

DOI: 10.13382/j.jemi.B2407552

基于重参数化广义金字塔与扩张残差的 遥感图像旋转框算法*

董宝鑫¹ 王江涛^{1,2}

(1. 淮北师范大学物理与电子信息学院 淮北 235000; 2. 智能计算及应用安徽省重点实验室 淮北 235000)

摘要: 由于遥感图像中目标数量多而密集,且背景信息复杂,导致现有检测算法对于小目标检测精度不够理想,针对该问题,提出了一种基于重参数化广义金字塔与扩张残差的遥感图像小目标旋转框检测算法 DRS-YOLO。首先,为克服主干网络对特征提取不足的缺点,以旋转算法 YOLOv8OBB 为基础,在颈部网络引入扩张式残差模块,以增强遥感目标语义信息。其次,为提高网络对于多尺度目标的检测性能,使底层特征信息流向高层,引入重参数化泛化特征金字塔网络替换颈部网络结构,更高效的融合多尺度特征,易于捕捉高层语义和低层空间细节。最后,为进一步提高网络对于小目标的检测性能,基于 SPPF 提出 SPPFI 对目标感受野进行扩展,提升了对遥感目标的检测精度。在公开的 DIOR 数据集和 HRSC2016 数据集上相较于原 YOLOv8sOBB 基线网络的检测精度分别提升了 1.5% 和 9.8%。实验表明改进后的算法显著增强了对遥感图像小目标的检测性能。

关键词: 深度学习; 遥感图像; 旋转框目标检测; 重参数广义金字塔; DRS-YOLO

中图分类号: TN29 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Remote sensing image rotation box algorithm leveraging reparameterized generalized pyramid and dilated residual

Dong Baoxin¹ Wang Jiangtao^{1,2}

(1. School of Physics and Electronic Information, Huaibei Normal University, Huaibei 235000, China;

2. Anhui Province Key Laboratory of Intelligent Computing and Applications, Huaibei 235000, China)

Abstract: Due to the dense presence of multiple targets and complex background information in remote sensing images, existing detection algorithms often struggle to achieve satisfactory precision in small target detection. To address this issue, we propose a small target rotation box detection algorithm for remote sensing images called DRS-YOLO, based on reparameterized generalized pyramids and dilated residuals. First, to overcome the shortcomings of the backbone network in feature extraction, we enhance semantic information by incorporating a dilated residual module into the neck of the network, building upon the YOLOv8OBB framework. Second, to improve the network's performance in detecting multi-scale targets and facilitate the flow of low-level feature information to high-level features, we replace the neck structure with a reparameterized generalized feature pyramid network for more efficient multi-scale feature fusion, which aids in capturing high-level semantics and low-level spatial details. Finally, to further enhance the network's performance in detecting small targets, we propose the SPPFI module to expand the receptive field, thereby improving detection accuracy for remote sensing targets. Experimental results demonstrate that the improved algorithm achieves an increase in detection precision of 1.5% and 9.8% on the public DIOR and HRSC2016 datasets, respectively, compared to the original YOLOv8sOBB baseline network, indicating a significant enhancement in small target detection performance for remote sensing images.

Keywords: deep learning; remote sensing images; rotated bounding box object detection; reparameterized generalized pyramid; DRS-YOLO

收稿日期: 2024-05-26 Received Date: 2024-05-26

* 基金项目: 国家自然科学基金 (61976101)、安徽省高校自然科学研究重点项目 (2023AH050319)、安徽省高校优秀科技创新团队项目 (2023AH010044) 资助

0 引言

随着当下科技的发展,人们在遥感领域的应用场景愈发广泛,遥感技术在土地利用和覆盖分类、航拍^[1]、农业生产^[2]和气候变化研究^[3]等领域有了显著发展。但于此同时,遥感图像所存在诸多问题。首先,随着遥感分辨率提高和背景区域扩大,图像包含了复杂且无用的背景,这些背景信息显著的干扰了检测精度。其次,同一分辨率下目标尺度变化大,即使是同一目标也会因为拍摄的角度和距离不同而呈现出不同的尺度,另外目标排列密集,使得检测精度下降,识别难度增加等。这些问题依然困扰着众研究者。

近年来,随着遥感领域的发展,众多研究者也提出了自己的算法,Xie 等^[4]提出了一种有效且简单的两阶段定向目标检测框架,即 R-CNN,在第一阶段,提出了一个定向区域提案网络,直接生成高质量的定向提案。第二阶段是定向 R-CNN 头,用于细化定向感兴趣区域并识别它们。然而但由于参数量较大,导致检测速度较低。Li 等^[5]提出了一种轻量级大型选择性核网络 (large selective kernel network, LSKNet) 主干网,LSKNet 可以动态调整其大的空间感受野,以更好地模拟遥感场景中各种物体的测距上下文。但该网络参数量较大,难以在移动平台上使用。Wang 等^[6]在 PP-YOLO 基础上开发 PP-YOLO-R,引入 ProbIoU 损失作为回归损失,以避免边界不连续问题,以适合旋转对象检测;设计了一个解耦的角度预测头,并通过 DFL 损失直接学习角度的一般分布,以获得更准确的角度预测。然而上述算法在特征融合阶段效果较差,在高分辨率图像下的小目标检测效果不理想,因此并不适用于遥感目标。同时,遥感图像目标检测通常部署在无人机等移动端平台,对实时性要求较高。

为了适用实际应用中检测速度及精度的要求,基于 YOLOv8sOBB 提出一种新型的旋转目标检测算法:DRS-YOLO,在密集目标检测、小目标检测等领域有显著提升。首先,为了绘制多尺度信息,融合多尺度特征图,在颈部网络引入膨胀式残差注意力模块 (dilation wise residual module, DWR)。其次,为了提升内部网络特征融合,加快模型的推理速度,提升检测精度,替换颈部结构为重参数化泛化特征金字塔网络 (reparameterized generalized fpn, RepGFPN)。最后,为了使网络聚焦于中小型目标,提出了新型空间池化金字塔 (spatial pyramid pooling fast improve, SPPFI),显著提升检测精度。

1 DRS-YOLO 模型设计思想

YOLOv8 模型是 YOLO^[7] 中的一个最新模型,相较于

之前的 YOLO 算法,其骨干网络和 Neck 部分参考了 YOLOv7 ELAN^[8] 设计思想,将 YOLOv5 的 C3 结构换成了梯度流更丰富的 C2f 结构,并对不同尺度模型调整了不同的通道数,在 Head 部分采用了目前主流的解耦头结构,将分类任务和回归任务解耦,同时将 Anchor-Based 换成了 Anchor-Free,在 Loss 中使用 BEC Loss 作为分类损失,使用 DEL Loss + CIOU Loss 作为回归损失,同时 YOLOv8 采用了 Task-Ailigned Assingner 正样本分配策略,最后在训练的后 10 轮关闭了 Mosaic 数据增强操作,有效的提升了算法精度。YOLOv8OBB 以 YOLOv8 为基础,采用边界范围为的长边定义法,边界范围为 $[-\frac{\pi}{4}, \frac{3\pi}{4}]$,此定义法处于边界范围的目标较少,有利于模型的检测。数据集采用四点标注法。矩形旋转框损失用 ProbIoU,即将旋转框建模为一个二维高斯分布,之后计算预测分布与真实分布之间的距离。

针对遥感目标尺度小,排列密集,背景信息复杂等因素,模型以 YOLOv8OBB 模型为基础,提出 DRS-YOLO。首先,为融合多尺度信息,引入膨胀式残差注意力模块 DWR,构建更全面的特征表示。其次,颈部网络替换为 RepGFPN,加快模型推理速度,提升检测精度。最后,将池化金字塔替 (spatial pyramid pooling-fast, SPPF) 换为 SPPFI,聚焦中小层目标,提升网络精度。网络结构如图 1 所示。

1.1 膨胀式残差注意力模块

针对遥感小目标像素低,难以检测,进一步增强小目标语义信息,在颈部网络引入扩张式残差模块 DWR^[9],DWR 结构分为两步:第一步根据输入特征生成残差特征,即区域残差化。输入特征通过 3×3 的卷积与批量归一化 (BN) 层和 ReLU 结合实现,用于初步特征提取。第二步模块分别采用多速率多速率扩张深度卷积来对不同大小的区域特征进行形态学过滤。即将区域特征图分为 3 个组,然后 3 组的区域特征图分别进行扩张速率为 1、3、5 的扩张深度卷积。通过以上两步,深度扩张卷积的作用由获取复杂语义信息简化为形态过滤,从而对学习过程进行排序,可以更有效地保护多尺度上下文信息。之后将 3 个输出通过 1×1 的卷积层后合并特征,形成最终残差,最后将残差添加到输入特征图中以构建更全面的特征表示,如式 (1) 所示。

$$x = \text{cat}(f^{1 \times 1}, f^{1 \times 1}(D^{3 \times 3, 1} + D^{3 \times 3, 3}, D^{3 \times 3, 5})) \quad (1)$$

其中, $f^{1 \times 1}$ 是卷积核大小为 1×1 , 激活函数为 SiLU 的 CBS 模块, $D^{3 \times 3, 1}$ 是卷积核大小为 3×3 , 激活函数为 SiLU, 膨胀系数为 1 的空洞卷积, $D^{3 \times 3, 3}$ 是卷积核大小为 3×3 , 激活函数为 SiLU, 膨胀系数为 3 的空洞卷积, $D^{3 \times 3, 5}$ 是卷积核大小为 3×3 , 激活函数为 SiLU, 膨胀系数为 5 的空洞卷积, Cat 代表通道级联操作。其网络结构如图 2 所示。

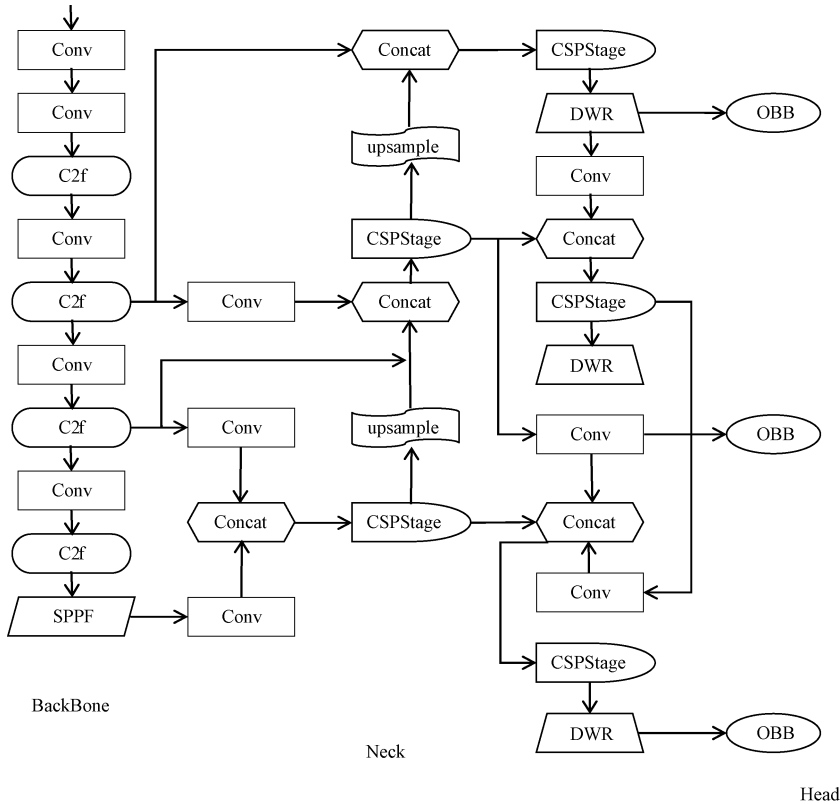


图 1 DRS-YOLO 网络结构
Fig. 1 DRS-YOLO network structure

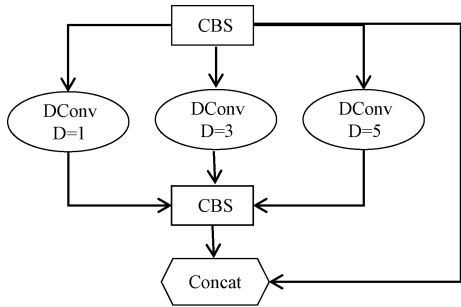


图 2 DWR 网络结构
Fig. 2 DWR network structure

1.2 重参数化泛化特征金字塔网络

在特征融合中,小目标物体用尺度较大的特征图检测,大目标物体用尺度较小特征图进行检测,而将两者结合的特征图对目标的检测更准确。重参数化通过改变模型参数的表示方式,使得在训练过程中可以使用复杂的模型结构或采样操作,而在推理或计算梯度时则转换为更简单、更直接的形式。这种转换有助于保留梯度信息,使得模型参数可以通过反向传播算法进行更新。在变分自编码器(VAE)中,使用重参数化技巧,可以将采样过程分解为确定性部分(均值和方差)和随机部分(噪声)。使得 encoder 计算得到的均值和方差网络就可以通过网

络梯度进行更新。而在模型提出的重参数泛化中,在卷积层和特种融合中采用了重参数泛化,不损失网络性能的前提下进行网络结构的优化。推理时可以被简化为更简单的结构,从而减少了计算量并提高了推理速度,保持较高的检测精度。YOLOv8 采用 FPN + PANet^[10] 结构,FPN 由自上而下的路径来获得语义丰富的特征,并将更准确的位置信息与特征进行结合,PANet 则引入了自底向上的路径,使得底层信息更容易传递到高层顶部。缩短信息传播路径,同时利用低层特征的精准定位信息。然而,在以上方法中,不同尺度的特征信息往往被平等对待、计算成本较高,且只关注特征融合,缺乏了内部连接等问题,Jiang 等^[11] 提出了 GFPN (Generalized FPN),GFPN 注重于网络内部的连接,设计了跳层连接与跨层连接,与此同时,GFPN 很难确保高级低分辨率特征和低级高分辨率特征具有相同的丰富表达能力,并且使用大量的采样操作和 3×3 卷积的跨尺度特征融合,极大地影响了推断速度与参数量。因此为了加快模型的推理速度,提升检测精度,模型提出重参数化泛化特征金字塔网络^[12] (RepGFPN)。首先,针对不同尺度特征图使用不同的通道数,控制高级特征和低级特征的表达能,其次,删除额外上采样操作,减少了模型推断延迟,最后将原始基于卷积的特征融合改进为 CSPStage 连接,并引入重参

数化和 ELAN 连接的思想,以提高模型的准确性而不增加更多的计算量。CSPStage 可以增强 CNN 的学习能力,能够轻量化的同时保持准确性、降低计算瓶颈。RepGFPN 结构如图 3 所示。CSPStage 如式(2)所示。

$$x = \text{Cat}(f^{1 \times 1}, f^{1 \times 1}(f^{3 \times 3} + R(f^{1 \times 1}))_n) \quad (2)$$

其中, $f^{1 \times 1}$ 是卷积核大小为 1×1 , 激活函数为 SiLU 的 CBS 模块, $f^{3 \times 3}$ 是卷积核大小为 3×3 , 激活函数为 SiLU 的 CBS 模块, R 是卷积核大小为 3×3 , 激活函数为 ReLU 的重复卷积神经网络(RepConv), Cat 代表通道级联操作。

CSPStage 采用残差设计, 内部使用 RepConv, 训练时采用双分支结构, 推理时采用单分支结构, 大大减少了推理时间, RepGFPN 结构如图 3 所示。

RepFPN

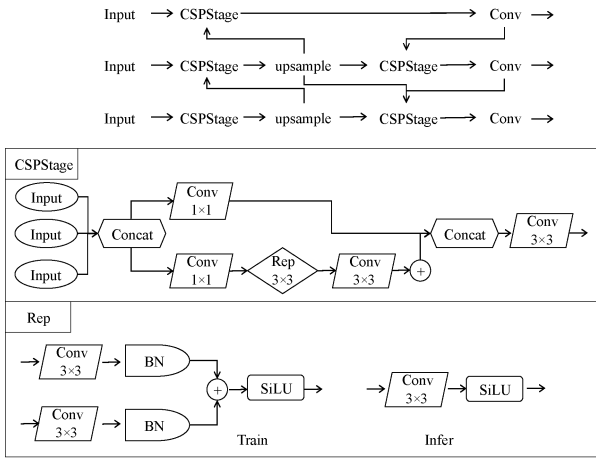


图 3 RepGFPN 网络结构

Fig. 3 RepGFPN network structure

1.3 改进空间金字塔池化

由于小目标空间位置小,对于周边信息感知不强,为了增强网络对于中小目标的关注度,提升特征提取能力,基于 SPPF 的结构提出适用于小目标的空间池化金字塔 SPPFI(spatial pyramid pooling fast improve),将 SPPF 中的池化和尺寸由 5 调整为 3 与 9,输入特征 x 经过 CBS 之后,分别经过 3×3 和 9×9 大小的池化层,得到 x_1 与 x_2 ,之后 x_1 经过池化核为 3×3 大小的池化层得到 x_3 , x_2 经过池化核为 9×9 大小的池化层得到 x_4 ,最后将 x, x_1, x_2, x_3, x_4 进行 Concat 融合,进一步捕获小目标特征信息,提升模型检测性能。公式如式(3)所示。

$$x = \text{Cat}(f^{1 \times 1}, f^{1 \times 1}(M^9), f^{1 \times 1}(M^9(M^9)), f^{1 \times 1}(M^3), f^{1 \times 1}(M^3(M^3))) \quad (3)$$

其中, $f^{1 \times 1}$ 是卷积核大小为 1×1 , 激活函数为 SiLU 的 CBS 模块, $f^{3 \times 3}$ 是卷积核大小为 3×3 , 激活函数为 SiLU 的 CBS 模块, M^9 代表着池化核为 9 的最大池化层, M^3 代表池化核为 3 的最大池化层, Cat 代表通道级联操作。网络

结构如图 4 所示。

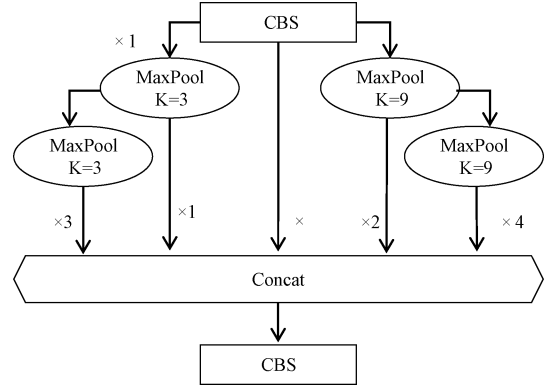


图 4 SPPFI 网络结构

Fig. 4 SPPFI network structure

2 实验结果与评估

2.1 实验设置

本实验所用的环境如表 1 所示。

表 1 本实验所用环境

Table 1 The environment used in this experiment

环境	参数配置
操作系统	Ubuntu 18.04
CPU	Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50 GHz
GPU	NVIDIA GeForce RTX3090
开发环境	Pycharm
CUDA	11.1
编程语言	Python

初始训练参数设置如下:初始学习率设置为 0.01, 学习率动量设置为 0.937, 输入图像大小设置为 640×640 , 批量大小设置为 24, Dropout 为 0.5, 优化器选择 SDG, 训练轮次为 800 epoch。

2.2 数据集

DIOR 数据集^[13]是一个用于光学遥感图像目标检测的大规模基准数据集。它有着 23 463 张遥感图像和 192 472 个目标实例,这些目标实例涵盖了 20 个对象类别,包括飞机、机场、棒球场、篮球场、桥梁、烟囱、水坝、高速公路服务区、高速公路收费站、港口、高尔夫球场、地面田径场、天桥、船舶、体育场、储罐、网球场、火车站、车辆和风车等。这个数据集的图像大小为 800×800 像素,空间分辨率为 $0.5 \sim 30$ m,并被分为训练验证集(11 725 张图像)和测试集(11 738 张图像)。这些图像中的目标实例是用轴向对齐的边界框手动标记的,总共有 192 472 个轴对齐的目标边界框注释。

HRSC2016 数据集^[14]是用于舰船目标检测的高分辨率遥感舰船数据集,其中包含海上舰船和近岸舰船两个场景的图像,共有 1 061 幅有效标注的图像,划分的训练

集、验证集和测试集分别为 436 幅、181 幅和 444 幅。实验中,在保持长宽比的前提下将图像尺寸调整为 1 024×1 024(单位为像素)

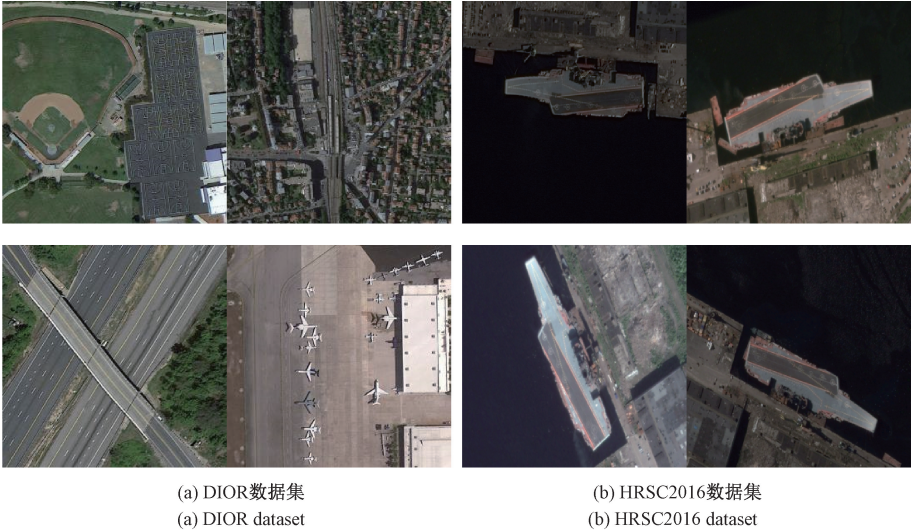


图 5 部分数据集展示
Fig. 5 Partial dataset display

2.3 评价指标

为了验证模型有效性,使用精度 (precision, P)、召回率 (recall, R)、平均精确度 (average precision, AP)、平均精密度 (mean average precision, mAP)、作为模型的评价指标, AP 是一种针对类别精度的评价指标, mAP 为不同类别的 AP 值相加再除以数据集的类别数而得到。而 $mAP_{0.5}$ 、 $mAP_{0.75}$ 分别以 $IoU = 0.5$ 和 0.7 时进行计算, $mAPs$ 为小目标 (像素小于 32×32) 检测的平均准确度。平均精度 AP 和平均精度均值公式如式 (4)~(7) 所示。

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$AP = \int_0^1 Precision(Recall) d(Recall) \tag{6}$$

$$mAP = \frac{AP}{N} = \frac{\int_0^1 p(r) dr}{N} \tag{7}$$

其中, N 代表数据集检测个数,精度 P 是所有检测到的目标中正确预测目标的比例, TP 表示正确的预测目标, FP 表示错误的预测目标。召回率 R 是所有目标中正确检测到目标的比例, FN 表示存在但未被正确检测的目标。

2.4 消融实验

DRS-YOLO 算法分别对 3 个模块进行消融实验,模型所提出的算法相较于 YOLOv8sOBb 而言,在 DIOR 数据集上和 HRSC2016 数据集上检测精度分别提升了

图 5 展示了该数据集中的部分图像,由于该数据集所含图像和类别数众多,且图像之间的相似程度较高,分类精度很难有大幅度的提升。

1.5%和9.8%。在 DIOR 数据集中,在 YOLOv8OBb 颈部网络加入 DWR 注意力机制后精度分别提升 0.6%,用 RepGFPN 替换颈部网络后精度提升 0.6%,将 SPPF 替换 SPPFI 后精度提升 0.3%,结合 3 种改进策略,网络对于小目标的特征提取效果更好,精度有明显提升,如表 2 所示。

表 2 不同模块的影响
Table 2 Effects of different modules

Baseline	DWR	RepGFPN	SPPI	mAP50 (DIOR)	mAP50 (HRSC2016)
✓				71.8%	65.8%
✓	✓			72.4%	75.4%
✓	✓	✓		73.0%	68.7%
✓	✓	✓	✓	73.3%	75.6%

2.5 对比试验

DIOR 数据集中包含了 20 个类别,分别为 Airplane、Airport、Baseball field、Basketball court、Bridge、Chimney、Dam、Expressway service area、Expressway toll station、Golf field、Ground track field、Harbor、Overpass、Ship、Stadium、Storage tank、Tennis court、Train station、Vehicle、Windmill,为了更好的对比改进算法的优势,将算法与当前 R-CNN、Faster R-CNN、SSD、PANet、RICAOD、RICNN 等经典算法在 DIOR 数据集上进行比较,DRS-YOLO 相较于其他经典算法在 DIOR 数据集上均有较大优势,而对于 Airport、Dam 两类别精度有所下降,这是因为这两个类别尺寸较大且具有较大的纵横比,算法将一个目标误认为多个目标。为方便比较,将数据种类编号,如表 3 所示。

对于比 YOLOv80BB, DRS-YOLO 在 Airport、Baseball field、Bridge、Expressway、Service area、Golf field、Stadium、Train station、Windmill 等 8 个类别中精度提升了 2.5%~4%, 其他类别均有不同程度的提升。如表 4 所示。为了

直观展示算法性能,分别在密集目标检测、复杂背景信息下的目标检测、较大纵横比目标检测、小尺度目标检测等4个不同场景与基线算法进行比较,如图6所示,由图可知相较于基线网络,改进算法在密集目标检测下可以检

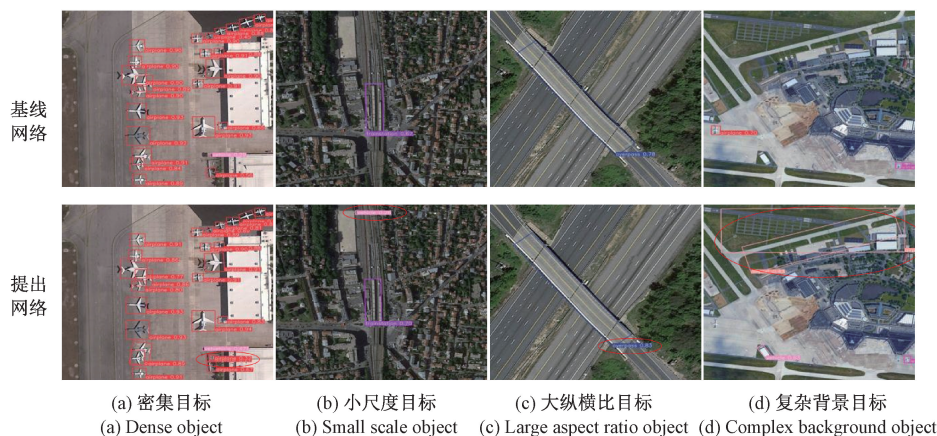


Fig. 6 Comparison between the proposed network and the baseline network

表 3 DIOR 数据集种类编号

Table 3 Category and number of the DIOR dataset

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Airplane	Airport	Baseball field	Basketball court	Bridge	Chimney	Dam	Expressway service ares	Expressway Toll station	Golf field
C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
Ground track field	Harbor	Overpass	Ship	Stadium	Storage tank	Tennis court	Train station	Vehicle	Wind mill

表 4 DIOR 各类别在模型中的精度

Table 4 Accuracy of each category in the DIOR dataset within the model

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	MAP
R-CNN ^[15]	35.6	43.0	53.8	62.3	15.6	53.7	33.7	50.2	33.5	50.1	49.3	39.5	30.9	9.1	60.8	18.0	54.0	36.1	9.1	16.4	37.7
RICNN ^[16]	39.1	61.0	60.1	66.3	25.3	63.3	41.1	51.7	36.6	55.9	58.9	43.5	39.0	9.1	61.1	19.1	63.5	46.1	11.4	31.5	44.2
RICAOD ^[17]	42.2	69.7	62.0	79.0	27.7	68.9	50.1	60.5	49.3	64.4	65.3	42.3	46.8	11.7	53.5	24.5	70.3	53.3	20.4	56.2	50.9
RIFD-CNN ^[18]	56.6	53.2	79.9	69.0	29.0	71.5	63.1	69.0	56.0	68.9	62.4	51.2	51.1	31.7	73.6	41.5	79.5	40.1	28.5	46.9	56.1
Faster R-CNN ^[19]	53.6	49.3	78.8	66.2	28.0	70.9	62.3	69.0	55.2	68.0	56.9	50.2	50.1	27.7	73.0	39.8	75.2	38.6	23.6	45.4	54.1
SSD ^[20]	59.5	72.7	72.4	75.7	29.7	65.8	56.6	63.5	53.1	65.3	68.6	49.4	48.1	59.2	61.0	46.6	76.3	55.1	27.4	65.7	58.6
Faster R-CNN FPN-ResNet-50	54.1	71.4	63.3	81.0	42.6	72.5	57.5	68.7	62.1	73.1	76.5	42.8	56.0	71.8	57.0	53.5	81.2	53.0	43.1	80.9	63.1
FasterR-CNN FPN-ResNet-101	54.0	74.5	63.3	80.7	44.8	72.5	60.0	75.6	62.3	76.0	76.8	46.4	57.2	71.8	68.3	53.8	81.1	59.5	43.1	81.2	65.1
Mask-R-CNN ^[21] FPN-ResNet-50	53.8	72.3	63.2	81.0	38.7	72.6	55.9	71.6	67.0	73.0	75.8	44.2	56.5	71.9	58.6	53.6	81.1	54.0	43.1	81.1	63.5
Mask-R-CNN FPN-ResNet-101	53.9	76.6	63.2	80.9	40.2	72.5	60.4	76.3	62.5	76.0	75.9	46.5	57.4	71.8	68.3	53.7	81.0	62.3	43.0	81.0	65.2
RetinaNet ^[22] ResNet-50	53.7	77.0	69.0	81.3	44.1	72.3	62.5	76.2	66.0	77.7	74.2	50.7	59.6	71.2	69.3	44.8	81.3	54.2	45.1	83.4	65.7
RetinaNet ResNet-101	53.3	77.0	69.3	85.0	44.1	73.2	62.4	78.6	62.8	78.6	76.6	49.9	59.6	71.1	68.4	45.8	81.3	55.2	44.4	85.5	66.1
PANet ^[23] ResNet-50	61.9	70.4	71.0	80.4	38.9	72.5	56.6	68.4	60.0	69.0	74.6	41.6	55.8	71.7	72.9	62.3	81.2	54.6	48.2	86.7	63.8
PANet ResNet-101	60.2	72.0	70.6	80.5	43.6	72.3	61.4	72.1	66.7	72.0	73.4	45.3	56.9	71.7	70.4	62.0	80.9	57.0	49.2	84.5	661
CornerNet ^[24]	58.8	84.2	72.0	80.8	46.4	75.3	64.3	81.6	76.3	79.5	79.5	26.1	60.5	37.6	70.7	45.2	84.0	57.1	43.0	75.9	64.9
YOLOv8OBB	88.1	56.7	80.4	90.9	46.7	81.8	39.6	68.4	63.9	81.0	69.7	60.4	62.1	92.7	71.7	81.6	90.9	68.4	58.6	82.4	71.8
DRS-YOLO	89.5	59.8	82.9	91.9	49.2	83.7	41.1	71.6	65.5	83.5	70.4	60.1	64.0	92.4	74.0	81.4	91.0	70.9	59.1	84.9	73.3

测出较难检测的物体,同时各个类别的物体的置信度较于基线网络高,在复杂背景信息下目标检测、大纵横比目标检测、小尺度目标检测等应用场景中较于基线网络漏检检错概率低。因此 DRS-YOLO 网络在复杂背景下、小目标检测以及排列密集目标检测下均有显著优势。

3 结 论

针对遥感图像中小目标像素低、背景复杂以及目标排列密集等问题,提出了 DRS-YOLO 算法。通过集成 DWR 模块、RepGFPN 模块和 SPPFI 模块以增强对遥感图像的小目标检测性能。实验结果证实 DRS-YOLO 在不增加其大小、参数的情况下,在 DIOR 数据集上实现了 73.3% 的 mAP_{50} ,在复杂背景下小目标检测精度明显改进。而 DIOR 和 HRSC2016 公开数据集上的实验结果表明,模型与所有比较方法相比精度有显著提升。在未来的工作中将继续研究无人机航拍图像的小目标检测算法,在保证小目标实时性检测的前提下进一步提升检测精度。

参考文献

- [1] 陈朋磊,王江涛,张志伟,等. 基于特征聚合与多元协同特征交互的航拍图像小目标检测[J]. 电子测量与仪器学报,2023,37(10):183-192.
CHEN P L, WANG J T, ZHANG ZH W, et al. Small object detection in aerial images based on feature aggregation and multiple cooperative features interaction [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(10): 183-192.
- [2] 史舟,梁宗正,杨媛媛,等. 农业遥感研究现状与展望[J]. 农业机械学报,2015,46(2):247-260.
SHI ZH, LIANG Z ZH, YANG Y Y, et al. Status and prospect of agricultural remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(2): 247-260.
- [3] TALOOR A K, GOSWAMI A, BAHUGUNA I M, et al. Remote sensing and GIS applications in water cryosphere and climate change[J]. Remote Sensing Applications: Society and Environment, 2022, 28: 100866.
- [4] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3520-3529.
- [5] LI Y X, LI X, DAI Y, et al. LSKNet: A foundation lightweight backbone for remote sensing [J]. International Journal of Computer Vision, 2024: 1-22, DOI: 10.48550/arXiv.2403.11735.
- [6] WANG X, WANG G, DANG Q, et al. PP-YOLOE-R: An efficient anchor-free rotated object detector [J]. ArXiv preprint arXiv:2211.02386, 2022.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Computer Vision & Pattern Recognition. IEEE, 2016.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [9] WEI H R, LIU X, XU SH CH, et al. DWRSeg: Rethinking efficient acquisition of multi-scale contextual information for real-time semantic segmentation [J]. ArXiv preprint arXiv:2212.01173, 2022.
- [10] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [11] JIANG Y, TAN Z, WANG J, et al. GiraffeDet: A heavy-neck paradigm for object detection [J]. ArXiv preprint arXiv:2202.04256, 2022.
- [12] XU X, JIANG Y, CHEN W, et al. Damo-yolo: A report on real-time object detection design[J]. ArXiv preprint arXiv:2211.15444, 2022.
- [13] ELSHORBANY Y, MIXSON S, MARCUM L, et al. The relation between atmospheric aerosol concentration and SARS-CoV-2 Variants' infection and mortality rates in the united states: A remote-sensing perspective [J]. Sustainability, 2024, 16(8): 3381.
- [14] LIU Z, YUAN L, WENG L, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines [C]. International conference on pattern recognition applications and methods. SciTePress, 2017, 2: 324-331.
- [15] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [16] YAO X, FENG X, CHENG G, et al. Rotation-invariant latent semantic representation learning for object detection in VHR optical remote sensing images [C]. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 1382-1385.
- [17] LI K, CHENG G, BU SH H, et al. Rotation-insensitive and context-augmented object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(4): 2337-2348.

- [18] CHENG G, ZHOU P, HAN J. Rfd-CNN: Rotation-invariant and fisher discriminative convolutional neural networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2884-2893.
- [19] REN SH 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, 2016, 39(6): 1137-1149.
- [20] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. Springer International Publishing, 2016: 21-37.
- [21] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [22] ROSS T Y, DOLLAR G. Focal loss for dense object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2980-2988.
- [23] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [24] LAW H, DENG J. Cornernet: Detecting objects as

paired keypoints [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 734-750.

作者简介



董宝鑫, 2023 年于山东理工大学获得学士学位, 现为淮北师范大学研究生, 主要研究方向为目标检测、深度学习。

E-mail: 2251169235@qq.com

Dong Baoxin received his B. Sc. degree from Shandong University of Technology in 2023. Now he is a M. Sc. candidate at Huaibei Normal University. His research interests include object detection and deep learning.



王江涛 (通信作者), 2002 年于青岛大学获得学士学位, 2005 年于青岛大学获硕士学位, 2008 年于南京理工大学获得博士学位, 现为淮北师范大学教授, 主要研究方向为图像处理、模式识别、计算机视觉。

E-mail: jiangtaowang@chnu.edu.cn

Wang Jiangtao (Corresponding author) received his B. Sc. degree from Qingdao University in 2002, M. Sc. degree from Qingdao University in 2005, and Ph. D. degree from Nanjing University of Science and Technology in 2008. Now he is a professor at Huaibei Normal University. His research interests include image processing, pattern recognition and computer vision.