DOI:10.19651/j.cnki.emt.2107024

基于注意力机制的NMS在目标检测中的研究*

张长伦　张翠文　王恒友　何　强　刘屹伟
(北京建筑大学 理学院 北京 100044)

摘　要：非最大值抑制算法是目标检测中选择定位准确框的主要算法。针对算法仅以分类得分作为标准可能去除得分低但定位准确的预测框,并且对于有遮挡的情况更不友好的情况,提出了A-NMS方法。该方法将注意力机制融入非极大值抑制算法中,利用位置信息与框的得分信息相结合来调整框的最后得分。此外,还提出了改进的基于距离的交并比损失函数,重新定义了损失项,并将其引入到非极大值抑制中代替IOU来计算框间的交并比。最后将两种改进算法融合到3种经典的目标检测中,在Pascal-VOC 2012和MS-COCO 2017数据集上对上述两种算法进行了验证,结果表明检测精度得到了1%～2%的提升。

关键词：非最大抑制;目标检测;注意力机制
中图分类号：TP391　　**文献标识码**：A　　**国家标准学科分类代码**：510.4050

Research on non-maximum suppression based on attention mechanism in object detection

Zhang Changlun　Zhang Cuiwen　Wang Hengyou　He Qiang　Liu Xiwei
(School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)

Abstract: Non maximum suppression algorithm(NMS) is the main algorithm to select the accurate positioning box in object detection. The algorithm only takes the classification score as the standard, which may remove the prediction frame with low score but accurate positioning, and is more unfriendly to the situation with occlusion, A-NMS method is proposed, which integrates the attention mechanism into the non maximum suppression algorithm, and adjusts the final score of the box by combining the position information with the score information of the box. In addition, an improved distance based intersection union ratio loss function is proposed, the loss term is redefined, and it is introduced into non maximum suppression to calculate the intersection union ratio between frames instead of IOU. Finally, the two improved algorithms are integrated into three classical target detection. The above two algorithms are verified on Pascal-VOC 2012 and MS-COCO 2017 data sets. The results show that the detection accuracy has been improved by 1%～2%.

Keywords: non maximum suppression;object detection;attention mechanism

0　引　　言

随着计算机视觉在行人检测[1-2]、自动驾驶[3-4]和视频跟踪[5-6]等领域的广泛应用,作为其基本算法的目标检测技术得到了迅速发展。

早期的目标检测算法都是基于手动设计特征的,包括Viola-Jones(VJ)检测器[7]、hog-pedden检测器[8]、可变形零件模型[9](deformable part model, DPM)等,但这些方法并没有摆脱主观经验,因此检测的精度和速度都比较低。

Girshick等[10]提出了区域卷积网络(regional convolution network, R-CNN),这是基于深度学习的两阶段目标检测算法的开始。为了提高算法的速度,简化算法规模,文献[11-12]又提出了Fast R-CNN和Faster R-CNN,使两阶段算法实现了实时检测速度(17帧/s)和更高的检测精度(VOC2007将精度从70.0%提高到78.8%)。为了检测速度更快,Redmon等[13]提出了YOLO(you only look once)算法,它是第1个一阶段目标检测算法,速度远高于两阶段算法(原始算法达到45帧/s,快速版本速度155帧/s)。

收稿日期:2021-06-23
* 基金项目:国家自然科学基金项目(62072024)、北京建筑大学北京未来城市设计高精尖创新中心项目(UDC2017033322,UDC2019033324)、北京建筑大学市属高校基本科研业务费专项(X20084)资助

为了解决 YOLO 算法中的一系列问题,文献[14-15]又相继提出了 YOLO v2 和 YOLO v3,这不仅保持了算法的速度,而且进一步提高了算法的精度。注意力机制与目标检测算法并行发展,同时也应用于目标检测算法网络中。

人们在观察图像时,会有选择地提取图像的主要特征。受人类注意力的启发,文献[16]提出了视觉注意循环模块,并在循环神经网络中加入了注意力机制用于图像分类。随后,Bahdanau 等[17]提出了基于联合学习对齐和翻译的神经机器翻译,将注意力思想引入自然语言处理领域。卷积神经网络(convolutional neural network, CNN)是目标检测的主要算法,将注意机制引入 CNN 已成为研究热点。如 Hu 等[18]提出的 SENet (squeeze-and-excitation networks),在特征通道之间插入注意模块,根据损失函数学习特征权重,使无效或影响小的特征具有较小的权重,从而训练模型以获得更好的结果。另一个方向是在空间位置和特征通道之间插入注意模块,如 Fu 等[19]提出的 Danet,其中空间位置注意模块捕获特征图中任意两个位置之间的空间依赖性,对于特定特征,它由所有位置上的特征加权和更新。权重是对应两个位置之间的特征相似度,任何两个现有相似特征的位置都可以相互贡献,而不管它们之间的距离如何。此外,Li 等[20]提出的 SKNet 使用注意模块自适应选择卷积核的大小,Hu 等[21]提出关系模块,在对框进行分类和预测时不独立检测目标,但要了解目标之间的关系,提高监控效果。将该方法应用于非最大值抑制(non maximum suppression, NMS)中,寻求分数信息、特征向量和框间位置信息之间的关系,提高得分的准确率。

传统的 NMS 方法以分类得分作为排序的指标,选择得分最大的框与其他框进行 IOU 值比较。比率大于阈值的框将被删除。然而,这种方法的消除机制过于严格。为了改进这种方法,Bodla 等[22]提出了 soft-NMS 来代替将分数设置为 0 的 NMS。在遮挡的情况下,这种方法可能会导致对定位良好但分数较低的框的惩罚大于定位不好分数较高的框的惩罚。因此,Jiang 等[23]提出了一种基于位置优先级的 NMS 过滤方法,但该方法需要在网络中增加 IOU 预测分支,造成计算开销。此外,文献[24-25]提出基于加权法和方差加权平均法的方法,用于调整框的位置和分数。在遮挡的情况下,这些方法可能会导致去除定位准确但分类不准确的框。因此,Liu 等[26]提出了一种自适应阈值方法,通过增加预测模块自适应调整阈值。以上方法都是基于 IOU 的,只考虑了两框间的重叠区域,对框间重叠关系的描述不够准确。为了更准确地描述框间的交集,Zheng 等[27]提出了 DIOU(distance IOU),引入两框间的中心距离和重叠区域来描述框间的重叠关系,并将其应用于 NMS 来更好地划分阈值。框回归的损失函数也采用了框间的重叠关系进行描述。

在目标检测过程中,框回归是定位预测框的重要手段。为了更准确地预测目标的位置,框回归损失函数经历了一系列的发展。Girshick 等在 Fast R-CNN 中引入了平滑 L1 损失函数。该损失函数已被用作两阶段算法中的框回归优化算法。YOLO 系列采用均方误差函数。这两个损失函数的设计思想是优化预测框坐标与目标框坐标之间的距离。但实际上,评估预测框的指标是 IOU,这两个指标并不等价。多个预测框可能具有相同大小的平滑 L1 损失值,但是 IOU 是完全不同的。为了解决这个问题,Yu 等[28]引入 IOU 损失函数,最大化预测框与目标框之间的交并比,规范预测框的位置。当预测框和目标框之间没有重叠区域时,目标损失函数值为 1,因此无法执行梯度回传,且 IOU 不能完全描述两框的交集。Rezatofighi 等[29]提出了 GIOU (generalized IOU)来解决上述问题。然而,当预测框和目标框相互包含时,GIOU 退化为 IOU,收敛速度较慢。Zheng 等提出了 DIOU,将两框之间的距离加入到损失函数中进行优化,解决了两框之间相互包含并退化为 IOU 的问题,加快了收敛速度。

基于注意力机制的思想,对传统的 NMS 方法进行了改进,为框的选择提供了一种更为精确的方法,提出将注意力模块引入 NMS 的 A-NMS,通过寻找框间的位置关系来调整框的得分。然后,本文还提出了改进的基于距离的损失函数 IDIOU(improved distance IOU loss)来解决 DIOU 中不服从优化方向的问题。最后,为了评估本文算法,将其应用到 3 个经典的目标检测算法中,包括 YOLO v3、SSD[30]和 Faster R-CNN,并在 VOC 2012 和 MS-COCO 2017 两个公共数据集上进行了测试。本文第 1 部分回顾了近年来的目标检测算法和注意机制。第 2 部分介绍了相关工作。第 3 部分介绍了 A-NMS 的方法和流程框架。第 4 部分对描述框架重叠的 IOU 进行了分析,并提出了自己的见解。最后通过实验验证了改进算法的正确性。

1 相关工作

1.1 非极大值抑制

传统的 NMS 方法首先对同一类别中所有预测框的分类得分进行排序,然后选择得分最高的预测框与该类别中的其他预测框一起计算 IOU。将大于阈值的框删除,直到计算出所有框。然而,当对象被遮挡时,如图 1 所示,3 个虚线框是与目标类型相同的 3 个对象,实线框是将 3 个相似对象框在一起的框,但其 IOU 阈值是最大的。NMS 运行时,第 1 个要保留的框是 IOU 值为 0.9 的实线框。如果其他框与该框的 IOU 阈值大于所选阈值,则会删除定位准确但分类得分较低的 3 个虚线框,导致框选择不准确。

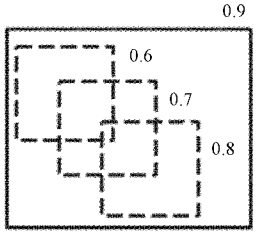
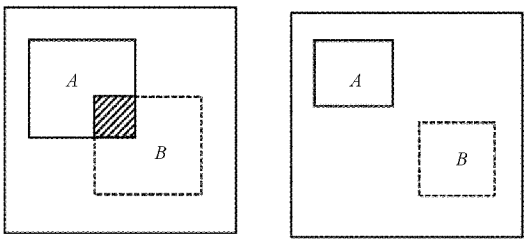


图 1 框间遮挡情况

1.2　DIOU

设置目标检测的损失函数设置为 1－IOU，如图 2(a)所示，最大化预测框和真实框之间的交集区域。然而，当两框不相交时，如图 2(b)所示，即两框的 IOU 值为 0，损失函数为 1，不能执行梯度回传，损失函数失效。



(a) 两框架相交　　(b) 两框架不相交

图 2　两框相交情况

Zheng 等提出了 DIOU，即基于距离的交并比。在损失函数中加入真实框和预测框中心点之间的距离，使两框相交面积最大化，且中心点之间的距离最小化。当两框不相交时，中心点之间的距离仍然可以作为梯度回传的惩罚。

DIOU 损失函数定义如式(1)所示。

$$L_{\mathrm{DIOU}}=1-\mathrm{IOU}+\frac{\rho^2(b,b_{\mathrm{gt}})}{C^2} \tag{1}$$

其中，$\rho^2(b,b_{\mathrm{gt}})$是预测框的中心点与目标框之间的距离，如图 3 虚线所示，是包括预测框与目标框之间的最小闭包对角线的距离。除了最大化相交部分的面积外，DIOU 损失函数还将两框中心点之间的距离加入损失函数，从面积和距离上标准化了预测框的位置，加快了损失函数的收敛速度。

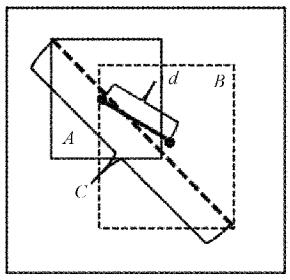


图 3　DIOU 优化

DIOU 损失函数的第 1 部分是最大化两框相交部分的面积，第 2 部分是最小化两框中心点之间的距离。损失函数将随着优化的进行惩罚两框之间的距离，从而缩短两框之间的距离。随着优化的进行，两框最小闭包对角线的距离变小，损失函数的第 2 部分变大，这违背优化方向，降低损失函数的优化速度。

2　A-NMS 方法

传统的 NMS 方法是对一个类别的所有预测框进行排序，然后找出得分最大的框，将剩余的框整合成一个集合 F 与得分最大的框进行 IOU 计算，最后从 F 中剔除，直到 F 变为空集。本文引入注意机制来寻找框间位置信息的相关性，然后进行 NMS 操作。具体过程如图 4 所示，矩阵 $\boldsymbol{A}=\mathrm{IOU}(M,F)\geqslant$ NMS-threhold 由得分最大的框 M 和 IOU 大于阈值的框组成。$\boldsymbol{A}'$由矩阵转置与原始矩阵相乘得到，其中 $\boldsymbol{C}$ 是所有框的得分矩阵。最后将矩阵 $\boldsymbol{C}$ 逐行进行 Softmax 归一化运算转化为权矩阵，并将权矩阵与得分矩阵矩阵相乘得到调整后的分数矩阵 $\boldsymbol{C}'$。这样就可以得到每个框调整后的新分数，然后进行 NMS 的正常运行。因此，可以将上述框架直接嵌入到 NMS 模块中，而不用在目标检测算法中引入额外的分支预测模块，以降低计算量。

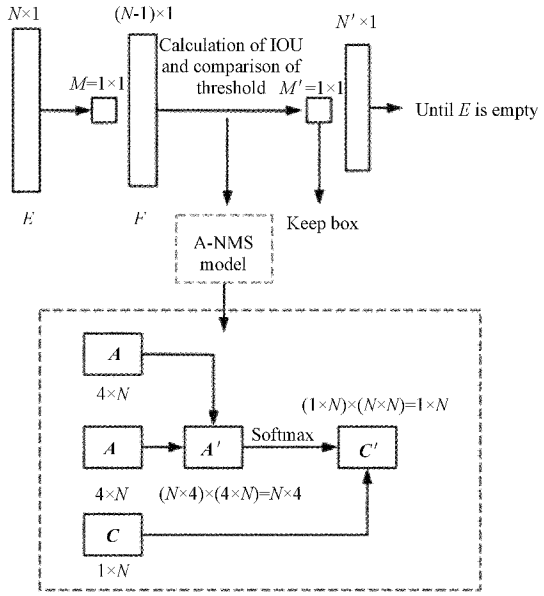


图 4　A-NMS 流程

本文将矩阵 $\boldsymbol{A}$ 与矩阵 $\boldsymbol{A}$ 的转置 $\boldsymbol{A}'$ 相乘得到相关性系数，也就是说，框之间的相似度越高，相关系数就越高，框的分数就越高。这样，对于类似的密集目标和遮挡目标，如图 1 所示，可以避免得分高但定位不准确的框。框的得分相对较低，但与得分大的框的 IOU 值大于所选阈值的框也被删除。将 A-NMS 模块插入 NMS 的具体过程如图 4 所示。

3　改进的 DIOU 损失函数

基于以上问题分析，提出了 IDIOU，如图 5 所示。将原公式中的对角线距离改进为两框相交部分的对角线距离。

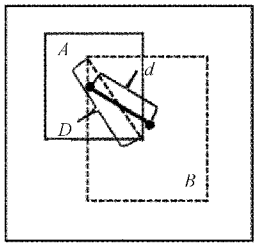


图 5　IDIOU 优化

具体公式如下:

$$L_{IDIOU}=\begin{cases}1-IOU+\dfrac{\rho^2(b,b_{gt})}{D^2}, & D\geqslant\rho^2(b,b_{gt})\\ 1-IOU+\dfrac{\rho^2(b,b_{gt})}{C^2}, & D<\rho^2(b,b_{gt})\end{cases}\tag{2}$$

其中,D 是两个框相交部分之间的对角线距离,如图 5 所示。当 D 的长度大于或等于两框中心点之间的距离时,损失函数使用式(2)中的上方子式,当 D 的长度小于两框中心点之间的距离时,损失函数使用式(2)中的下方子式。这是为了标准化对角线长度,使损失函数中的对角线长度小于 1。

在上述损失函数中考虑了两部分优化,即最大化两框相交部分的面积和最小化两框中心点之间的距离。利用式(2)作为损失函数,将原最小闭包的对角线长度改为相交部分的对角线长度,损失项随两框间距离的增大而变化。在损失函数的惩罚下,两框相交部分中心点之间的距离变长,两框之间的距离变短,收敛速度加快。将改进后的方法与 DIOU 损失函数进行了比较,对同一预测框进行了优化。优化速度如图 6 所示。

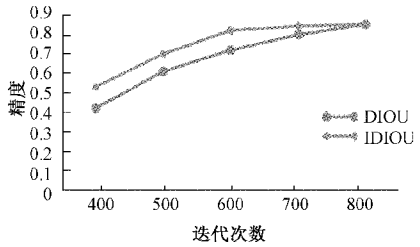


图 6 DIOU 和 IDIOU 的收敛速度

4 实验分析

本文将注意力机制引入 NMS 中,构建了 A-NMS 模块,并将其应用于 3 种目标检测框架:Faster R-CNN、SSD 和 YOLO v3,并将重新定义的指标应用于 NMS 的阈值。然后将 IDIOU 分别应用于 Faster R-CNN、SSD 和 YOLO v3 的丢失函数。改进方法在 Pascal-VOC 2012 和 MS-COCO 2017 两个数据集上进行了验证,其中 Pascal-VOC 2012 数据集共包含 11 540 幅图像,20 个类别,其中注释对象 27 450 个,其中 50%用于训练,50%用于测试。在训练和测试集中,图像和对象按类别分布大致相同。2017 年的 MS-COCO 训练集共有 80 个类别的 123 287 张图片,其中培训集 118 287 张,测试集 5 000 张。

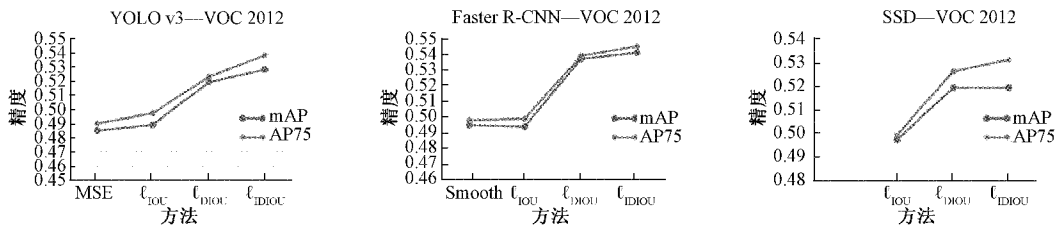
本文采用 2012 年最新定义的平均精度作为评价标准,即将 IOU 阈值设为{0.5, 0.55, …, 0.95}计算 AP 除以 10 得到 mAP,AP75 作为单独的评价标准。

4.1 Pascal-VOC 2012 实验结果比较

表 1 为基于 Faster R-CNN、SSD 和 YOLO v3 框架。损失函数由 MSE 和光滑 L1 分别变为 IOU、DIOU 和 IDIOU。计算了 VOC 2012 数据集的平均精度和阈值为 0.75 时的精度。从表 1 分析可以看出,在 YOLO v3 框架上,与 DIOU 相比,IDIOU 的平均精度增长分别为 0.9%,在 Faster R-CNN 框架下为 0.4%,在 SSD 框架下为 0%。也就是说,IDIOU 的平均精度提高在 0~0.9%之间。当 IOU 设置为 0.75 时,IDIOU 在 YOLO v3 框架上的精度提高了 1.5%,在 Faster R-CNN 框架上的精度提高了 0.6%,在 SSD 框架上的精度提高了 0.5%。IDIOU 的平均准确度提高了 0.5%~1.5%。与原方法相比,平均准确度和阈值 0.75 的准确度都有所提高。与基于欧氏距离的方法相比,改进后的方法平均精度提高了 2.2%~4.6%,也证明了基于面积和中心点距离的方法比基于坐标点的方法能更好地优化框的位置,提高了目标检测的精度。从以上分析可以看出,IDIOU 能够提高目标检测的测试精度,具有可行性和泛化能力。为了进一步反映该方法精度的提高,绘制了精度趋势图,如图 7 所示。

表 1 基于 IDIOU 的精度比较

YOLO v3——VOC 2012			Faster R-CNN——VOC 2012			SSD——VOC 2012		
方法/精度	mAP	AP75	方法/精度	mAP	AP75	方法/精度	mAP	AP75
MSE	0.485	0.490	Smooth	0.495	0.498			
ℓ_{IOU}	0.489	0.497	ℓ_{IOU}	0.494	0.499	ℓ_{IOU}	0.497	0.499
ℓ_{DIOU}	0.519	0.523	ℓ_{DIOU}	0.537	0.539	ℓ_{DIOU}	0.519	0.526
ℓ_{IDIOU}	**0.528**	**0.538**	ℓ_{IDIOU}	**0.541**	**0.545**	ℓ_{IDIOU}	**0.519**	**0.531**



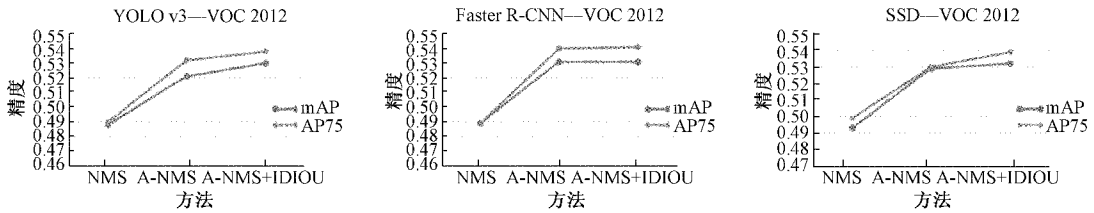
(a) 基于YOLO v3框架的精度趋势 (b) 基于Faster R-CNN框架的精度趋势 (c) 基于SSD框架的精度趋势

图 7 基于 VOC 2012 数据集的目标检测精度对比

表 2 中，在 NMS 中添加注意力机制，其中阈值是通过使用传统 IOU=0.5 来选择的。与传统的 NMS 方法相比，3 种目标检测框架的平均增长率分别为 3.3%、4.2%、3.6%。将重新定义的 IDIOU 损失函数添加到 A-NMS 中，可以更准确地描述框间的重合度。3 种目标检测框架的平均增长率为 0～0.9%，而与原 NMS 相比，增长率分别为 4.2%、4.2%和 3.9%，即平均增长 4%。因此，该方法在目标检测中的精度有了显著提高。精度趋势如图 8 所示。

表 2　基于 A-NMS 模块的精度比较

方法/精度	YOLO v3——VOC 2012		Faster R-CNN——VOC 2012		SSD——VOC 2012	
	mAP	AP75	mAP	AP75	mAP	AP75
NMS	0.488	0.490	0.489	0.489	0.493	0.499
A-NMS	**0.521**	**0.532**	**0.531**	**0.540**	**0.529**	**0.530**
A-NMS+IDIOU	0.530	0.538	0.531	0.541	0.532	0.539



(a) 基于YOLO v3框架的精度趋势　(b) 基于Faster R-CNN框架的精度趋势　(c) 基于SSD框架的精度趋势

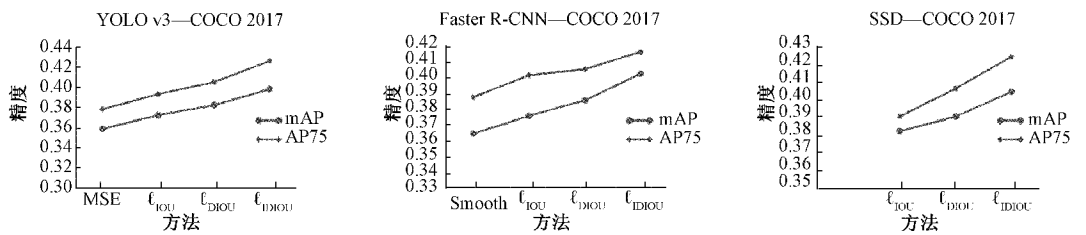
图 8　基于 VOC 2012 数据集的 NMS 精度对比

4.2　MS-COCO 2017 精度对比

与表 3 中的 DIOU 相比，在 YOLO v3 框架下，IDIOU 的平均准确率提高了 1.6%，在 Faster R-CNN 框架下提高了 1.7%，在 SSD 框架下提高了 1.4%。当 IOU 阈值设置为 0.75 时，该方法分别提高了 1.1%、1.1%和 1.8%。与基于欧氏距离的方法相比，改进后的方法精度提高了 2.2%～3.9%。根据 COCO 2017 数据集绘制了目标检测精度对比图，如图 9 所示。

表 3　基于 IDIOU 的精度比较

YOLO v3——COCO 2017			Faster R-CNN——COCO 2017			SSD——COCO 2017		
方法/精度	mAP	AP75	方法/精度	mAP	AP75	方法/精度	mAP	AP75
MSE	0.353	0.372	Smooth	0.361	0.384	——		
ℓ_{IOU}	0.366	0.387	ℓ_{IOU}	0.372	0.398	ℓ_{IOU}	0.379	0.387
ℓ_{DIOU}	0.376	0.399	ℓ_{DIOU}	0.382	0.402	ℓ_{DIOU}	0.387	0.403
ℓ_{IDIOU}	**0.392**	**0.420**	ℓ_{IDIOU}	**0.399**	**0.413**	ℓ_{IDIOU}	**0.401**	**0.421**



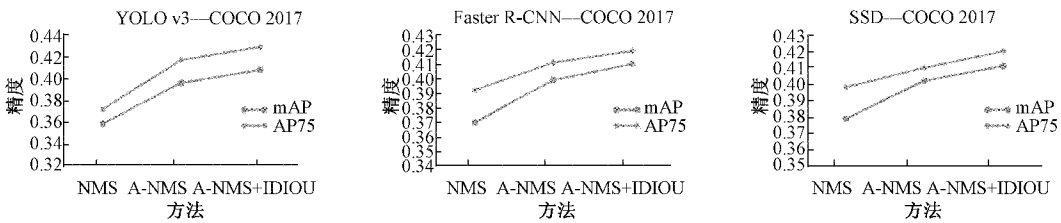
(a) 基于YOLO v3框架的精度趋势　(b) 基于Faster R-CNN框架的精度趋势　(c) 基于SSD框架的精度趋势

图 9　基于 COCO 2017 数据集的目标检测精度对比

COCO 2017 数据集中的 A-NMS 的精度如表 4 所示。与原始 NMS 方法相比，3 种目标检测框架中 A-NMS 的平均增长率分别为 3.7%、2.9%和 2.3%。A-NMS+IDIOU 3 种目标检测框架的平均增长率分别为 1.2%、1.1%和 0.9%，比原 NMS 分别增长 4.9%、4%和 3.2%，平均增长 6%。精度趋势如图 10 所示。

表 4　基于 A-NMS 模块的实验结果比较

方法/精度	YOLO v3——COCO 2017		Faster R-CNN——COCO 2017		SSD——COCO 2017	
	mAP	AP75	mAP	AP75	mAP	AP75
NMS	0.359	0.372	0.370	0.392	0.379	0.398
A-NMS	**0.396**	**0.417**	**0.399**	**0.411**	**0.402**	**0.410**
A-NMS+IDIOU	0.408	0.429	0.410	0.419	0.411	0.420



(a) 基于YOLO v3框架的精度趋势　(b) 基于Faster R-CNN框架的精度趋势　(c) 基于SSD框架的精度趋势

图 10　基于 COCO 2017 数据集的 NMS 精度对比

5　结　　论

本文提出了一种用于目标检测的 A-NMS 方法。它利用框间的坐标信息进行相关性提取和框分数调整,解决了分数低但定位准确的框作为冗余框被剔除的问题。为了准确地描述框的重合信息,提出了 IDIOU 并将其引入框回归损失函数中。在 DIOU 的基础上,通过优化选择对角线距离的定义方法,将 DIOU 中最小闭包的对角线距离转化为相交部分的对角线距离。改进后的方法提高了计算精度,表明了改进方法的有效性。但在实验过程中,将 A-NMS 与 IDIOU 一起应用于经典的目标检测算法中后,对比于原始经典目标检测算法精度得到提升,但速度降低,在后续的研究中,应在保持精度提升的前提下提升检测速度。

参考文献

[1] 廖畅,马秀丽. 基于 TOF 相机的俯视行人检测[J]. 电子测量技术,2019,42(6):109-113.

[2] 王颖, 金若辰, 金志刚. 支持行人检测的智能车载监控终端[J]. 电子测量技术, 2019,42(6):17-21.

[3] LI Y P, HOU L Y, WANG CH. Moving objects detection in a automatic driving based on YOLOv3[J]. Computer Engineering Design, 2018, 40 (12): 2812-2819.

[4] SHEN L F, ZHANG R, ZHU Y P, et al. High-precision and real-time localization algorithm for automatic driving vehicles[J]. Computer Engineering Design, 2020(1):28-35.

[5] 李力行,黄永梅,王强,等.点目标视频跟踪中的噪声自适应卡尔曼滤波器[J]. 电子测量技术,2017,40(6):170-174.

[6] HUANG H, BI D, GAO SH, et al. Visual tracking via locality-sensitive kernel sparse representation[J]. Computer Engineering Design, 2016, 38 (4): 993-999.

[7] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [C]. IEEE Conference on Computer Vision and Pattern Recognitino, 2001:511-518.

[8] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2005:886-893.

[9] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model[C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2008: 1984-1991.

[10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Computer Vision and Pattern Recognition, IEEE, 2013:580-587.

[11] GIRSHICK R. Fast R-CNN[C]. IEEE, 2015,ArXiv: 1504.08083, 2015.

[12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time objectdetection with region proposal networks[C]. IEEE, 2015,ArXiv:1506.01497, 2015.

[13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016:779-788.

[14] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 6517-6525.

[15] REDMON J, FARHADI A. YOLO v3: An incremental improvement[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2018, ArXiv:1804.02767,2018.

[16] MNIH V, HEESS N, GRAVES A, et al. Recurrent models of visual attention[C]. Advances Inneural Information Processing Systems, 2014:2204-2212.

[17] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. Computer Ence, 2014, DOI: 10.1007/S00521-021-06444-2.

[18] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.

[19] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, DOI: 10.1109/CVPR.2019.00326.

[20] LI X, WANG W, HU X, et al. Selective kernel networks[J]. CVPR,2019,DOI: 10.1049/CIT2.12008.

[21] HU H, GU J, ZHANG Z, et al. Relation networks for object detection[J]. 2017, DOI: 10.1007/S11276-020-02391-3.

[22] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS—improving object detection with one line of code[J]. ICCV,2017,DOI: 10.1016/j.patcog.2017.107131.

[23] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[J]. ArXiv,2018,ArXiv:1807.11590.

[24] NING C, ZHOU H, SONG Y, et al. Inception single shot multibox detector for object detection[C]. 2017 IEEE International Conference on Multimedia & Expo Workshops(ICMEW), IEEE, 2017:549-554.

[25] HE Y, ZHU C, WANG J, et al. Softer-NMS: Rethinking bounding box regression for accurate object detection[J]. ArXiv Preprint,2018, ArXiv: 1809.08545.

[26] LIU S, HUANG D, WANG Y. Adaptive NMS: Refining pedestrian detection in a crowd[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019:6452-6491.

[27] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020,34(7):12993-13000.

[28] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network[C]. Proceedings of the 2016 ACM on Multimedia Conference, 2016:516-520.

[29] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bound in regression[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2019:658-666.

[30] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[J]. ECCV,2016:21-37.

作者简介

张长伦,副教授,主要研究方向为深度学习。

E-mail:zclun@bucea.edu.cn

张翠文,硕士,主要研究方向为机器学习与深度学习在目标检测中的应用。

E-mail:1074102915@qq.com

王恒友,副教授,主要研究方向为人脸识别、深度学习。

E-mail:wanghengyou@bucea.edu.cn

何强,副教授,主要研究方向为机器学习、深度学习。

E-mail:heqiang@bucea.edu.cn

刘屹伟,硕士,主要研究方向为深度学习。

E-mail:liuyiwei@bucea.edu.cn