

基于 YOLOv3 的光学遥感图像目标检测算法

汪鹏^{**}, 辛雪静^{*}, 王利琴, 刘瑞

河北工业大学人工智能与数据科学学院, 天津 300100

摘要 针对目前算法对遥感图像中背景复杂、目标小而密集的复杂场景下的目标检测精度低的问题,提出了一种基于 YOLOv3 的改进算法,在 YOLOv3 的基础上,结合了密集连接网络,利用密集连接块来提取深层特征,增强特征传播,同时引入 Distance-IoU(DIoU) loss 作为坐标预测的损失函数,使边界框的定位更加准确,此外针对目标间相互遮挡的情况,改进了传统的非极大值抑制算法,使用 DIoU 代替 IoU 来克服虚假抑制的问题。对所提算法在三个经典的遥感数据集上进行了测试,实验结果证明,所提方法具有更高的检测精度。

关键词 遥感; 遥感图像; 卷积神经网络; 目标检测; YOLOv3 网络; 密集连接

中图分类号 TP753

文献标志码 A

doi: 10.3788/LOP202158.2028006

Object Detection Algorithm of Optical Remote Sensing Images Based on YOLOv3

Wang Peng^{**}, Xin Xuejing^{*}, Wang Liqin, Liu Rui

School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300100, China

Abstract To solve the problems of low detection accuracy for remote sensing images in complex scenes with complex background and small and dense objects, an improved YOLOv3 algorithm is proposed in this paper. Based on YOLOv3, our algorithm is combined with dense connection network (DenseNet) and uses the dense connection blocks to extract deep features and enhance feature propagation. Meanwhile, Distance-IoU (DIoU) loss is introduced as the loss function of coordinate prediction, making the location of the bounding box more accurate. Besides, aiming at the situation of mutual occlusion between targets, we use DIoU instead of IoU in the improved non-maximum suppression algorithm to overcome the problem of false suppression. The proposed algorithm is tested on three classical remote sensing datasets, and the experimental results show that the detection method in this paper has higher detection accuracy.

Key words remote sensing; remote sensing image; convolutional neural network; object detection; YOLOv3 network; dense connection

OCIS codes 280.4750; 100.5010; 100.3008; 100.2000

1 引言

近年来,随着遥感技术的不断发展,遥感卫星提供了大量高分辨率的遥感图像,遥感图像目标检测也在军事和民用领域得到了广泛的应用,面对当前大量多样化的遥感数据的生成,如何更精确地检测

所需要的目标区域并将其投入到应用中,这是一个亟待解决的问题。

随着深度学习的发展,卷积神经网络在目标检测领域得到了广泛的应用,并出现了越来越多的目标检测算法。比如经典的 Faster RCNN^[1] 算法自提出之后在遥感图像领域获得了很好的运用。徐志

收稿日期: 2020-11-11; 修回日期: 2020-12-07; 录用日期: 2021-01-20

基金项目: 天津市自然科学基金重点项目(19JCZDJC40000)、天津市科技计划项目(18YFCZCC00060, 18ZXZNGX00100)

通信作者: *1306014217@qq.com; **wangpeng1027@126.com

京等^[2]基于双路组合的网络提取特征,提出了一种自适应旋转区域生成网络,提升了遥感图像中舰船的检测精度。欧攀等^[3]在 Faster RCNN 的基础上,提出了一种具有旋转不变性自学习能力的模型,以解决遥感图像中旋转角度过大的问题。这些算法虽然在一定程度上提高了遥感图像的检测精度,但是,由于 Faster RCNN 是两阶段检测算法,需要先提取候选区域,因此它的网络结构比较复杂,参数较多,检测速度慢。在此背景下,YOLO^[4]系列的单阶段算法受到了更多学者的青睐。YOLO 算法直接从输入图像中获取预测结果,省去了区域检测过程,从而大幅提高了算法的检测速度。目前 YOLO 系列算法中应用较广泛的是 YOLOv3^[5]算法,与之前的版本相比,该算法在检测精度和速度上都取得了更好的结果。李成跃等^[6]对 YOLOv3 进行改进,在略微降低精度的同时,大幅度提升了检测速度。Li 等^[7]在 YOLOv3 的基础上,采用多级特征提取的方法提高了遥感图像中舰船的检测精度。Liu 等^[8]对 YOLOv3 的目标检测层进行修改,并使用滑动窗口方法将原始图像划分成多个固定的小尺寸图像用于检测,以减少漏检情况。

上述改进算法虽然都取得了一定的效果,但是 YOLOv3 仍然存在特征利用率低的缺点,致使网络在特征提取过程中容易丢失小目标的信息,此外,该算法在复杂场景下的检测精度也并不理想。综上,本文提出了一种基于 YOLOv3 的遥感图像目标检测算法。为了加强特征传递,本文引入 DenseNet 结构作为骨干网络的连接方式,并在边界框回归时使用 DIoU loss 解决了原损失函数的不足,此外还将 DIoU 代替 IoU 作为 NMS 算法的评价标准,并通过调整锚点框来进一步优化模型。

2 基本原理

2.1 相关工作

YOLO 是基于回归的单阶段算法,它通过网络处理,直接输出目标的位置和所属类别。YOLOv3 是 Redmon 等^[5]在 2018 年提出的,其主干网络是 Darknet-53,其结构如图 1 所示。Darknet-53 借鉴了 ResNet 的残差结构,其中 1, 2, 8, 8, 4 代表重复的残差模块的个数。该网络大量使用 3×3 与 1×1 的卷积层依次连接的形式,整个网络共由 53 个卷积层组成,为了改善池化带来的负面效果,该网络使用步长为 2 的卷积进行降采样。

YOLO 算法是将输入图像划分为 $K \times K$ 的网

	Type	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	$3 \times 3/2$	128×128
1×	Convolutional	32	1×1	128×128
	Convolutional	64	3×3	
	Residual			
	Convolutional	128	$3 \times 3/2$	64×64
2×	Convolutional	64	1×1	64×64
	Convolutional	128	3×3	
	Residual			
	Convolutional	256	$3 \times 3/2$	32×32
8×	Convolutional	128	1×1	32×32
	Convolutional	256	3×3	
	Residual			
	Convolutional	512	$3 \times 3/2$	16×16
8×	Convolutional	256	1×1	16×16
	Convolutional	512	3×3	
	Residual			
	Convolutional	1024	$3 \times 3/2$	8×8
4×	Convolutional	512	1×1	8×8
	Convolutional	1024	3×3	
	Residual			
	Avgpool		Global	
	Connected		1000	
	Softmax			

图 1 Darknet-53 结构

Fig. 1 Darknet-53 structure

格,如果一个物体的中心落入网格单元,该网格单元就负责检测该物体,在预测过程中,每个网格需要预测一定数量的边界框,包括预测边界框的位置信息、置信度及类别预测,最后利用极大值抑制算法(NMS)去掉重复的边界框,从中选择最好的结果。为了提高小物体检测的准确性,YOLOv3 引入了类似特征金字塔网络的概念,使用多尺度融合进行检测,共生成了 3 种不同比例的目标检测层,若将输入的图像归一化为 416×416 ,则三种比例的特征图大小分别为 13×13 、 26×26 和 52×52 ,YOLOv3 给每种比例的特征图设定了三种不同尺寸的锚点框(Anchor box),共聚类了 9 个锚点框。COCO 数据集上的 9 个锚点框分别是 $(10, 13)$, $(16, 30)$, $(33, 23)$, $(30, 61)$, $(62, 45)$, $(59, 119)$, $(116, 90)$, $(156, 198)$, $(373, 326)$ 。

2.2 实验原理与方法

2.2.1 结合密集连接结构

YOLOv3 中使用 ResNet 作为基本网络,ResNet 模型的核心是通过建立前面层与后面层之间的跳跃连接,使网络更深。在此思想的基础上,为了最大化网络中所有层之间的信息流,本文使用 DenseNet 网络结构的连接方式,DenseNet 将网络中的所有层都两两进行了连接,使得网络中每一层

都接收它前面所有层的特征作为输入,并且它自己的特征也被用作所有后续层的输入,本文借鉴 DenseNet 的密集连接结构, YOLOv3 各模块除第一个残差块外,均使用密集连接块代替原有的残差结构,密集连接结构一方面使特征的传输更加有效,另一方面可以缓解卷积神经网络中经常出现的梯度消失的问题。

本文中以 1×1 卷积、 3×3 卷积和 Leaky ReLU 作为网络的一个组成部件,则两种结构的连接方式对比如图 2、3 所示。

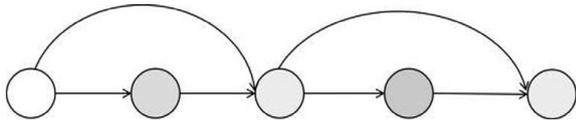


图 2 残差结构连接方式

Fig. 2 Residual structure connection mode

如图 3 所示, DenseNet 网络中所有的模块之间都两两互联,其密集的连接方式有利于最大程度地进行特征传输。所有模块在传统的 L 层卷积神经网络中具有 L 个连接,但在 DenseNet 中会有 $L(L+1)/2$ 个连接,每个密集连接块可表示为

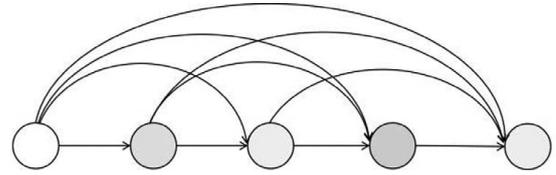


图 3 密集连接结构

Fig. 3 Dense connection structure

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \quad (1)$$

式中: x_0, x_1, \dots, x_{l-1} 表示各层生成的特征图的连接,使第 l 层接收所有先前的特征图,并将其输入;定义 $H_l(\cdot)$ 由卷积层、批标准化 (BN) 和 Leaky ReLU 三部分组成。

改进后的网络结构如图 4 所示,骨干网络由 1 个残差块和 4 个密集连接块构成,进行一系列卷积操作后得到第一个尺度的目标检测层,再执行上采样操作,然后将结果分别与第 2 个和第 3 个密集连接块的特征图融合,从而实现多尺度检测。本文神经网络结构中使用密集连接结构替代原始网络的连接方式,实现了特征复用,减少了模块之间特征传递的损失。

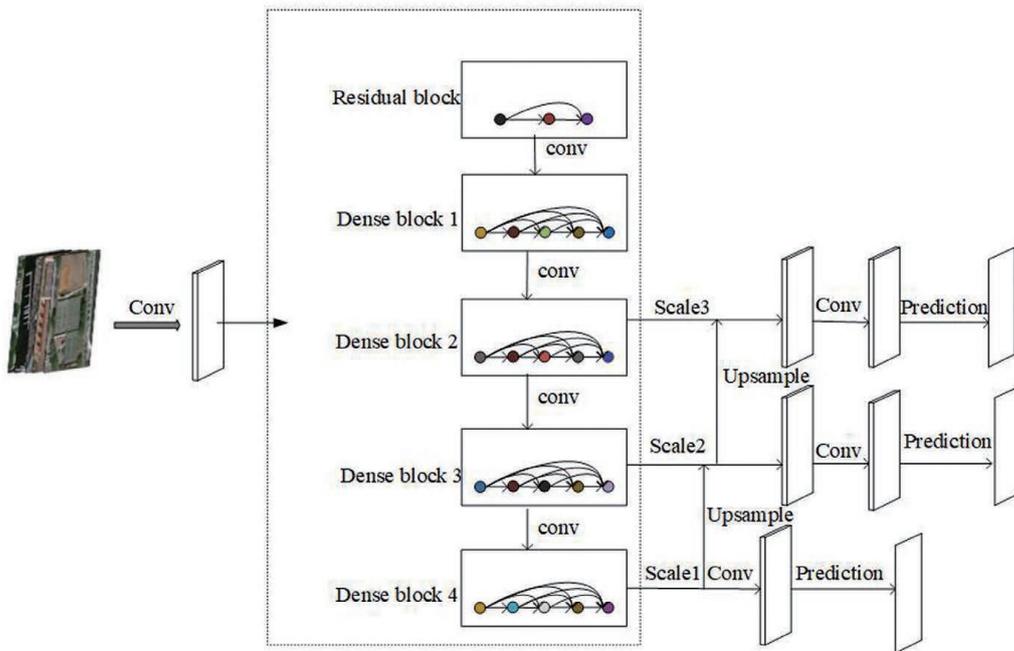


图 4 改进后的网络结构

Fig. 4 Improved network structure

2.2.2 使用 DIoU loss 作为边界框回归的损失函数

边界框回归是目标检测中重要的一部分,原始 YOLOv3 根据 L2 范数计算边界框损失函数,并利用 IoU 作为评测标准。IoU 为预测框和真实框的交集与并集的比例(其大小可用 R_{IoU} 表示),可用于

评价预测框的优劣。 R_{IoU} 可表示为

$$R_{IoU} = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}, \quad (2)$$

式中: B 为预测框; B^{gt} 为真实框。

但 L2 范数与 IoU 并不完全一致,故还有待优

化。之后研究人员提出直接使用 IoU loss 作为损失函数,通过 IoU 计算 loss 并进行回归操作,损失函数的计算式为

$$L_{IoU} = 1 - R_{IoU} \quad (3)$$

此时存在两个问题:1)当两个框之间无交集时,很近与很远的无交集框的 IoU 值均等于 0,这种情况下无法进行进一步优化;2)假设预测框和真实框的 IoU 值是相同时,IoU 值不能反映两个框的相交情况,它只适用于两框之间有重叠的情况。针对这些问题,Rezatofighi 等^[9]提出了 GIoU 这一概念,它考虑到了非重叠的区域,克服了 IoU 的缺点。GIoU loss 可表示为

$$L_{GIoU} = 1 - R_{IoU} + \frac{|C - B \cup B^{gt}|}{|C|} \quad (4)$$

式中: C 为包含 B 与 B^{gt} 的最小矩形框。

但是 GIoU 仍然存在收敛速度慢、回归不准确的问题。综上,Zheng 等^[10]提出了 Distance-IoU (DIoU) loss,相对于 GIoU loss,DIoU loss 不仅考虑了非重叠区域,还考虑了中心点距离,使得目标框回归变得更加稳定且准确。DIoU loss 可表示为

$$L_{DIoU} = 1 - R_{IoU} + R_{DIoU} \quad (5)$$

其中

$$R_{DIoU} = \frac{\rho^2(b, b^{gt})}{c^2} \quad (6)$$

式中: b, b^{gt} 分别表示 B, B^{gt} 的中心点; $\rho(\cdot)$ 表示欧氏距离; c 表示最小外接矩形的对角线距离。DIoU loss 可以直接最小化两个目标框间的距离,因此比 GIoU loss 收敛速度快。此外,两个框在水平方向和垂直方向上的对比结果如图 5 所示(其中,小框与大框分别代表真实框与预测框),当目标框完全包裹真实框时,即使两个框的位置不同, L_{IoU} 和 L_{GIoU} 的值也相等,均为 0.75,此时 GIoU 退化为 IoU,无法区分两框的相对位置关系,从而导致回归不准确,而 DIoU 弥补了这个不足。

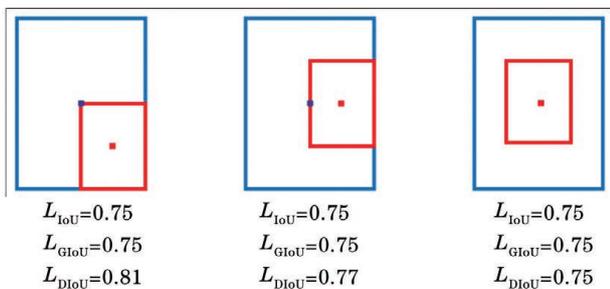


图 5 L_{IoU}, L_{GIoU} 和 L_{DIoU} 的对比示意图

Fig. 5 Comparison of $L_{IoU}, L_{GIoU},$ and L_{DIoU}

2.2.3 改进的非极大值抑制算法

通常在预测过程中,多个预测框可能围绕同一目标,造成大量重叠,为了筛选预测框,根据预测框的目标分类得分,使用非极大值抑制(NMS)算法消除得分较低的候选框,从而减少冗余框。本质上,NMS 是一种迭代遍历的消除过程,通过设置 IoU 阈值可大幅减少候选框的数量。

原始的 NMS 算法可表示为

$$s_i = \begin{cases} s_i, & L_{IoU}(M, b_i) < T \\ 0, & L_{IoU}(M, b_i) \geq T \end{cases} \quad (7)$$

式中: s_i 为分类得分; M 为得分最高的预测框; b_i 为除 M 之外的其他预测框; T 是 NMS 的阈值参数,本文取 $T=0.45$ 。

但是,IoU 只考虑重叠区域,多个目标互相遮挡时,IoU 可能会对遮挡情况产生虚假抑制,误删有目标的预测框,而 DIoU 不仅考虑重叠区域,还考虑两个框中心点之间的距离,与 IoU 相比能更好地度量遮挡情况。因此,本文使用 DIoU 来替换普通的 IoU 评价策略,并将其应用于 NMS 中。

DIoU-NMS 的定义为

$$s_i = \begin{cases} s_i, & L_{IoU} - R_{DIoU}(M, b_i) < T \\ 0, & L_{IoU} - R_{DIoU}(M, b_i) \geq T \end{cases} \quad (8)$$

3 实验结果分析

3.1 实验数据集

为了验证本文方法的有效性,本文在以下三个经典的遥感图像目标检测数据集上进行了实验。

DIOR 数据集^[11]是近年发布的具有挑战性的大规模数据集,该数据集包含 23463 张图像和 190288 个实例,覆盖 20 个目标类别,每个类别包含约 1200 幅图像。与其他目标检测数据集相比,DIOR 数据集总体图像、目标类别和目标实例的数量都很大,且跨物体的类别间和类别内尺寸变化范围也较大,小物体较为密集。其中,随机选取了 11725 幅遥感图像(即 50%的数据集)作为训练集和验证集,剩下的 11738 幅图像作为测试集(按照原始数据集中对训练集和测试集的划分),各个类别的分布如表 1 所示。

RSOD 数据集^[12-13]共包含飞机、操场、立交桥和油桶 4 类目标,一共 976 张图像,主要包括 446 张图像中的 4993 架飞机,189 张图像中的 191 个操场,176 张图像中的 180 座立交桥和 165 张图像中的 1586 个油桶,其中有 936 个带注释的图像,本文大致将该数据集分为 25%的训练集、25%的验证集和

50% 的测试集,其中具体的分布如表 2 所示。

表 1 各个类别的训练集、验证集及测试集的图像数

Table 1 Numbers of images in training set, verification set, and test set of each category

Category	Training set	Verification set	Test set
Airplane	344	338	705
Airport	326	327	657
Baseball field	551	577	1312
Basketball court	336	329	704
Bridge	379	495	1304
Chimney	202	204	448
Dam	238	246	502
Expressway service area	279	281	565
Expressway toll station	285	299	634
Golf course	216	239	491
Ground track field	536	454	1322
Harbor	328	332	814
Overpass	410	510	1099
Ship	650	652	1400
Stadium	289	292	619
storage tank	391	384	839
Tennis court	605	630	1347
Train station	244	249	501
Vehicle	1556	1558	3306
Wind mill	404	403	809
Total	5862	5863	11738

表 2 各个类别的训练集、验证集及测试集的图像数

Table 2 Numbers of images in training set, verification set, and test set of each category

Category	Training set	Verification set	Test set
Playground	37	34	78
Overpass	44	44	88
Oiltank	41	41	83
Aircraft	111	112	223
Total	233	231	472

TGRS-HRRSD 数据集^[14]是在 2018 年被发布,用于研究高分辨率遥感图像目标检测的数据集。该数据集共有 21761 张图像和 55740 个对象实例,包含 13 类物体。本文从此数据集中选取了 500 张图片,其中 300 张为飞机图片,200 张为轮船图片,并随机选取 50% 作为训练集和验证集,其余作为测试集。具体的分布如表 3 所示。

表 3 各个类别的训练集、验证集及测试集的图像数

Table 3 Numbers of images in training set, verification set, and test set of each category

Category	Training set	Verification set	Test set
Airplane	75	75	150
Ship	50	50	100
Total	125	125	250

3.2 参数设置

本实验中,为了更好地检测小物体,将输入图像规格化为 608×608 ,此操作会增加检测时间,但会大大提高检测精度。在模型训练阶段,初始学习率设置为 10^{-4} ,结束学习率为 10^{-6} ,迭代次数为 10^5 ,优化方式为动量优化, momentum 取值为 0.9。

本文使用 K-means 算法分析数据集,旨在生成适用于所使用数据集的锚点框。YOLOv3 中通常设定 9 个锚点框,故本文选取数量为 1~9 的锚点框对训练集进行聚类,得到锚点框数量与其平均 IoU 值的关系图,如图 6~8 所示。

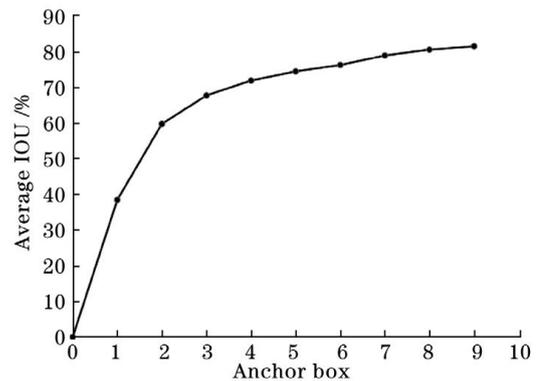


图 6 RSOD 遥感图像数据集聚类结果图

Fig. 6 Clustering results of RSOD remote sensing image dataset

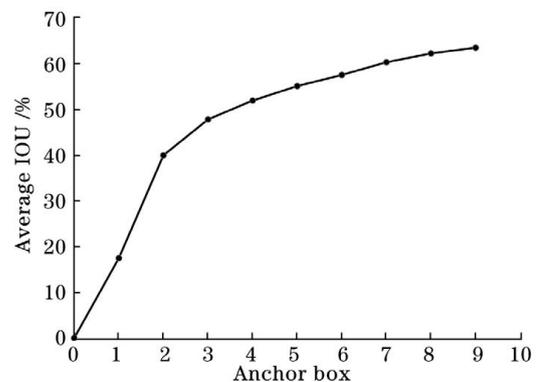


图 7 DIOR 遥感图像数据集聚类结果图

Fig. 7 Clustering results of DIOR remote sensing image dataset

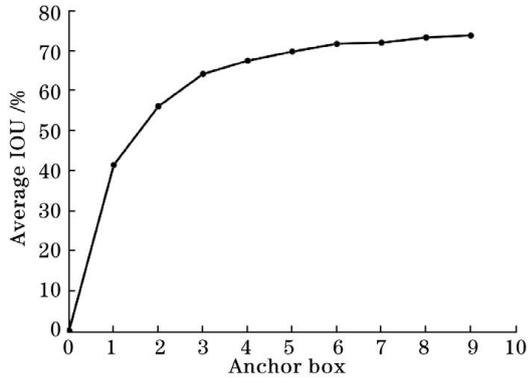


图 8 TGRS-HRRSD 部分遥感图像数据集聚类结果图

Fig. 8 Clustering results of partial TGRS-HRRSD remote sensing image dataset

根据折线图,平均 IoU 值均呈上升趋势,尤其是 DIOR 和 RSOD 数据集,锚点框数量达到最大之前上升趋势还较明显,为达到较好的检测结果,本文为这两个数据集设定 9 种不同比例的锚点框。在 RSOD 数据集上,这 9 组数据分别为(12,13),(17,17),(23,24),(30,29),(36,38),(45,48),(59,65),(75,86),(213,262),相比于使用原始的锚点框,mAP(mean Average Precision)值提高了将近 1 个百分点。但是在 DIOR 数据集上直接使用 K-means 算法聚类得到锚点框的效果并不好,本文对聚类得到的锚点框进行了放大和缩小,根据经验选取一组效果较好的锚点框:(9,6),(10,14),(27,12),(15,27),(45,27),(25,51),(63,75),(131,154),(339,344),聚类得到的锚点框相比于原始锚点框的 mAP 值提高了 0.6 个百分点。对于选取的部分 TGRS-HRRSD 数据集,其平均 IoU 值在锚点框为 6 之后,上升较为平缓,故选取锚点框数量为 6,尺度 1 上对应(37,36)和(60,62),尺度 2 上对应(92,92)和(140,129),尺度 3 上对应(207,201)和(86,247)。

3.3 实验结果与分析

3.3.1 评价标准

在目标检测中通常使用 AP (Average Precision)与 mAP 作为评价指标。AP 值是衡量某一个类别检测结果的好坏,它是利用召回率(R)和精确度(P)计算得到的。mAP 值是对多个类别检测结果的衡量,它的计算与 AP 值有关。本实验采用 AP 值与 mAP 值来共同评价算法的有效性。AP 值与 mAP 值的具体计算公式如下。

精确度的计算式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (9)$$

式中: N_{TP} 为被正确分类的正例样本数; N_{FP} 为本来

是负例,被错分为正例的样本数。

召回率的计算式为

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (10)$$

式中: N_{TP} 为被正确分类的正例样本数; N_{FN} 为本来是正例,被错分为负例的样本数。

AP 值为以精确度和召回率作为纵横坐标即 P - R 曲线围成的面积,其计算式为

$$A_P = \int_0^1 P(r) dr. \quad (11)$$

再对所有 AP 值取平均得到 mAP 值:

$$V_{mAP} = \frac{\sum A_P}{N_{all_classes}}, \quad (12)$$

式中: $N_{all_classes}$ 为样本总数。此外,本文还使用检测时间来评估算法的检测速度,通过统计算法中每张图片的检测时间来对比算法的性能。

3.3.2 实验结果

为了验证本文算法的有效性,分别以两种参数的 SSD 网络、Faster RCNN 网络、原始的 YOLOv3 网络和本文方法在上述三个遥感数据集上进行实验,表 4~6 分别列出了检测结果的对比。

从表 4~6 中可以看出,在 DIOR 数据集中,本文方法与 SSD(300×300)、SSD(512×512)、Faster RCNN 和原始的 YOLOv3 相比,mAP 值分别提升了 11.4, 4.0, 8.5, 5.5 个百分点,在 RSOD 数据集中分别提高了 4.7, 1.1, 2.3, 2.8 个百分点,在部分 TGRS-HRRSD 数据集上分别提高了 6.1, 4.7, 5.1, 4.3 个百分点。从各类别的 AP 值来看,本文算法增强了原算法对小目标检测的优势,从表 4 中可以看出,本文算法对车辆、储罐、船舶、飞机的检测结果相比原始的 YOLOv3 都得到了提高,也明显高于 Faster RCNN 以及其他算法。此外,表 4 中本文算法对机场和立交桥的检测结果相比原始的 YOLOv3 提高尤为明显,分别提升了 10.1 和 10.0 个百分点。这是因为本文算法提高了特征的利用率,能更大程度地增强特征传输,减少信息在传输过程中的丢失,并使用更合理有效的 DIoU loss,使边界框回归过程中网络的优化方向更加准确,所以本文算法在复杂场景下更具有优势,但是也有某些类别的检测结果与其他算法还有一定差距,这有待后续的研究。在检测时间上,SSD(300×300)的检测时间最短,但是检测精度不佳,本文算法虽然比 YOLOv3 算法的检测时间长,但是远少于 Faster RCNN 的检测时间,基本也可满足实时检测。

表 4 不同模型在 DIOR 数据集的检测结果

Table 4 Test results of different models on DIOR dataset

Parameter	SSD	SSD	Faster	YOLOv3 ^[5]	Ours
	(300×300) ^[15]	(512×512) ^[15]	RCNN ^[1]		
Airplane	49.1	59.5	53.6	72.2	75.9
Airport	62.1	72.7	49.3	29.2	39.3
Baseball field	66.2	72.4	78.8	74.0	77.8
Basketball court	72.0	75.7	66.2	78.6	82.7
Bridge	26.1	29.7	28.0	31.2	38.6
Chimney	63.3	65.8	70.9	69.7	72.7
Dam	54.0	56.6	62.3	26.9	34.9
Expressway service area	62.7	63.5	69.0	48.6	53.7
Expressway toll station	46.6	53.1	55.2	54.4	58.5
Golf course	64.8	65.3	68.0	31.1	40.8
Ground track field	53.1	68.6	56.9	61.1	64.2
Harbor	44.2	49.4	50.2	44.9	48.3
Overpass	34.7	48.1	50.1	49.7	59.7
Ship	44.4	59.2	27.7	87.4	91.7
Stadium	58.3	61.0	73.0	70.6	74.9
Storage tank	42.1	46.6	39.8	68.7	73.2
Tennis court	72.6	76.3	75.2	87.3	90.9
Train station	37.4	55.1	38.6	29.4	36.7
Vehicle	22.7	27.4	23.6	48.3	54.8
Wind mill	47.1	65.7	45.4	78.7	81.9
mAP / %	51.2	58.6	54.1	57.1	62.6
Time / s	0.021	0.032	0.180	0.024	0.042

表 5 不同模型在 RSOD 数据集的检测结果

Table 5 Test results of different models on RSOD dataset

Parameter	SSD	SSD	Faster RCNN ^[1]	YOLOv3 ^[5]	Ours
	(300×300) ^[15]	(512×512) ^[15]			
Aircraft	68.0	86.1	76.2	89.7	91.6
Oiltank	90.5	90.6	94.3	90.5	92.6
Playground	90.3	90.4	96.0	90.2	91.8
Overpass	77.3	73.3	69.1	63.3	68.7
mAP / %	81.5	85.1	83.9	83.4	86.2
Time / s	0.021	0.034	0.180	0.024	0.042

表 6 不同模型在部分 TGRS-HRRSD 数据集的检测结果

Table 6 Test results of different models on partial TGRS-HRRSD dataset

Parameter	SSD	SSD	Faster RCNN ^[1]	YOLOv3 ^[5]	Ours
	(300×300) ^[15]	(512×512) ^[15]			
Ship	85.0	85.3	80.3	86.1	92.1
Airplane	87.4	89.9	94.1	90.0	92.5
mAP / %	86.2	87.6	87.2	88.0	92.3
Time / s	0.024	0.034	0.183	0.026	0.040

3.3.3 测试结果展示

使用改进的 YOLOv3 和原始的 YOLOv3 分别在 DIOR 数据集、RSOD 数据集和选出的部分 TGRS-HRRSD 数据集上进行测试。随机选取 5 组图片的测试结果进行对比,第 1~5 组图片分

别是对飞机、油罐、车辆、轮船和港口,以及风磨进行检测的结果,对比结果如图 9 所示,图 9(a)均为原始图片,图 9(b)均为原始 YOLOv3 的检测结果,图 9(c)均为使用本文改进的 YOLOv3 的检测结果。

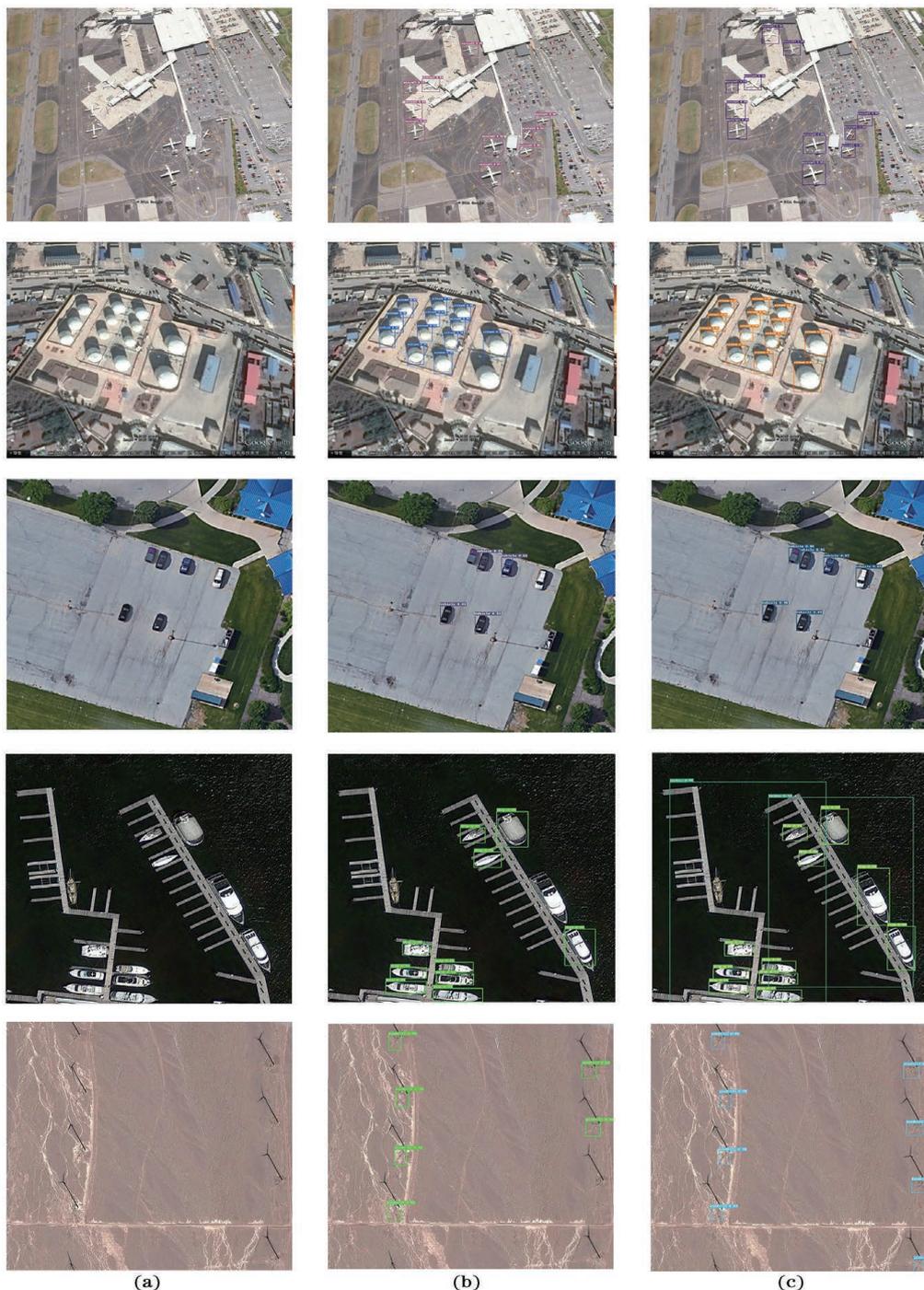


图 9 检测结果对比。(a)原图;(b)YOLOv3 检测结果;(c)本文方法的检测结果

Fig. 9 Comparison of detection results. (a) Original pictures; (b) detection results of YOLOv3; (c) test results of proposed method

从以上 5 组图像中可以看出,在背景复杂的情况下,目标与背景的区分度低,此时原始的

YOLOv3 容易产生漏检,而本文改进后的算法在目标较多的情况下的检测结果更准确,由此可以直观

地发现本文算法优于改进前的算法。

4 结 论

基于 YOLOv3 算法进行一系列改进,提出了一种新的光学遥感图像目标检测算法。通过结合密集连接结构,使卷积层两两互联,增强了原始的特征提取网络,并使用 DIoU loss 使最终的预测框与真实框的重叠度更高,且用 DIoU 对原始 NMS 算法进行了改进。实验结果表明,改进后的算法有效地提高了对遥感图像目标的检测精度。

本文得到的检测精度虽然有了一定提升,但是在检测时间上没有得到有效的缩短,后续希望针对检测速度进行改进,并结合上下文信息对检测精度进行进一步提升,争取同时实现轻量化与高精度。

参 考 文 献

- [1] 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.
- [2] Xu Z J, Ding Y. Ship object detection of remote sensing images based on adaptive rotation region proposal network[J]. Laser & Optoelectronics Progress, 2020, 57(24): 242805.
徐志京, 丁莹. 自适应旋转区域生成网络的遥感图像舰船目标检测[J]. 激光与光电子学进展, 2020, 57 (24): 242805.
- [3] Ou P, Zhang Z, Lu K, et al. Object detection in of remote sensing images based on convolutional neural networks [J]. Laser & Optoelectronics Progress, 2019, 56(5): 051002.
欧攀, 张正, 路奎, 等. 基于卷积神经网络的遥感图像目标检测 [J]. 激光与光电子学进展, 2019, 56 (5): 051002.
- [4] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [5] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-08-08) [2020-11-05]. <https://arxiv.org/abs/1804.02767>.
- [6] Li C Y, Yao J M, Lin Z X, et al. Object detection method based on improved YOLO lightweight network [J]. Laser & Optoelectronics Progress, 2020, 57(14): 141003.
- [7] Li X, Cai K Y. Method research on ship detection in remote sensing image based on Yolo algorithm[C] // 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), August 14-16, 2020, Xi'an, China. New York: IEEE Press, 2020: 104-108.
- [8] Liu R W, Wang T, Zhou Y, et al. A satellite image target detection model based on an improved single-stage target detection network [C] // 2019 Chinese Automation Congress (CAC), November 22-24, 2019, Hangzhou, China. New York: IEEE Press, 2019: 4931-4936.
- [9] 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, June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 658-666.
- [10] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 12993-13000.
- [11] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: a survey and a new benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
- [12] Long Y, Gong Y P, Xiao Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2486-2498.
- [13] Xiao Z F, Liu Q, Tang G F, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images [J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.
- [14] Zhang Y L, Yuan Y, Feng Y C, et al. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5535-5548.
- [15] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M] // Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.