标尺寸更小的 MS-COCO 数据集上展现出了更好的检测结果并解



(a) 静止物体



(b) 动物



(c) 运动物体

图 5 检测示例

决了之前方法收敛速度慢、回归不确定的问题,也表明本文方法具有较好的泛化能力和检测精度,能够推广到其他流行的网络框架中。图片风格、光照、物体的形状、背景不同时,测试数据和训练数据存在较大的跨域偏移问题,面对测试集和训练集存在域偏移问题时,OIoU 检测速度和精确度会受到不同程度的影响。接下来的工作会针对以上问题对本文模型进行改进以提高 OIoU 的泛化能力。

参考文献

- [1] Girshick R. Fast R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision, San diego, Chile, 2015 : 1440-1448
- [2] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016 : 779-788
- [3] Cai Z, Vasconcelos N. Cascade R-CNN: delving into — 752 —
- [4] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C] // European Conference on Computer vision, Amsterdam, Netherlands, 2016 : 21-37
- [5] 刘畅, 王鹏钧, 赵潇, 等. 监控视频 ROI 完整分割技术研究[J]. 高技术通讯, 2019, 29(11) : 1053-1062
- [6] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017 : 2961-2969
- [7] Huang Z, Huang L, Gong Y, et al. Mask scoring R-CNN [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019 : 6409-6418
- [8] 刘畅, 王鹏钧, 张美玲, 等. 基于 IOU 分析的稀疏视频检测技术研究[J]. 高技术通讯, 2019, 29(10) : 943-950
- [9] Yu J, Jiang Y, Wang Z, et al. Unitbox: an advanced object detection network[C] // Proceedings of the 24th ACM International Conference on Multimedia, Berlin, Germany, 2016 : 516-520
- [10] Jiang B, Luo R, Mao J, et al. Acquisition of localization confidence for accurate object detection[C] // Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018 : 784-799
- [11] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019 : 658-666
- [12] Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: faster and better learning for bounding box regression[C] // American Association for Artificial Intelligence Conference, New York, USA, 2020 : 12993-13000
- [13] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C] // European Conference on Computer Vision, Zurich, Switzerland, 2014 : 740-755
- [14] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2) : 303-338
- [15] He Y, Zhu C, Wang J, et al. Bounding box regression

- with uncertainty for accurate object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 2888-2897
- [16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv*: 1409.1556, 2014
- [17] Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems [J]. *arXiv*: 1603.04467, 2016
- [18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778

Target detection based on optimized bounding box regression

Zhao Wencang, Xu Changkai, Wang Chunxin

(College of Automation Engineering, Qingdao University of Science and Technology, Qingdao 266061)

Abstract

Target detection is one of the key problems in computer vision tasks, and bounding box regression is an important step in target detection. In the existing methods, intersection over union (IoU) loss is considered to be beneficial to IoU measurement, but there are still problems such as slow convergence and inaccurate regression. In addition, large-scale target detection data sets try to define the true bounding box as clearly as possible, but biases are still introduced when marking the bounding box. Therefore, this paper proposes an optimized bounding box optimize-IoU (OIoU) loss based on complete-IoU (CIoU) loss and bounding box regression Kullback-Leibler (KL) loss, and solves the above problems mainly by learning boundary transformation and positioning variance. OIoU is not only used to evaluate the localization confidence, but also to calculate the standard deviation and the position of the bounding box, which solves the problems of deviation and inaccurate regression caused by the bounding box labeling while speeding up the convergence speed. On the two data sets of PASCAL VOC and MS-COCO, it is proved that the optimized bounding box loss is superior to the previous methods in detection accuracy and convergence speed.

Key words: target detection, optimized bounding box regression, Kullback-Leibler (KL) loss, faster R-CNN