

DOI:10.19651/j.cnki.emt.2209862

基于自适应校准和多分支注意力的遥感目标检测^{*}雷帮军^{1,2} 耿红彬^{1,2} 吴正平^{1,2}

(1.三峡大学计算机与信息学院 宜昌 443000; 2.水电工程智能视觉监测湖北省重点实验室 宜昌 443000)

摘要: 面向复杂多变的遥感场景下目标检测易受干扰的问题,提出了结合自校准模块和 D_Triplet Attention 的任意方向目标检测模型 SD-Centernet。该方法在网络结构中引入旋转角度,为检测框提供角度信息。在 Dlanet 特征提取网络中引入 self-Calibrated 模块,通过自适应校准操作融合来自两个不同空间尺度的信息,增大输出特征的感受野。同时为了加强图像局部信息的聚焦,引入 D_Triplet Attention,更好的解决了跨维度交互问题。SD-Centernet 在 HRSC-2016 数据集上的检测精度达到 86.25%,检测速度达到 14.9 帧/秒,有效提高了遥感航拍中多方位目标的检测效果。

关键词: 航拍图像;自适应校准;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:** 520.2040

Remote sensing object detection based on adaptive calibration and multi-branch attention

Lei Bangjun^{1,2} Geng Hongbin^{1,2} Wu Zhengping^{1,2}

(1. School of Computer and Information Science, China Three Gorges University, Yichang 443000, China;

2. Hubei Provincial Key Laboratory of Intelligent Visual Monitoring of Hydropower Engineering, Yichang 443000, China)

Abstract: In order to overcome the problem of vulnerable disturbance of target detection in complex and variable remote sensing situations, this paper suggests SD-Centernet, which is an optional orientation target detection model combining self-Calibrated module and D_Triplet Attention. This new method introduces rotation angle in the network structure, which provides angular information to the detection box. Self-Calibrated module is introduced into the Dlanet feature extraction network to increase the perceptual field of the output features by fusing information from two different spatial scales through an adaptive calibration operation. Meanwhile, D_Triplet Attention is introduced to enhance the focus of image-based local information, which better solves the cross-dimensional interaction problem. 86.25% detection accuracy and 14.9 fps detection speed have been achieved on SD-Centernet in HRSC-2016 Dataset, that effectively improves the multi-directional target detection in remote sensing aerial photography.

Keywords: aerial image; self-calibrated; object detection; attention mechanism

0 引言

随着科学技术的发展,遥感场景下的目标检测已经开始在城市道路、海域感知、农田检测等领域进行使用,但航拍情况因目标密集、尺度变化以及背景干扰等问题,使自动航拍检测在实际场景应用中收效甚微。因此遥感图像的目标精准检测对于地面观测等任务有着重要指导意义。

当前主流目标检测算法主要是 Anchor-based 和 Anchor-free 检测方式。为了解决遥感目标的多方位问题,

研究人员利用 Anchor-based 的检测算法通过目标特征在候选区域引入大量锚框,对特征图筛选出合适的锚框,再完成分类和定位工作。例如 Jianqi Ma 等^[1]提出了旋转区域建议网络(RRPN),利用旋转感兴趣区域池化层生成带有方向角度信息的倾斜建议,并将角度信息用于边界框回归; Minghui Liao 等^[2]利用旋转卷积滤波器对不同网络分支提取的特征进行回归和分类,提取旋转敏感特征和不变特征,提高了面向边框的旋转目标回归性能; Yingying Jiang 等^[3]基于 Faster R-CNN 框架,利用区域建议网络和倾斜非最大抑制法,汇聚不同尺寸轴向的特征,从而得到旋转锚框;

收稿日期:2022-05-06

^{*} 基金项目:国家自然科学基金(U1401252)、湖北省重点实验室开放基金(2018SDSJ07)、2020年产学研合作协同育人项目(202002286038)资助

陈彦彤等^[4]利用端对端的语义分割,将主干网络与条件随机场相结合,解决了检测结果可靠性低的问题;Jian Ding等^[5]通过对 RoI 进行空间变换,有效避免区域特征与目标间的不对称,解决了目标分类置信度和定位精度间不一致问题;张宏群等^[6]利用 DIOU_NMS 算法降低了遥感目标被遮挡后的影响,增强了检测效果;Xingxing Xie 等^[7]提出了面向区域提案网络(面向 RPN)和定向 R-CNN 头部,生成高质量的面向区域,并对感兴趣的区域进行提炼和识别;李明阳等^[8]利用多尺度可变形卷积解决复杂场景中旋转不变性问题,提高了遥感目标检测精度。然而对遥感卫星采集到的图像而言,目标小且密集,易出现目标遮挡、漏检等情况,使 Anchor-based 的目标检测算法受到较大的影响。Anchor-based 的检测算法在锚框筛选过程中,导致正负样本不平衡,出现大量冗余,造成了计算资源的过多浪费。

为了减少锚框的非必要冗余,Anchor-free 的旋转目标检测思路被提出,通过检测关键点或者中心区域来确定目标区域,在目标检测领域取得了较好的检测精度,并逐渐被应用到遥感场景。例如牛戈等^[9]引入注意力机制和特征重建网络,扩大旋转目标关注度,缓解目标特征不对齐问题;Pengming Feng 等^[10]在单阶段无锚框的网络基础上,引入新的几何变换方法回归参数,并使用 Student's-t 分布作为联合分布,避免了旋转锚框带来的复杂计算;王阳明等^[11]利用特征注意力和多尺度池化信息,聚合多个区域信息,进一步提取目标特征;Zhifeng Xiao 等^[12]提出了一种新的宽高比感知方向中心法,利用像素预测的方式检测定向目标,避免了设置太多锚点参数,并对大宽高比目标检测带来了改进;张涛等^[13]在基于关键点的目标检测模型上,引入方向不变模型 OIM 结合 SimAM 注意力机制,增强网络对目标旋转特征的适应能力,实现旋转目标检测。但 Anchor-free 的检测思路应用在航拍领域中,小目标的关键点、中心区域以及轮廓信息随着分辨率的降低而大量丢失,后处理过程效果较差,检测精度仍有不足。

遥感目标在方位上任意排列,背景区域复杂多变,严重影响了目标的检测效果。而现有的遥感多方位检测方法大多忽略了背景区域对目标的干扰。为了解决遥感图像中目标排列紧密、方向多变等问题,本文采用基于中心点的 Anchor-free 检测模型 Centernet 作为基准模型,相较于基于锚框的检测网络,无需进行后处理,提高了检测速度。为降低在无人机航拍检测中背景因素的干扰,提高任意方向目标的检测精度,对 Centernet^[14]提出了以下改进:首先在特征提取网络中引入自适应校准卷积模块,来扩大特征的感受野;并在主干网络结构中引入旋转角度分支;最后在主干网络之后添加 D_Triplet Attention,增加目标特征的像素偏移量,使其更有效地学习目标的角度特征。基于大型舰船遥感数据集 HRSC-2016,本文算法与其他旋转目标检测算法进行了多组遥感目标对比试验,验证了 SD-Centernet 算法的有效性。

1 Centernet 目标检测模型改进设计

1.1 Centernet 目标检测模型的基本原理

Centernet 模型的组成有编、解码网络以及检测网络,结构简单,检测流程如图 1 所示。与 Anchor-based 的检测模型相比,直接在 heatmap 图上执行过滤操作,避免进行复杂的 NMS 后处理,并且不需要设置 Anchor 以及其超参数,显著缩减了网络的计算参数,具有更快的处理速度和检测精度。

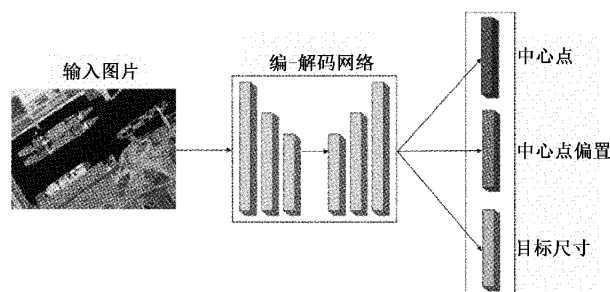


图 1 Centernet 模型

Centernet 利用中心点代替锚框,首先利用主干网络生成的关键点热力图寻找局部最大值,来预测目标中心点的位置,然后根据中心点周围的特征来预测目标的类别以及位置信息。最后通过中心点偏置微调中心点位置坐标,使边框更加贴合目标区域。

1.2 改进的 Centernet 目标检测模型

如引言所述,遥感影像目标具有任意方向排列、密集、尺度变化大等特点,且背景区域干扰性强。针对上述问题,本文基于 Centernet 目标检测模型,结合任意方向目标的特点,提出了一种基于自适应校准模块和多分支注意力机制的任意方向遥感检测方法 SD-Centernet,整体流程如图 2 所示。

SD-Centernet 目标检测算法对 Centernet 网络的改进主要如下:

- 1) 为避免背景区域干扰,本文在 DLANet 特征提取网络的基础上,引入自校准卷积模块^[15],融合来自不同空间的尺度信息,扩大目标特征的感受野;
- 2) 在 Centernet 网络结构中引入旋转角度和角度损失函数,为检测目标提供角度信息,增加图像信息的可读性;
- 3) 在 Triplet Attention 中引入可变形卷积,增加像素偏移量,用于解决目标角度变换问题,并将改进的多分支注意力机制 D_Triplet Attention 附加到指定卷积层,在以可忽略的计算开销下,有效提高检测精度。

1.3 改进特征提取网络

Centernet 采用 DLANet^[16]作为特征提取网络,该网络先利用迭代深度聚合(hierarchical deep aggregation, HDA)合并相同分辨率下、不同卷积层的语义特征,使语义信息融合,然后利用分层深度聚合(iterative deep aggregation,

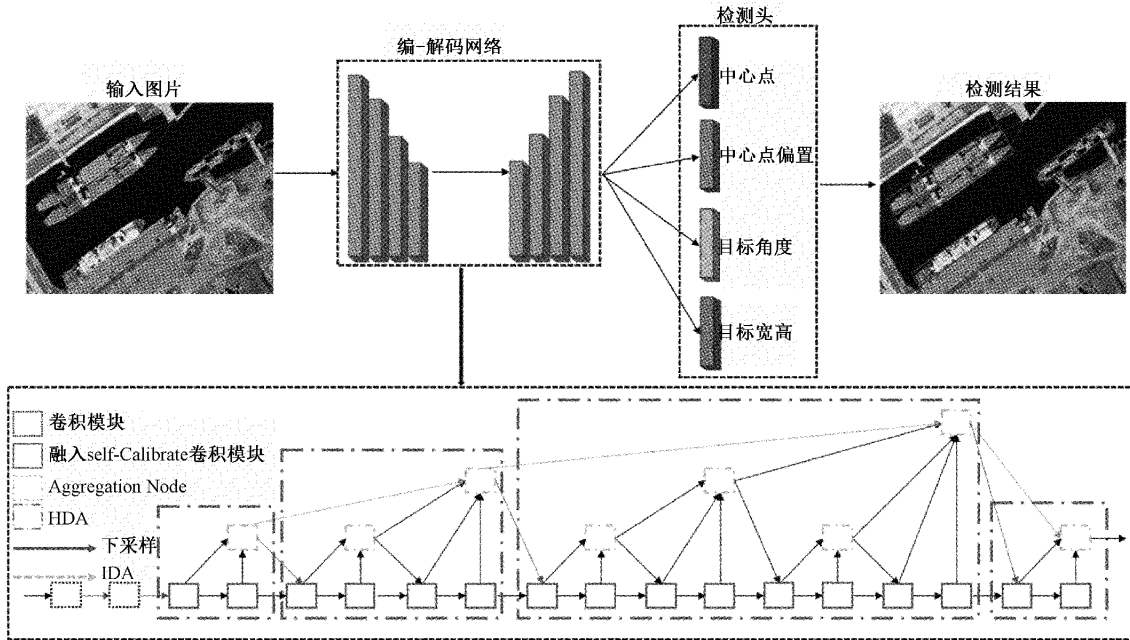


图 2 改进的目标检测模型 SD-CenterNet

IDA)将不同分辨率的特征图进行融合,从而得到丰富的语义信息和空间信息。为了进一步整合特征图的上下文信息,本文通过自适应校准模块(self-calibrated, SC)扩大感受野,关注更多的上下文信息。

图 3 为自适应校准模块的架构图,其中 X 为特征输入图, Y 为特征输出图, F 为不同 kernel_size 的卷积层, K 为对应的 kernel 值。输入特征图 X 的大小为 C×H×W,通过 Self-Calibrated 操作,对特征 X 采用平均池化下采样 4 倍,再利用双线性插值进行上采样操作,最后利用 Sigmoid

激活函数对经过 K2 卷积提取后的特征进行校准,从而得到输出特征 Y。自适应校准模块的公式如下:

$$Y = F_2(X) \cdot \sigma(X + Up(AugPool_r(X) * K_1)) * K_3 \quad (1)$$

与传统卷积不同, self-Calibrated 模块在不引入多余参数的情况下,只关注局部上下文信息,不关注全局。在检测过程中,避免了背景区域的干扰,使目标的检测更精准,能对多尺度信息进行编码,可以更有效地学习不同分辨率下的目标特征。

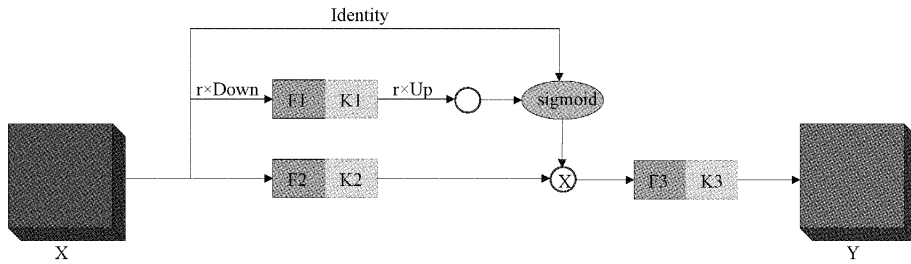


图 3 自适应校准模块

1.4 改进的 Triplet Attention

Diganta Misra 等^[17]在 Triplet Attention 中提出了跨纬度交互的概念,通过旋转操作和残差变换分别捕获通道维度和空间维度之间的交互关系,建立通道和空间维度间的连接。

利用 Z-pool 将通道 C 维度的张量 Tensor 缩减到二维,降低计算量。并将维度上的平均汇集和最大汇集特征连接起来,保留实际张量,公式如下:

$$Z - pool(X) = [MaxPool(X), AugPool(X)] \quad (2)$$

Triplet-Attention 利用三支结构捕获跨纬度交互信

息,来计算注意力权重,最终输出的张量为:

$$y = \frac{1}{3} ((\hat{\chi}_1 \sigma(\psi(\hat{\chi}_1^*))) - \hat{\chi}_2 \sigma(\psi_2(\hat{\chi}_2)) + \chi \sigma(\psi_3(\hat{\chi}_3))) \quad (3)$$

本文在 Triplet Attention 引入 DCN 模块,组成新的注意力机制 D_Triplet Attention。与 Triplet Attention 相比, D_Triplet Attention 增加了像素偏移量,使卷积核在学习过程中扩展到更大的范围,特征学习更明显,如图 4 所示。

1)第 1 个分支:通道 C 与空间 H 维度交互捕获分支, C×H×W 维度的特征经过跨维度相互作用,转变为

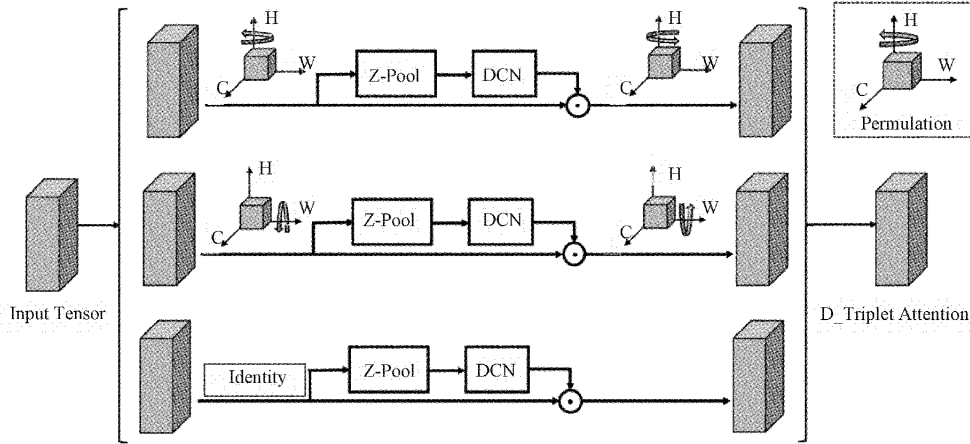


图 4 D_Triplet Attention 结构

$W \times H \times C$ 维度的特征,接着 Z-pool 对输入的 W 维度进行池化操作,最后经过跨维度相互作用还原为 $C \times H \times W$ 维度,如图 5 所示。

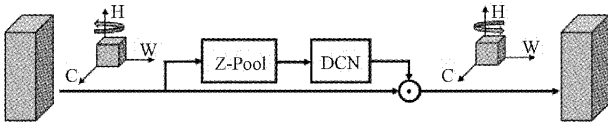


图 5 通道 C 与空间 H 维度交互分支

2)第 2 个分支:通道 C 与空间 W 维度交互捕获分支,经过跨维度相互作用转变为 $H \times C \times W$ 维度特征,对输入的 H 维度进行池化操作,最后还原为 $C \times H \times W$ 维度,如图 6 所示。

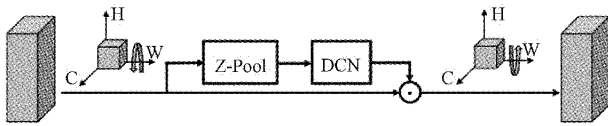


图 6 通道 C 与空间 W 维度交互分支

3)第 3 个分支:空间注意力计算分支,输入 $C \times H \times W$ 维度的特征,经过 Z-pool 和可变形卷积,生成空间注意力权重,如图 7 所示。

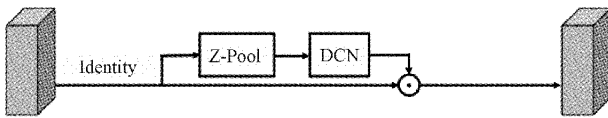


图 7 空间注意力计算分支

4)对 3 个不同跨维度的分支相加求平均,得到最终的注意力权重。

1.5 改进损失函数

针对目标任意方向的角度检测难以确定的问题,本文在原 Centernet 网络的基础上,引入针对目标角度的损失函数。新的组合损失函数由中心点预测损失、中心点偏移损失、目标角度损失以及目标宽高损失四部分组成:

$$L_{loss} = L_k + \lambda_{offset} L_{offset} + \lambda_{angle} L_{angle} + \lambda_{size} L_{size} \quad (4)$$

其中, $\lambda_{offset} = 0.5, \lambda_{angle} = 0.1, \lambda_{size} = 0.1$ 。

中心点预测采用 Focal Loss 损失函数,具体表现形式为:

$$L_k = -\frac{1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}), Y_{xyc} = 1 \\ (1 - \hat{Y}_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \log(1 - \hat{Y}_{xyc}), \text{其他} \end{cases} \quad (5)$$

其中, α 和 β 是超参数, N 代表的是图像关键点个数,

\hat{Y}_{xyc} 是中心点预测值。

中心点偏移损失函数的具体形式为:

$$L_{offset} = \frac{1}{N} \begin{cases} 0.5 \times \left(\hat{O}_p - \left(\frac{p}{R} - \hat{p} \right) \right)^2, |\hat{O}_p - \left(\frac{p}{R} - \hat{p} \right)| < 1 \\ |\hat{O}_p - \left(\frac{p}{R} - \hat{p} \right)| - 0.5, \text{其他} \end{cases} \quad (6)$$

其中, p 代表目标框中心点, R 代表下采样倍数。

目标角度损失采用 SmoothL1 Loss 损失函数,具体形式如下:

$$L_{angle} = \frac{1}{N} \begin{cases} 0.5 \times (\hat{A}_p - a_k)^2, |\hat{A}_p - a_k| < 1 \\ |\hat{A}_p - a_k| - 0.5, \text{其他} \end{cases} \quad (7)$$

其中, \hat{A}_p 为目标的角度预测值, a_k 为目标的实际角度值。

目标宽高损失函数,具体形式如下:

$$L_{size} = \frac{1}{N} \sum_{k=1}^N \begin{cases} 0.5 \times (\hat{S}_p - s_k)^2, |\hat{S}_p - s_k| < 1 \\ |\hat{S}_p - s_k| - 0.5, \text{其他} \end{cases} \quad (8)$$

其中, \hat{S}_p 为目标的宽高预测值, s_k 为目标的真实宽高值。

2 实验内容与结果分析

本文的实验硬件环境为 NVIDIA GeForce GTX1080,

Intel(R)Core(TM)i7-8700K CPU@3.70 GHz,16 GB 运行内存;软件环境为 Python3.7.0,PyTorch1.2.0。实验数据集选择的是航拍舰船目标检测数据集 HRSC2016,评估指标选用 VOC2007 平均精度(average precision,AP)和检测速度(FPS)。

2.1 数据集介绍

HRSC2016 舰船数据集目标密集且变化较大,包含 Google Earth 上多个港口的图像,场景主要为近岸舰船和海域舰船。图像大小在 300×300~1 500×900 像素之间,图像分辨率处于 0.4~2 m 之间。此外为了提高模型训练的准确性和鲁棒性,本文使用镜像翻转、添加噪声等增强方法做数据扩充。

2.2 实验设计

为进一步比较改进的网络结构对算法检测精度的影响,本文以添加角度分支的 Centernet 作为基线模型,在 HRSC2016 数据集上设计了 6 组消融实验,消融实验效果

如图 8 所示。消融实验 1 是添加角度分支的 Centernet 基线模型;消融实验 2 在基线模型中添加自适应校准模块,提高了模型对上下文信息的关注度,增大特征的感受野;消融实验 3 在基线模型中添加 Triplet Attention,通过跨维度交互作用,提高目标的聚焦,但对于多方位的舰船目标,受背景干扰严重,精度反而降低;消融实验 4 在基线模型中添加 D_Triplet Attention,引入像素偏移量,更适用于带有角度的舰船目标,增强了局部信息聚焦;消融实验 5 在基线模型中添加自适应校准模块和 D_Triplet Attention,即 SD-Centernet。模型通过关注上下文信息,扩大输出特征的感受野,降低背景干扰。在检测过程中利用 DCN 引入像素偏移量,增强了图像局部信息的聚焦,提高目标在任意方向上的检测精度。在消融实验中,均采用 DLANet 作为骨干网络,输入图像统一为 512×512,添加自适应校准模块和 D_Triplet Attention 的模型效果最佳。实验结果如表 1 所示。

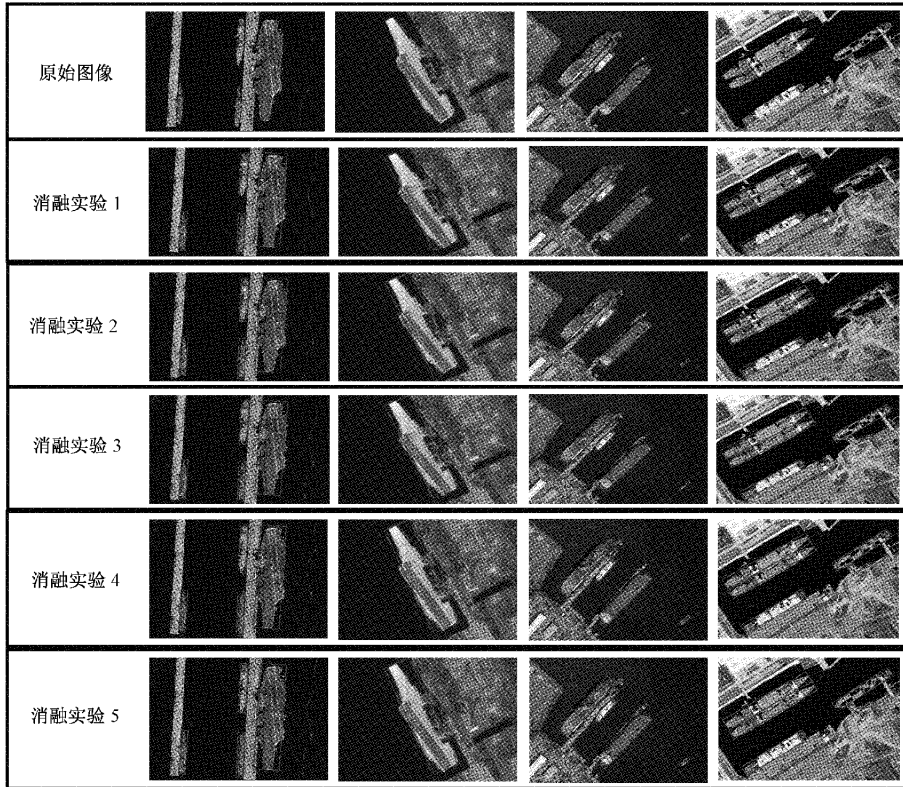


图 8 消融实验检测效果

表 1 消融实验结果

| 实验 | SC | Tri | D-Tri | mAP/% |
|----|----|-----|-------|-------|
| 1 | | | | 82.37 |
| 2 | ✓ | | | 85.72 |
| 3 | | ✓ | | 81.78 |
| 4 | | | ✓ | 83.82 |
| 5 | ✓ | | ✓ | 86.25 |

注:SC 是指 self Calibrated 模块, Tri 是指 Triplet Attention, D-Tri 是指 D_Triplet Attention。

为验证改进后的 SD-Centernet 模型有效性,将改进算法 SD-Centernet 与 R²CNN、RR-CNN、RRPN、RNet-H、RRD 等多种旋转目标检测算法在 HRSC-2016 数据集上做对比实验,实验结果如表 2 所示。

从表 2 可以看出,本文方法在推理速度和检测精度具有较高的性能。改进模型 SD-Centernet 的 mAP 值超过 R²CNN、RR-CNN、RRPN 等算法,比 R²CNN 高 13.18%,比 RRPN 高 7.17%,比 RRD 高 1.95%。SD-Centernet 的

表 2 在 HRSC2016 数据集上不同旋转目标检测算法的性能对比

| | 旋转目标检测模型 | | | | | |
|-----------|----------|------------------------|---------|---------|---------|---------|
| | R2CNN | RR-CNN ^[18] | RRPN | TOSO | RRD | Ours |
| Backbone | R101 | VGG16 | R101 | R101 | VGG16 | S_DLA |
| Imagesize | 800×800 | — | 800×800 | 800×800 | 384×384 | 512×512 |
| 检测速度(FPS) | 5 | — | 1.5 | 17 | — | 14.9 |
| mAP/% | 73.07 | 75.70 | 79.08 | 79.29 | 84.30 | 86.25 |

注:R101 指的是 Resnet101 网络。

检测速度达到 14.9 fps,高于 R2CNN、RRPN 等算法,仅次于 TOSO 的 17 fps。

3 结 论

针对任意方向的目标分布,本文提出了基于自校准模块和多分支注意力的任意方向目标检测方法,结合自适应校准模块和融入可变性卷积的多分支注意力机制,通过增大特征的感受野和加强图像局部信息的聚焦,进一步解决了目标的角度问题,实现了任意方向目标检测。

在后续的工作中,将进一步优化检测算法,针对自然场景中航拍目标遮挡、密集、尺度变化以及背景干扰等情况,提高算法的鲁棒性。

参考文献

- [1] MA J Q, SHAO W Y. IEEE Transactions on multimedia[C]. IEEE,2018;3111-3122.
- [2] LIAO M H, ZHU Z. Otation-sensitive regression for oriented scene text detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE,2018;5909-5918.
- [3] JIANG Y Y, ZHU X Y. R2CNN: Rotational region CNN for orientation robust scene text detection[J]. ArXiv Preprint,2017, ArXiv:1706.09579.
- [4] 陈彦彤,李雨阳,陈伟楠,等.基于深度语义分割的遥感图像海面舰船检测研究[J]. 仪器仪表学报, 2020, 41(1):233-240.
- [5] DING J, XUE N. Learning RoI transformer for oriented object detection in aerial images [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE,2019;2849-2858.
- [6] 张宏群,班勇苗,郭玲玲,等.基于 YOLOv5 的遥感图像舰船的检测方法[J]. 电子测量技术, 2021, 38(8): 87-92.
- [7] XIE X X, CHENG G. Oriented R-CNN for object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3520-3529.
- [8] 李明阳,胡显,雷宏.基于可变形卷积神经网络的遥感图像飞机目标检测[J]. 国外电子测量技术, 2020, 7: 212-126.
- [9] 牛戈,陈小前,季明江,等.基于注意力机制特征重建

网络的舰船目标检测[J]. 上海航天(中英文), 2021, 4(38):128-136.

- [10] FENG P M, LIN Y T. TOSO: Student' s-T distribution aided one-stage orientation target detection in remote sensing images[C]. International Conference on Acoustics, Speech, and Signal Processing, IEEE, 2020;4057-4061.
- [11] 王明阳,王江涛,刘琛.基于关键点的遥感图像旋转目标检测[J]. 电子测量与仪器学报, 2021, 6(35): 102-108.
- [12] XIAO Z F, QIAN L J. Axis learning for orientated objects detection in aerial images[J]. Remote Sensing, 2020,12(6):1-20.
- [13] 张涛,杨小冈.基于关键点的遥感图像舰船目标检测[J]. 系统工程与电子技术, 2022,44(08):2437-2447.
- [14] ZHOU X Y, WANG D Q. Objects as Points [J]. ArXiv Preprint,2019, ArXiv:1904.07850.
- [15] LIU J J, HOU Q B. Improving convolutional networks with self-calibrated convolutions[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE,2020;10096-10105.
- [16] YU F, WANG D Q. Deep layer aggregation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE,2018;2403-2412.
- [17] MIRSA D, NALAMADA T. Rotate to attend: Convolutional triplet attention module[C]. 2021 IEEE Winter Conference on Applications of Computer Vision, IEEE,2021:3139-3148.
- [18] LIU Z K, HU J G. Rotated region based CNN for ship detection[C]. 2017 IEEE International Conference on Image Processing, IEEE,2017:900-904.

作者简介

雷帮军,博士,博士生导师,欧盟 2020 计划特聘顾问,欧盟高级人才计划会审专家,IEEE 高级会员,湖北省楚天学者特聘教授,湖北省百人计划人才,研究方向为计算机视觉、图像处理、人工智能、模式识别。

E-mail: Bangjun. Lei@ctgu. edu. cn

耿红彬,硕士研究生,研究方向为计算机视觉、图像处理。

E-mail: 2697587306@qq. com

吴正平(通信作者),博士,硕士生导师,讲师,研究方向为计算机视觉、图像处理、人工智能、模式识别。

E-mail: 34673637@qq. com