

使用 GIoU 改进非极大值抑制的目标检测算法

侯志强¹, 刘晓义¹, 余旺盛², 蒲磊², 马素刚¹, 范九伦³

(1. 西安邮电大学计算机学院, 陕西西安 710121; 2. 空军工程大学信息与导航学院, 陕西西安 710077;
3. 西安邮电大学通信与信息工程学院, 陕西西安 710121)

摘要: 针对单阈值-非极大值抑制算法中出现的目标漏检和重复检测问题, 本文提出了一种使用全局交并比指标 GIoU (Generalized Intersection over Union) 衡量目标相似度的双阈值非极大值抑制算法 GDT-NMS (Generalized Dual Threshold NMS, GDT-NMS). 使用双阈值改进 NMS 算法和 soft-NMS 算法, 抑制多余的检测框; 在此基础上, 使用 GIoU 替换传统的 IoU 计算目标间的相似度, 使目标的定位更加准确; 进一步, 使用非线性函数赋予检测框不同比例的权值惩罚, 使检测框的得分随距离呈非线性变化, 目标区分度更高. 改进算法在 PASCAL VOC 和 MSCOCO 上的检测精度分别为 74.8% 和 25.9%, 与使用 NMS 算法作为后处理的 Faster R-CNN 算法相比, 性能分别提升了 1.6% 和 1.5%. 同时本文算法具有较快的检测速度.

关键词: 双阈值; 非极大值抑制算法; 重复检测; 后处理

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2021)04-0696-10

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20200132

Object Detection Algorithm for Improving Non-Maximum Suppression Using GIoU

HOU Zhi-qiang¹, LIU Xiao-yi¹, YU Wang-sheng², PU Lei², MA Su-gang¹, FAN Jiu-lun³

(1. School of Computer, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China;

2. School of Information and Navigation, Air Force Engineering University, Xi'an, Shaanxi 710077, China;

3. School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China)

Abstract: Aiming at the problem of missed detection and repeated detection in the single-threshold-non-maximum suppression algorithm, this paper proposes a dual-threshold Non-Maximum Suppression algorithm using GIoU (Generalized Intersection over Union). Using dual thresholds to improve the NMS algorithm and the soft-NMS algorithm, suppressing redundant detection boxes, not only balances the relationship between the object missed detection problem and the object false detection problem caused by the single threshold algorithm, but also reduces the occurrence of the soft-NMS algorithm. Based on the above, using GIoU instead of IoU to calculate the similarity between objects, so that the positioning of the object is more accurate; the non-linear function is used to give different weights to the proposal boxes, which makes the proposal boxes' scores change non-linearly with distance, and the target discrimination is higher, which is more conducive to suppressing the proposal boxes. The detection accuracy of the improved algorithm on PASCAL VOC and MSCOCO is 74.8% and 25.9%, respectively. At the same time, the algorithm in this paper has a fast detection speed.

Key words: double threshold; non-maximum suppression algorithm; repeated detection; post processing

1 引言

目标检测^[1,2]是计算机视觉领域中重要的研究方

向和研究热点, 在行人检测、文字检测、商标检测等^[3-8]多个领域有着重要的研究意义. 随着深度学习的发展, 使用神经网络的目标检测算法取得了很大的突破, 一

般分为一阶段检测算法和两阶段检测算法. 其中, 一阶段检测算法中以 YOLO 系列^[9-11]、SSD^[12-14] 系列等算法为代表, 两阶段检测算法以 R-CNN^[15]、Fast R-CNN^[16]、Faster R-CNN^[17] 等算法为代表.

在上述算法中, 大部分^[10-14,16,17] 都使用了非极大值抑制算法 NMS (Non-Maximum Suppression) 对检测框进行筛选和抑制, 只保留一部分检测框作为最终的目标位置, 因此 NMS 算法作为一种重要的后处理算法, 对目标检测的性能有着很大的影响. 传统的 NMS 算法在抑制检测框时采用贪心算法的思想, 在迭代选择中, 对于其 IoU (Intersection over Union) 值大于预设阈值的检测框都进行滤除, 这在很大程度上导致了目标的漏检和误检, 尤其是在目标高度重叠的场景中.

所以, 出现了很多针对 NMS 算法的改进算法. Henderson P 等人^[18] 提出直接在网络训练中调整 NMS 算法的参数, 即端到端的训练网络. 后来, Hosang J 等人^[19] 直接利用卷积神经网络的方式来代替 NMS 算法生成检测框. 随着循环神经网络 (Recurrent Neural Network, RNN) 的发展, Stewart R 等人^[20] 提出使用长短时记忆网络 (Long Short-Term Memory, LSTM) 来代替 NMS 算法. 这些方法都对 NMS 算法进行了一定的改进, 但往往无法兼顾检测性能和检测速度, 部分算法所需的计算量远远超过了 NMS 算法. soft-NMS 算法^[21] 采用“权值惩罚”策略去解决 NMS 算法带来的漏检问题, 在一定程度上取得了较好的效果. 但是实验表明, 使用 soft-NMS 算法会产生同一目标被重复检测和错分目标的新问题.

针对以上问题, 本文提出了一种使用全局交并比指标 GIoU^[22] 衡量目标相似度的双阈值非极大值抑制算法 GDT-NMS (Generalized Dual Threshold NMS, GDT-NMS). 具体工作有:

(1) 使用双阈值改进 NMS 算法和 soft-NMS 算法, 抑制多余的检测框, 既很好地平衡了单阈值算法带来的目标漏检问题和目标误检问题的关系, 又针对性地减小 soft-NMS 算法中出现的重复检测问题.

(2) 使用 GIoU 去替换 IoU 计算目标间的相似度, 相比 IoU, GIoU 能更好地刻画出目标间的相对位置和重叠情况, 使目标的定位更加准确.

(3) 使用非线性函数赋予检测框不同的权值惩罚, 使检测框的得分随距离呈非线性变化, 目标区分度更高, 更有利于对检测框进行抑制.

经过上述改进后的算法在 VGG16-Net 的网络结构上重新进行了网络训练.

(4) 提出了重复检测率和重复检测的目标错分率作为衡量指标, 以更好地对目标重复检测问题进行评价.

2 相关工作

目前, 在目标检测领域中, 人们通常使用 NMS 和 soft-NMS 算法进行后处理, 移除多余的检测框.

其中, NMS 算法^[23] 使用式 (1) 对检测框进行多轮过滤:

$$s_j = \begin{cases} s_i, & \text{IoU}(M, b_i) < N_i \\ 0, & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (1)$$

式中, b_i 为第 i 个待检测框, s_i 为 b_i 原始得分, s_j 为 b_i 的最后得分, M 为每一轮得分最高的检测框, 即抑制检测框. $\text{IoU}(M, b_i)$ 为 b_i 和 M 的交并比, N_i 为预设阈值.

但是, NMS 算法很容易产生目标的漏检和误检问题. 如图 1(a1) 所示, 使用 NMS 算法无法将目标都检测到, 产生了目标的漏检问题. 在图 1(a2) 中, 使用 NMS 算法错将目标检测成人, 产生了目标的误检问题.

soft-NMS 算法^[21] 采用“权值惩罚”决了漏检问题, 如式 (2) 所示.

$$s_j = \begin{cases} s_i, & \text{IoU}(M, b_i) < N_i \\ s_i \times (1 - \text{IoU}(M, b_i)), & \text{IoU}(M, b_i) \geq N_i \end{cases} \quad (2)$$

式中的符号含义与式 (1) 相同. 如图 1(b1) 所示, 使用 soft-NMS 算法可以检测到更多的目标. 在图 1(b2) 中, 使用 soft-NMS 算法修正了目标误检的问题.

但是实验发现, soft-NMS 算法会出现在同一目标上的重复检测, 造成目标的“假正例”. 如图 2(b1) 和图 2(b2) 所示, 使用 soft-NMS 算法对人和船产生了重复的检测.

造成上述问题的原因有两点:

(1) NMS 算法和 soft-NMS 算法都属于单阈值-非极大值抑制算法. 阈值的选择非常重要, 阈值设置过大, 容易出现目标漏检问题, 阈值设置过小, 又容易出现目标误检问题. 单阈值很难平衡好这两者之间的关系.

(2) NMS 算法与 soft-NMS 算法都使用 IoU ^[24] 来计算两个检测框的相似度. 但是, 根据文献[22], IoU 的值不能准确地刻画出两个目标间的距离和重叠方式, 会影响目标定位的准确度.

3 本文算法

本文提出了一种使用 GIoU 衡量目标相似度的双阈值非极大值抑制算法 GDT-NMS. 算法首先使用双阈值去改进 NMS 算法和 soft-NMS 算法; 在此基础上, 使用 GIoU 替换 IoU 计算目标间的相似度, 更准确地刻画出目标间的相对位置和重叠情况; 最后, 使用非线性函数值赋予检测框不同的权值惩罚, 使目标区分度更高.

3.1 双阈值非极大值抑制算法

在本文中, 我们使用式 (3) 对检测框进行抑制和滤除.

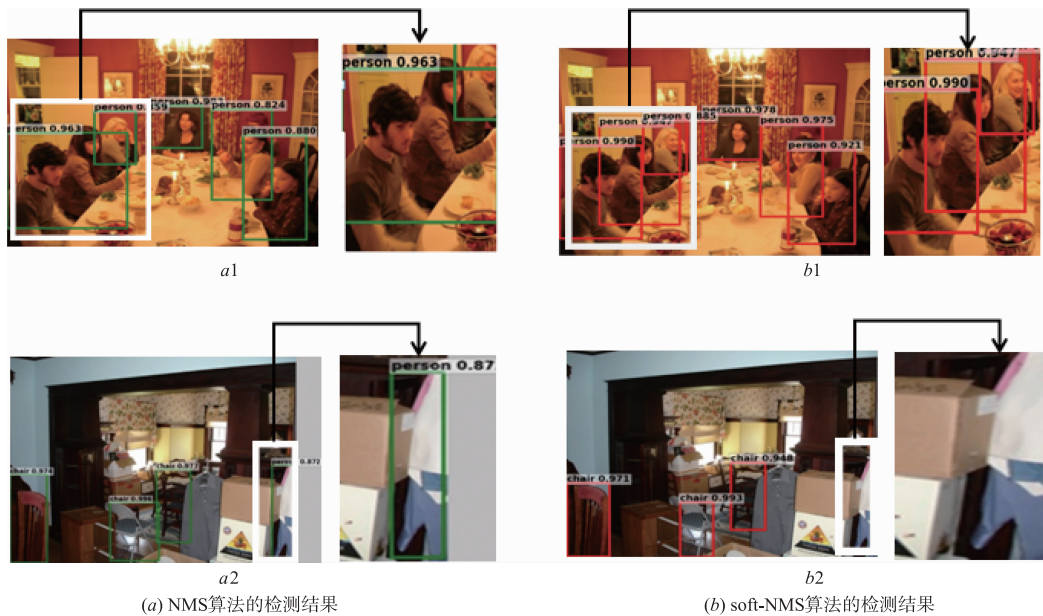


图1 不同的后处理算法在目标漏检和目标错检问题上的结果对比

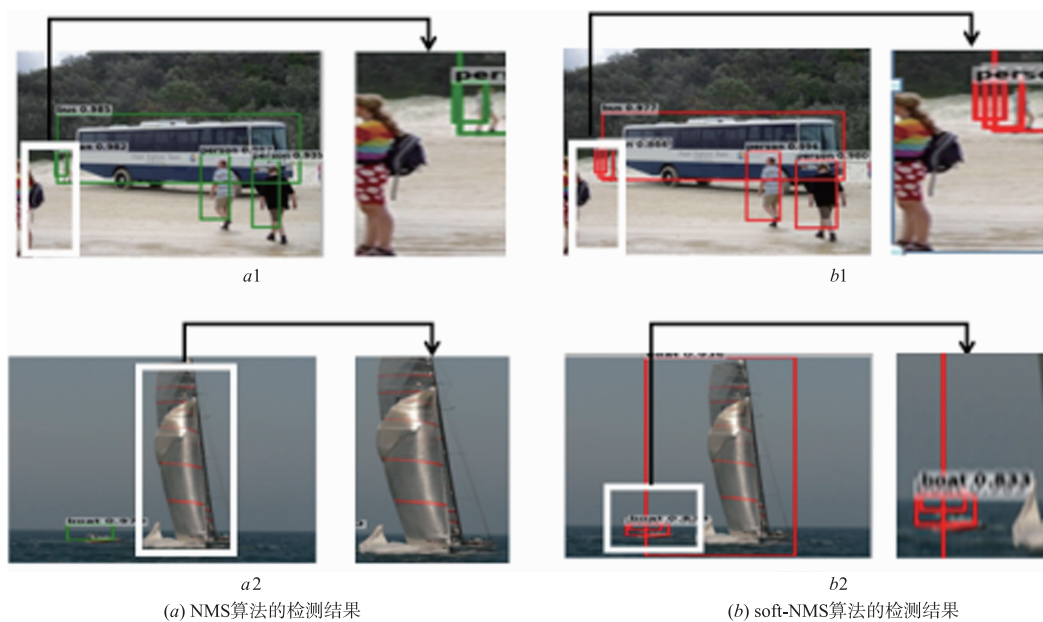


图2 不同的后处理算法在目标重复检测上的结果对比

$$s_f = \begin{cases} s_i, & Glou(M, b_i) \leq N_i \\ s_i \times (1 - f(Glou(M, b_i) + 1)), & N_i < Glou(M, b_i) < N_i \\ 0, & Glou(M, b_i) \geq N_i \end{cases} \quad (3)$$

式(3)中, b_i 为第 i 个待检测框, s_i 为 C 的原始得分, s_f 为 b_i 的最后得分, M 为此时得分最高的检测框, $Glou(M, b_i)$ 为 b_i 和 M 的 $GIoU$ 值, N_i 和 N_i 为阈值。

本文算法与 NMS 算法, soft-NMS 算法的不同之处在于:

(1) 使用两个阈值去处理检测框. 当 b_i 与 M 的 $GIoU(M, b_i)$ 大于 N_i 时, 进行权值惩罚, 降低了目标在高度重叠时的漏检和误检的概率. 并且, 删除与 M 的 $GIoU(M, b_i)$ 值大于 N_i 的待检测框, 当 $GIoU(M, b_i)$ 过大时, M 与 b_i 为同一目标的可能性非常大, 此时会加大目标被重复检测的概率。

(2) 在计算检测框的得分时, 使用 $GIoU$ 代替传统的 IoU 计算目标间的相似度, 在 3.2 节进行详细介绍。

(3) 相比于 soft-NMS 算法, 本文使用非线性的权值惩罚策略, 更好地区分出目标, 这部分工作将在 3.3 节

详细介绍.

由图 3 中的局部放大图(d1)和(d2)可以看出,本

文算法既解决了 NMS 算法中的目标漏检问题,又修正了 soft-NMS 算法的重复检测问题.

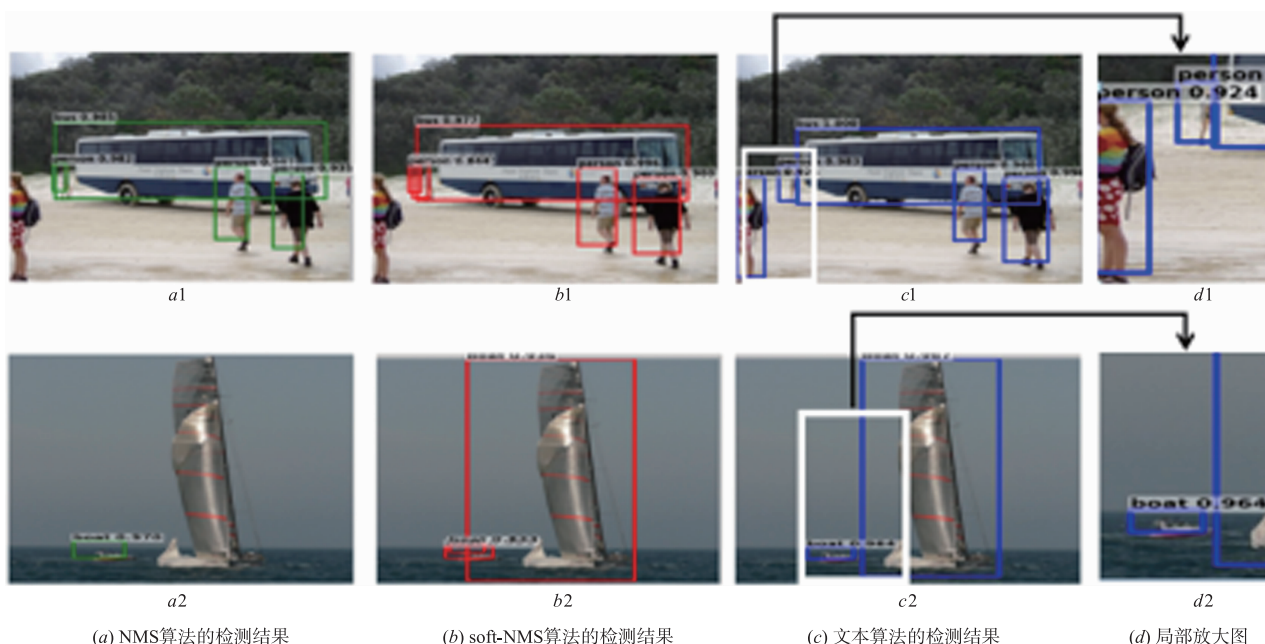


图3 本文算法在重复检测问题上与其他算法的结果对比

3.2 使用 GIoU 替换 IoU 作为衡量两个检测框相似度的指标

GIoU 使用式(4)和式(5)去计算两个检测框的相似度.

$$D = C \setminus (A \cup B) \tag{4}$$

$$GIoU = IoU - \frac{|D|}{|C|} \tag{5}$$

在式(4)中, A 和 B 代表两个检测框, C 为 A, B 的最小外接矩形, D 为矩形 C 中不含 A 和 B 并集的部分. 符号“ \setminus ”代表去掉的意思. 式(5)中 D, C 的含义与式(4)相同. IoU 和 GIoU 计算方式的不同之处如图 4 所示.

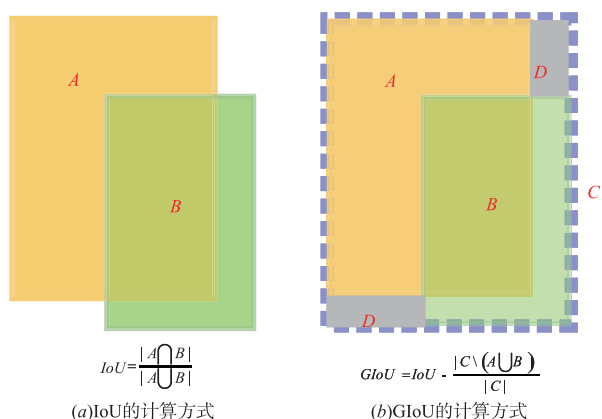


图4 IoU和GIoU计算方式的对比

利用 IoU 计算检测框的相似度,有以下问题:

(1) IoU 的值只能刻画两个目标的重叠程度,而无法说明两个目标之间的距离. 当两个目标不发生重叠时,它们的 IoU 值将等于零,此时,无法说明两个检测框的距离远近.

(2) IoU 的值不能说明两个目标间的重叠形式. 根据文献[22],当两个检测框在许多不同方向上有重叠,并且它们的交集水平相同时, IoU 值相同. 因此,仅使用 IoU 指标去衡量两个检测框的相似度,无法保证被保留下来的检测框更接近真实位置,进而影响算法的检测性能.

相比于 IoU, GIoU 的优势是:

(1) 两个框之间的距离不再仅由 A, B 本身的信息决定. 通过引入一个最小外接矩形 C , 增加了 D 的信息. 此时,两个框之间“距离”通过 A, B, D 共同刻画得出. 即使 A, B 不发生重叠, GIoU 也能更好地反映出两个框的远近.

(2) 当两个目标没有完全重合时,它们的重叠方式显示了对齐程度和相似度. 两个目标之间的对齐程度越接近,目标的相似度越高. 当 IoU 的值不能反映出目标之间如何重叠时, GIoU 通过引入 C 和 D 的信息可以更好地反映出检测框的对齐程度. 因此,使用 GIoU 作为度量目标相似度的指标,能够保证被保留下来的检测框的定位更加准确,从而更好地提升算法的检测性能.

图 5 给出了两组重叠方式不同检测框,图中红色的矩形框分别代表两个候选目标 $A1$ 和 $A2$,绿色的矩形框 B 代表该目标的真实位置,红色阴影部分代表信息 D ,紫色

线条围成的矩形是同时包含这两个目标的最小矩形 C 。显然,当无法使用 IoU 值判断哪个检测框更接近真实位置时,GIoU 的值可以更好地选择对齐度高的检测框。

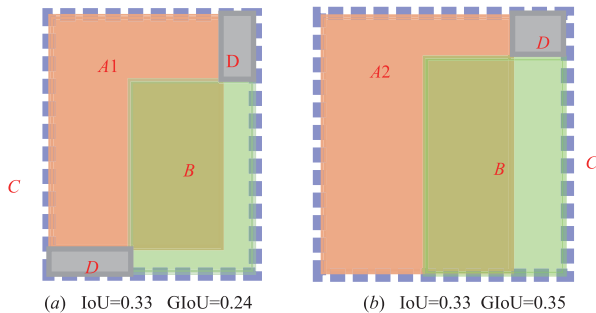


图5 IoU值相同但是GIoU值不同的两组检测框的重叠情况比较

因此,使用 GIoU 替换 IoU 作为衡量两个检测框相似度的指标可以弱化 IoU 的缺点,不仅能够利用增加的信息刻画出两个框之间的距离,而且可以更好的说明两个框的重叠情况。

3.3 非线性权值惩罚

在 soft-NMS 算法中,检测框的得分根据 IoU 值得到线性比例的惩罚项。实验发现,使用线性比例标准对检测框进行抑制,不能更好地区分目标。因此,在本文中,我们使用 e^x 函数和 \log 函数对检测框的得分 s_i 进行非线性的权值惩罚,具体见式(6)与式(7)。

$$s_f = \begin{cases} s_i, & GIoU(M, b_i) \leq N_i \\ s_i \times (1 - e^{-(GIoU(M, b_i) + 1)}), & N_i < GIoU(M, b_i) < N_i \\ 0, & GIoU(M, b_i) \geq N_i \end{cases} \quad (6)$$

$$s_f = \begin{cases} s_i, & GIoU(M, b_i) \leq N_i \\ s_i \times (1 - \log(GIoU(M, b_i) + 1)), & N_i < GIoU(M, b_i) < N_i \\ 0, & GIoU(M, b_i) \geq N_i \end{cases} \quad (7)$$

式(6)和(7)中, b_i 为第 i 个待检测框, s_i 为 C 的原始得分, s_f 为 b_i 的最后得分, M 为此时得分最高的检测框, $GIoU(M, b_i)$ 为 b_i 和 M 的 GIoU 值, N_i 和 N_i 为阈值。

对于一个检测框 b_i , 如果它和 M 的 GIoU 值处于阈值 N_i 和 N_i 之间, 我们给它的得分赋予如下惩罚: 如果它距离 M 越远, 那么它和 M 成为同一个目标的可能性越低, 惩罚程度越小, 且惩罚因子按照非线性比例减小; 如果它距离 M 越近, 那么它和 M 成为同一个目标的可能性越大, 因此需要加大惩罚程度, 且惩罚因子按照非线性比例增大。这样, 距离 M 很近的检测框 b_i 和距离 M 很远的检测框 b_j 对 M 重要程度不再相同, 它们的得分呈非线性变化, 目标区分度更高, 保留下来的检测框更准确。在文献[25]中也给出了相似的结论。

在进行阈值选择的实验中, 我们发现两种改进算法都能够提升检测性能, 但使用 \log 函数进行改进的算法性能更好一些。

当 $N_i = 0.4$, $N_i = 0.9$ 时, 两种算法的平均检测精度都达到最高。由于篇幅有限, 我们只以 \log 函数改进的算法为代表, 给出阈值选择的过程, 如表 1 所示。表中数据代表了算法在对应阈值下的平均准确度, 由于 $N_i > N_i$, 因此 N_i 取值从 0.2 开始, “-” 表示未取值的部分。

表 1 不同阈值组合下的算法的准确度对比

$N_i \backslash N_i$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	73.50	-	-	-	-	-	-	-	-
0.3	73.61	73.10	-	-	-	-	-	-	-
0.4	74.30	73.33	73.51	-	-	-	-	-	-
0.5	74.51	74.53	73.56	74.50	-	-	-	-	-
0.6	74.55	74.69	73.49	74.51	73.17	-	-	-	-
0.7	74.51	74.73	74.53	74.57	74.50	68.61	-	-	-
0.8	74.60	74.78	74.75	74.66	74.72	69.53	62.70	-	-
0.9	74.67	74.81	74.79	74.86	74.79	69.90	62.73	50.20	-
1.0	74.63	74.79	74.81	74.54	73.20	69.64	62.66	50.20	33.3

4 分析与讨论

本文算法在处理器为 Intel(R) Core i5 8400CPU 和 GTX1080GPU 的计算机上运行, 以 Faster R-CNN 算法作

为检测框架。分别在 PASCAL VOC 和 MSCOCO 数据集进行测试, 算法使用平均准确度 (mean Average Precision, mAP) 和检测速度作为性能评测指标。

4.1 定性分析

为验证本文算法的检测性能,将本文两种改进算法分别在 PASCAL VOC 和 MSCOCO 数据集的检测结果与 NMS 算法,soft-NMS 算法的检测结果进行比较.在本文中,仅选取部分代表性的结果进行展示.由图 6 和图

7 可以看出,当图像中存在多个目标时,本文提出两种改进的算法均比 NMS 算法和 soft-NMS 算法更好一些:有效地减少了目标漏检和误检问题,又更好地修正了重复检测问题.

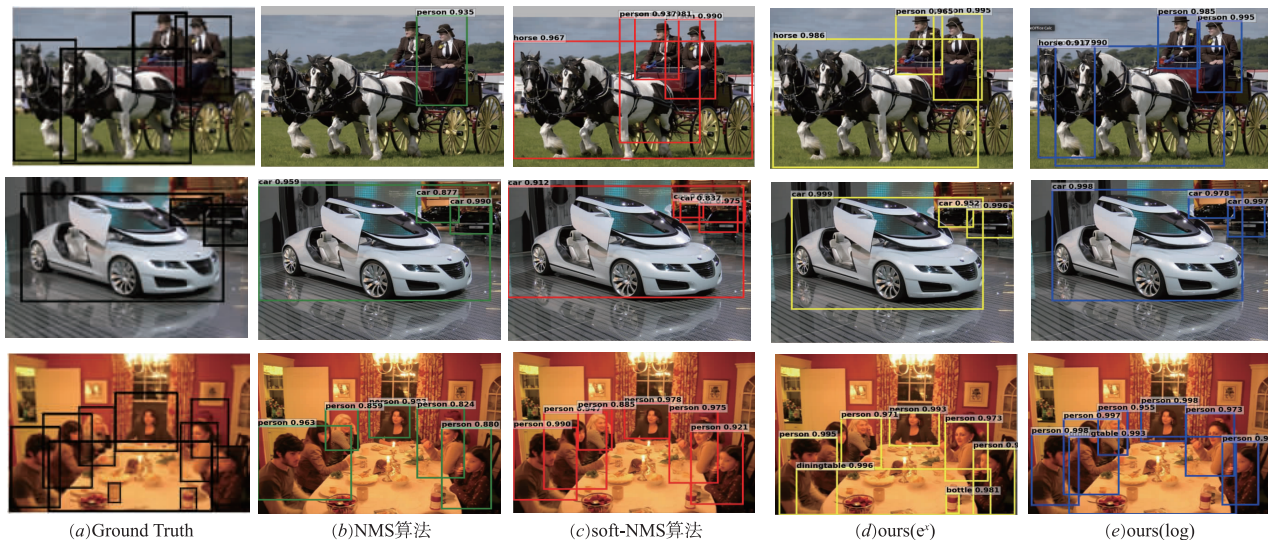


图6 本文算法与其他算法在PASCAL VOC上的实验结果对比



图7 本文算法与其他算法在MSCOCO上的实验结果对比

4.2 定量分析

4.2.1 对目标重复检测问题的分析

目标的重复检测问题会直接影响最终的检测准确率,目标被重复检测的次数增多,会导致“假正例”的问题出现,最终检测准确率出现假高.但这种情况应该被避免.因此,本文提出了重复检测率和重复检测的目标错分率作为衡量指标,以更好地对目标重复检测问题进行评价.重复检测率和错分率的计算如式(8)和式(9).

$$Rd_{rate} = \frac{Rd_{total}}{pic-total} \times 100\% \quad (8)$$

$$C_{error\ rate} = \frac{C_{error}}{Rd_{total}} \times 100\% \quad (9)$$

式(8)中, Rd_{total} 表示目标被重复多次检测的总数, $pic-total$ 表示测试图片总数, Rd_{rate} 为重复检测率.式(9)中, C_{error} 表示目标被重复检测且每次分类不同的次数, $C_{error\ rate}$ 表示被重复检测且目标分错的目标错分率.

表 2 中的 Rd_2 和 Rd_3 分别表示在同一目标上重复

检测 2 次和 3 次的总图像数,其他符号含义与式(8)式(9)相同.由表 2 可知,相比于 soft-NMS 算法,使用 FasterR-CNN 为检测框架的两种改进算法的结果都更好一

些.使用 log 函数进行改进的算法在 PASCAL VOC2007 上的重复检测率比 soft-NMS 算法降低了 2.6%,重复检测的目标错分率降低了 2.2%.

表 2 本文算法和 soft-NMS 算法针对重复检测问题在 PASCAL VOC2007 数据集的结果对比

Algorithm	Rd_{total}	Rd_2	Rd_3	C_{error}	Rd_{rate}	$C_{error\ rate}$
soft-NMS	592	543	49	64	11.9%	10.8%
DT-NMS + IoU + e^x	487	455	45	50	9.8%	10.2%
DT-NMS + IoU + log	479	450	39	52	9.7%	10.8%
DT-NMS + GIoU + e^x	465	438	29	44	9.3%	9.4%
DT-NMS + GIoU + log	461	433	28	40	9.3%	8.6%

4.2.2 本文算法与其他的非极大值抑制算法的比较

以使用 Faster R-CNN 作为检测框架的改进算法为例,表 3 显示了改进算法和其他的非线性的 NMS 改进算法在 PASCAL VOC 数据集上检测结果对比.根据表 3 可以看出,使用 log 函数和 e^x 函数改进的算法在 PAS-

CAL VOC2007 上的检测精度达到 74.8%.使用 log 函数进行改进的算法性能更好一些,比 Center-point NMS^[26] 算法的 mAP 提升 5.8%,比 Faster R-CNN 算法的 mAP 提升 1.6%,比使用指数加权法^[27] 的 NMS 的 mAP 高 1.5%,比 soft-NMS 算法的 mAP 提升 0.5%.

表 3 PASCAL VOC 数据集上的检测结果对比

Algorithm	mAP (%)	bird	boat	bottle	bus	Car	chair	cow
Center-point NMS	69.0	66.6	54.7	55.1	75.6	80.1	50.3	73.6
Faster R-CNN	73.2	70.9	65.5	52.1	83.1	84.7	52.0	81.9
soft-NMS	74.3	74.8	64.9	60.2	81.8	86.2	55.8	80.0
NMS-指数加权法	73.3	73.7	63.7	56.9	79.6	83.9	56.5	80.7
OUR(e^x)	74.8	75.0	64.5	61.6	82.1	86.7	55.8	81.7
OUR(log)	74.8	75.1	63.5	61.9	85.1	87.8	54.6	78.6
Algorithm	Table	dog	horse	person	plant	sheep	sofa	train
Center-point NMS	68.0	76.5	80.3	77.2	40.3	65.5	65.5	76.1
Faster R-CNN	65.7	84.8	84.6	76.7	38.8	73.6	73.9	83.0
soft-NMS	67.5	82.8	82.3	81.0	43.4	77.1	71.7	79.8
NMS-指数加权法	69.4	82.8	82.7	80.8	48.0	70.5	66.8	79.2
OUR(e^x)	67.9	85.1	83.8	80.9	43.0	79.0	70.4	81.2
OUR(log)	67.8	84.5	83.6	81.9	45.3	77.8	71.0	82.3

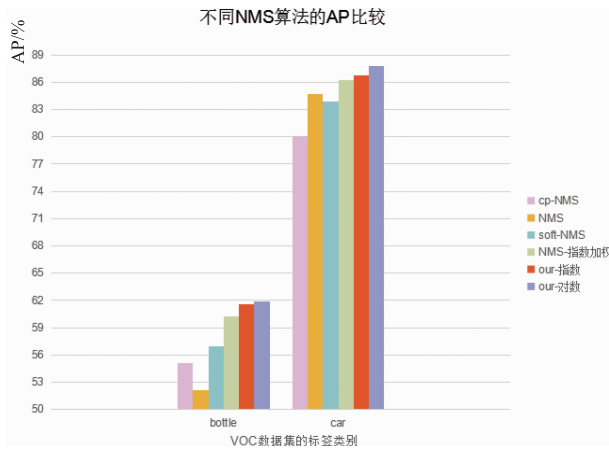
另外,为了更加直观地说明本文算法的优势,以 car、bottle 两类的检测结果为例,图 8(a)给出了不同 NMS 算法的检测性能 AP 对比柱状图,图 8(b)给出了不同 NMS 算法在 PASCAL VOC 数据集上的 mAP 折线图.结合表 3 可以很清楚地看到,本文算法在多个类别上都领先于其他算法,并且,在 PASCAL VOC 数据集上的 mAP 值也具有一定的优势.

4.2.3 本文算法的消融实验

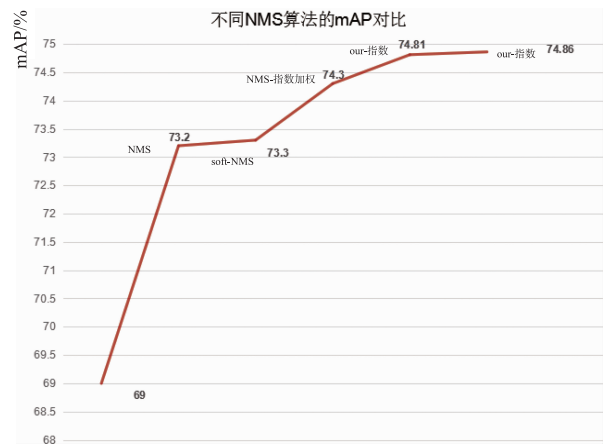
为了进一步说明本文算法的三个改进部分的优越性,表 4 以使用 Faster R-CNN 作为检测框架的本文算法为例,显示了改进算法在 PASCAL VOC 数据集上的消融实验.

以使用式(6)的改进算法为例,根据表 4 的第 4~6 行可以得出:首先,仅通过增加双阈值去改进非极大值抑制算法,算法的 mAP 比 Faster R-CNN 算法提升 1.1%,此时,算法仍然使用 IoU 衡量目标相似度,与 Faster R-CNN 算法相同;其次,在此基础上,将权值惩罚项中的函数替换为非线性函数 e^x ,算法的 mAP 继续提升 0.2%;最后,再引入 GIoU 指标,算法的 mAP 值达到了最高,为 74.81%.

同理,根据表 4 的第 7~9 行可以得出:通过逐步引入本文算法的三个改进部分,使用式(7)的改进算法的检测性能仍然高于对比算法, mAP 值达到 74.86%.



(a)不同NMS算法的AP对比柱状图



(b)不同NMS算法的mAP对比折线图

图8 不同NMS算法的检测性能对比

表 4 以 Faster R-CNN 为检测框架的改进算法在 PASCAL VOC 数据集上的消融实验

Algorithm	Training data	mAP/%	Speed(FPS)
Faster R-CNN + IoU	07 + 12	73.20	15.3
soft-NMS + IoU	07 + 12	74.31	15.1
DT-NMS + IoU	07 + 12	74.33	15.9
DT-NMS + IoU + e ^x	07 + 12	74.53	15.6
DT-NMS + GIoU + e^x	07 + 12	74.81	15.6
DT-NMS + IoU	07 + 12	74.33	15.9
DT-NMS + IoU + log	07 + 12	74.55	15.6
DT-NMS + GIoU + log	07 + 12	74.86	15.6

4.2.4 算法的综合性能分析

根据表 5 可以看出,与相同网络结构下的其他 the-state-of-art 的检测算法相比,使用 Faster R-CNN 作为检测框架的改进算法在 PASCAL VOC 数据集上的 mAP 为 74.8%,高于 RON320^[28] 算法 0.6%,高于 Yolov2 算法 1.1%,高于 SSD 算法 0.5%。在 MSCOCO 数据集上的平均检测率为 25.9%,高于 Faster R-CNN 算法 1.5%,高于 Yolov2 算法 4.3%,高于 SSD 算法 2.7%。同时,使用 SSD 作为检测框架的改进算法在 PASCAL VOC 数据集上的 mAP 为 77.4%,高于 Yolov2 算法 3.7%,高于 SSD 算法 0.4%。在 MSCOCO 数据集上的平均检测率为 24.8%,高于 SSD 算法 1.6%。

表 5 本文算法和其他算法在 PASCAL VOC 和 MSCOCO 数据集上的检测结果对比

Algorithm	Network	PASCAL VOC		MSCOCO	
		Training data	mAP/%	AP@0.5:0.95/%	AP@0.5/%
YoLov2	VGG16	07 + 12	73.7	21.6	44.0
SSD	VGG16	07 + 12	77.0	23.2	41.2
Faster R-CNN	VGG16	07 + 12	73.2	24.4	45.7
RON320	VGG16	07 + 12	74.2	22.7	44.7
soft-NMS	VGG16	07 + 12	74.3	25.5	46.7
Faster R-CNN + OURS(e ^x)	VGG16	07 + 12	74.8	25.8	47.0
Faster R-CNN + OURS(log)	VGG16	07 + 12	74.8	25.9	47.1
SSD + OURS(e ^x)	VGG16	07 + 12	77.3	24.7	43.0
SSD + OURS(log)	VGG16	07 + 12	77.4	24.8	43.1

5 结论

本文提出了一种使用 GIoU 改进的非极大值抑制算法,在后处理阶段有效地降低了目标漏检和误检问题,减小了目标被重复检测的概率,并提升了算法的性能.改进算法在 PASCAL VOC 和 MSCOCO 数据集上进

行了大量实验验证,均取得了不错的结果.下一步工作中,我们将通过使用更好的网络结构,如 Res-Net^[29] 和 Dense-Net^[30] 网络,以获得更好的目标检测性能.

参考文献

[1] Liu L, Ouyang W L, Wang X. Deep learning for generic

- object detection; A survey[J]. arXiv Preprint, 2018, arXiv: 1809.02165.
- [2] 黄凯奇,任伟强,谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6): 1225 - 1240.
HUANG Kai-qi, REN Wei-qiang, TAN Tie-niu. Overview of image object classification and detection algorithms[J]. Chinese Journal of Computers, 2014, 37(6): 1225 - 1240. (in Chinese)
- [3] Shi B, Bai X, Belongie S. Detecting oriented text in natural images by linking segments[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. New York, USA: IEEE, 2017. 2550 - 2558.
- [4] 周炫余,刘娟,卢笑,等. 一种联合文本和图像信息的行人检测方法[J]. 电子学报, 2017, 45(1): 140 - 146.
ZHOU Xuan-yu, LIU Juan, LU Xiao, et al. A pedestrian detection method combining text and image information [J]. Acta Electronica Sinica, 2017, 45(1): 140 - 146. (in Chinese)
- [5] 姜维,张重生,殷绪成. 基于深度学习的场景文字检测综述[J]. 电子学报, 2019, 47(5): 1152 - 1161.
JIANG Wei, ZHANG Chong-sheng, YIN Xu-cheng. Summary of scene text detection based on deep learning [J]. Acta Electronica Sinica, 2019, 47(5): 1152 - 1161. (in Chinese)
- [6] 辛鹏,许悦雷,唐红,等. 全卷积网络多层特征融合的飞机快速检测[J]. 光学学报, 2018, 38(3): 315003.
XIN Peng, XU Yuelei, TANG Hong, et al. Rapid detection of aircraft with multi-layer feature fusion of fully convolutional networks [J]. Acta Optics Sinica, 2018, 38(3): 315003. (in Chinese)
- [7] 丁鹏,张叶,贾平,等. 基于视觉显著性的海面舰船检测技术[J]. 电子学报, 2018, 46(1): 127 - 134.
DING Peng, ZHANG Ye, JIA Ping, et al. Sea-based ship inspection technology based on visual saliency [J]. Acta Electronica Sinica, 2018, 46(1): 127 - 134. (in Chinese)
- [8] Tian Z, Huang W, He T, et al. Detecting text in natural image with connectionist text proposal network[A]. European Conference on Computer Vision [C]. Berlin, Cham: Springer, 2016. 56 - 72.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. New York, USA: IEEE, 2016. 779 - 788.
- [10] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu, USA: IEEE, 2017. 6517 - 6525.
- [11] Redmon Joseph, Farhadi Ali. Yolov3: An incremental improvement[J]. arXiv Preprint, 2018, arXiv: 1804.02767.
- [12] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[A]. European Conference on Computer Vision [C]. Berlin, GER: Springer International Publishing, 2016. 21 - 37.
- [13] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector [J]. arXiv Preprint, 2017, arXiv: 1701.06659.
- [14] Shen Z Q, Liu Z H, Li J G, et al. DSOD: Learning deeply supervised object detectors from scratch[A]. Proceedings of the IEEE International Conference on Computer Vision [C]. New York, USA: IEEE, 2017. 1919 - 1927.
- [15] Girshick R, Donahue J, Darrell T. Rich feature hierarchies for accurate object detection and semantic segmentation [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Columbus, USA: IEEE, 2014. 580 - 587.
- [16] Girshick R. Fast R-CNN[A]. Proceedings of the IEEE International Conference on Computer Vision [C]. New York, USA: IEEE, 2015. 1440 - 1448.
- [17] Ren S, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [A]. International Conference on Neural Information Processing Systems [C]. Cambridge USA: MIT Press, 2015. 91 - 99.
- [18] Henderson P, Ferrari V. End-to-end training of object class detectors for mean average precision [A]. Proceedings of the 13th Asian Conference on Computer Vision [C]. Taipei, China: Springer, 2016. 198 - 213.
- [19] Hosang J, Benenson R, Schiele B. Learning non-maximum suppression [A]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu, USA: IEEE, 2017. 6469 - 6477.
- [20] Stewart R, Andriluka M, Ng A Y. End-to-end people detection in crowded scenes [A]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas, USA: IEEE, 2016. 2325 - 2333.
- [21] Bodla N, Singh B, Chellappa R, et al. Soft-NMS-improving object detection with one line of code [A]. IEEE International Conference on Computer Vision [C]. New York, USA: IEEE, 2017. 5562 - 5570.
- [22] Hamid Reza Tofighi, Nathan Tsoi, JunYoung Gwak, et al. Generalized intersection over union: A metric and a loss for bounding box [J]. arXiv Preprint, 2019, arXiv: 1902.09630.
- [23] Neubeck A, Gool L V. Efficient non-maximum suppression [A]. International Conference on Pattern Recognition [C]. New York, USA: IEEE Computer Society, 2006. 850 - 855.
- [24] Rahman M A, Wang Y. Optimizing intersection-over-u-

- nion in deep neural networks for image segmentation [A]. Advances in Visual Computing [M]. Berlin, GER: Springer International Publishing, 2016. 234 – 244.
- [25] 赵文清, 严海, 邵绪强. 改进的非极大值抑制算法的目标检测 [J]. 中国图象图形学报, 2018, 23 (11): 1676 – 1685.
- ZHAO Wen-qing, YAN Hai, SHAO Xu-qiang. Target detection by improved non-maximum suppression algorithm [J]. Journal of Image and Graphics, 2018, 23 (11): 1676 – 1685. (in Chinese)
- [26] 牛斌, 张怡迪, 马利, 等. 一种融合多尺度特征的多物体检测方法 [J]. 辽宁大学学报 (自然科学版), 2019, 46 (02): 118 – 124.
- NIU Bin, ZHANG Yi-di, MA Li, et al. A multi-object detection method combining multi-scale features [J]. Journal of Liaoning University (Natural Science Edition), 2019, 46 (02): 118 – 124. (in Chinese)
- [27] 张焯, 许艇, 冯定忠, 等. 基于难分样本挖掘的快速区域卷积神经网络目标检测研究 [J]. 电子与信息学报, 2019, 41 (06): 1496 – 1502.
- ZHANG Ye, XU Ting, FENG Ding-zhong, et al. Research on target detection of fast regional convolutional neural network based on mining difficult samples [J]. Journal of Electronics and Information Technology, 2019, 41 (06): 1496 – 1502. (in Chinese)
- [28] Kong T, Sun F C, Yao A B, et al. RON: Reverse connection with objectness prior networks for object detection [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu, USA: IEEE, 2017. 5244 – 5252.
- [29] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. New York, USA: IEEE, 2016. 770 – 778.
- [30] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. New York, USA: IEEE, 2017. 2261 – 2269.

作者简介



侯志强 男, 1973 年 10 月生, 陕西省眉县人. 现为西安邮电大学计算机学院教授, 主要从事图像处理、计算机视觉、无人机应用以及信息融合等领域的研究.



刘晓义 女, 1995 年 8 月生, 陕西省澄城县人. 现为西安邮电大学硕士研究生, 主要从事计算机视觉、目标检测以及图像处理等领域的研究.
E-mail: 18829290763@163.com