

基于改进级联算法的遥感图像目标检测

王友伟^{1,2}, 郭颖^{1,2*}, 邵香迎^{1,2}¹南京信息工程大学江苏省大气环境与装备技术协同创新中心, 江苏 南京, 210044;²南京信息工程大学自动化学院, 江苏 南京 210044

摘要 针对遥感图像中小目标规模大、分布不均匀、尺度比例变化大和背景复杂等问题,提出一种改进的级联算法 SA-Cascade。该算法使用循环特征金字塔使产生的特征表示逐步增强,提高小目标的检测率。使用基于可学习锚的建议区域生成网络,对遥感目标进行精确定位,并且引入特征自适应模块和特征融合模块,以提高模型对复杂背景图像的检测性能。在级联的基础上引入双分支检测头以提高模型对小目标的检测性能。在 TGRS-HRRSD-Dataset 和 VisDrone-DET 数据集上对不同算法进行对比实验,实验结果表明:改进后的级联算法可以更精确地对遥感图像目标进行检测和定位,相比改进前的级联算法在两个数据集上的精度分别提高 2.94% 和 9.71%。

关键词 遥感; 目标检测; 深度学习; 特征融合; 级联算法

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/AOS202242.2428004

Target Detection in Remote Sensing Images Based on Improved Cascade Algorithm

Wang Youwei^{1,2}, Guo Ying^{1,2*}, Shao Xiangying^{1,2}¹Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;²School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China

Abstract Given the large scale, uneven distribution, and large scale changes of small targets and the complex background in remote sensing images, an improved cascade algorithm SA-Cascade is proposed. This algorithm uses a recurrent feature pyramid network to strengthen the feature representation generated step by step, thereby improving the detection rate of small targets. The region proposal generation network based on the learnable anchor is utilized to locate the remote sensing target accurately. The feature adaptation module and feature fusion module are introduced to improve the performance in detecting images with complex backgrounds. On the basis of the cascade algorithm, the two-branch detection head is adopted to improve the performance of the model for detecting small targets. A comparative experiment of various algorithms is performed on TGRS-HRRSD-Dataset and VisDrone-DET dataset. The experimental results show that the improved cascade algorithm can detect and locate remote sensing image targets more accurately. Compared with the original cascade algorithm, the improved one increases the accuracy on the two datasets by 2.94% and 9.71%, respectively.

Key words remote sensing; target detection; deep learning; feature fusion; cascade algorithm

1 引言

遥感图像目标检测是遥感图像处理领域的首要任务,其目的是针对给定的遥感图像数据集,自动找出感兴趣区域(ROI),并给出目标的具体类别。遥感图像目标通常有机场、飞机、港口、船舶等,在民生建设、城市规

划、国防安全等方面都有重要的用途^[1-2]。因此,遥感图像目标检测具有重要的研究意义和广阔的应用前景。

传统的可见光遥感图像目标检测方法分为基于模板匹配的检测算法、基于知识表示的检测算法、基于图像目标分析的检测算法和基于传统机器学习的检测算法。这些算法都需要基于人工设定的规则或是模板,

收稿日期: 2022-03-16; 修回日期: 2022-04-27; 录用日期: 2022-05-28

基金项目: 国家自然科学基金(61971229)

通信作者: *yguo@nuist.edu.cn

没有统一的标准且鲁棒性较差。近年来,深度学习在目标检测领域取得了很大进展,遥感图像目标检测算法可以借鉴其成功经验。但遥感图像相较于自然图像具有小目标规模大,目标尺寸、方向、姿态不一的特点,加上季节光照变化和环境影响导致的背景复杂等特点,将处理自然图像的目标检测模型直接引入遥感数据集效果不佳。

针对遥感图像各类目标相似性较高的问题,文献[3]提出一种跨尺度的特征融合检测框架 CSFF;针对遥感图像中背景模糊的问题,文献[4]提出一种多模型决策融合目标检测框架,通过融合上下文信息提高目标检测精度;针对遥感图像中目标尺度不一的问题,文献[5]提出了基于上下文特征金字塔的多尺度目标检测框架,通过增强场景与目标之间的联系,提高多尺度目标检测的性能;针对遥感图像中目标角度不一的问题,文献[6]提出一种学习旋转不变的模型 RICNN,通过引入并学习新的旋转不变层来提高检测性能;为了解决不同尺寸目标的检测问题,文献[7]提出一种基于注意力机制的多尺度目标检测算法,通过引入注意力机制对不同通道的特征图进行权重的重新分配;针对目标探测算法对混合像元难以区分以及阈值不易选择的问题,文献[8]提出对抗生长(AG)算法;针对现有基于平视图像的目标检测算法在航拍图像上检测精度不高的问题,文献[9]提出一种具有动态感受野的单阶段目标检测算法,通过在 RetinaNet 结构中添加 bottom-up 短连接通路和全局上下文上采样模块,增强了检测层特征的结构性和语义性;文献[10]提出的 Cascade R-CNN 级联算法延续了基于候选区域的二阶段思想,采用级联的检测头结构可以逐级提升检测性能,对小目标有较好的检测效果。

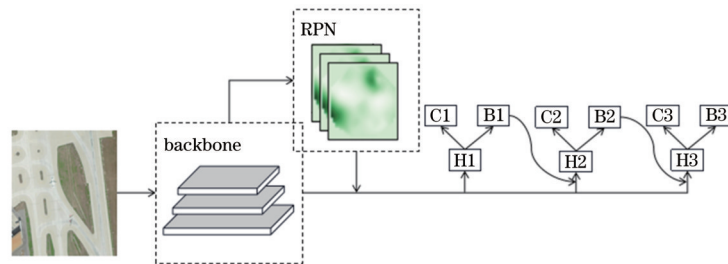


图1 Cascade R-CNN 框架

Fig. 1 Framework of Cascade R-CNN

3 改进级联算法

本文通过改进 Cascade R-CNN 模型,针对遥感图像存在的小目标规模大,背景占据较大比重,目标方向、尺寸、姿态不一等问题,提出了一种基于级联的遥感图像目标检测算法 SA-Cascade。首先将图片送入主干网络 ResNet^[12]进行特征提取,针对遥感图像中小目标规模大、特征不明显的问题,本文使用循环特征金字塔(R-FPN)对输入图片进行循环处理,以得到更全面的特征细节。将处理好的特征送入基于可学习锚的建议区域生成网络(LA-RPN),进行粗

本文针对传统算法对遥感图像的检测存在小目标规模大,背景占据较大比重,目标方向、尺寸、姿态不一等问题,提出一种改进的级联算法 SA-Cascade。该算法在级联模型的基础上引入了循环特征金字塔模块和双分支检测头,可以有效提升小目标的检测率。同时通过使用可学习锚生成建议区域,减小大量背景对检测的干扰,该算法在目标尺度变化较大、方向任意的情况下仍然可以有较高的精确度,同时可以有效减小计算成本。

2 级联网络 Cascade R-CNN 介绍

Cai 等^[10]在 2018 年提出了 Cascade R-CNN 级联算法,如图 1 所示。Cascade R-CNN 延续了 Faster R-CNN^[11]基于候选区域的二阶段思想,与 Faster R-CNN 的不同之处在于:Cascade R-CNN 使用级联的检测头,每个 stage 采用递增的交并比(IoU)阈值。

通过使用级联检测头逐步增加 IoU 阈值,解决了以下问题:1)在训练时如果使用较低的 IoU 阈值则算法会学习到大量的背景框、产生大量的噪声干扰预测;2)如果使用较高的 IoU 阈值则检测器的表现同样会变得很差。产生以上问题的主要原因有两点:一是因为随着阈值的增加,正样本数量会随之减少,导致模型存在过拟合的风险;二是因为推理过程中输入 proposals 的 IoU 和模型设置的 IoU 不匹配。实验表明,级联网络在每个阶段均有足够数量的 proposals,并不会出现严重的过拟合现象。前一个检测头输出一个良好的数据分布作为输入,接着训练下一个高质量检测头,这样可以有效缓解假阳性的问题。在推理阶段可以使用同样的网络结构提高 IoU 阈值并且不会出现之前提到的 IoU 不匹配问题。

分类,得到较为精确的锚框和分为前景或背景的二分类结果。最后将 LA-RPN 的输出和特征图送入级联双分支检测头(db-head),通过全连接层(FC)和卷积层得到最终分类和位置回归结果。本文还引入比例自适应模块,加强特征表示,减少遥感图像中信息的丢失;通过引入特征自适应模块和特征融合模块,使得到的锚框更接近真实框,减小遥感图像背景比重较大造成的干扰。最后将上述的模块整合到一个统一的框架中,得到基于级联网络的遥感图像目标检测算法 SA-Cascade。SA-Cascade 算法结构如图 2 所示。

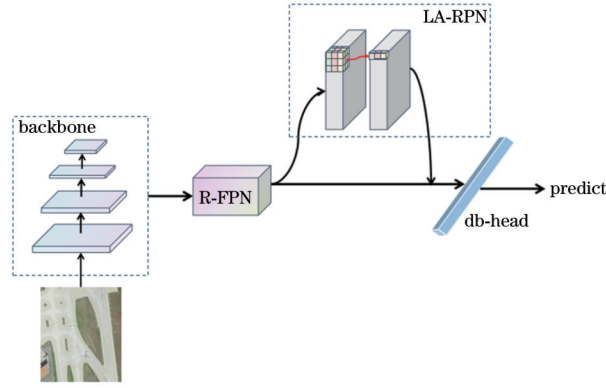


图 2 SA-Cascade算法框架
Fig. 2 Framework of SA-Cascade

3.1 循环特征金字塔模块

本文采用 ResNet 作为特征提取网络, ResNet 在 ImageNet 上进行预训练。由于遥感图像的目标通常较小或者特征不明显, 本文使用循环特征金字塔 (R-FPN)^[13], 通过递归特征金字塔 (FPN) 得到更强的特征表示, 使其能够应对遥感图像这类复杂的目标。具体来说是将 FPN 模块进行堆叠, 第一次 FPN 输出的结果经过 ASPP (atrous spatial pyramid pooling)^[14] 连接和主干网络后作为第二次 FPN 的输入。ASPP 连接是一个多分支结构, 混合了卷积核大小不同的空洞卷积和普通卷积, 在扩大感受野的同时丰富了特征, 如图 4 所示。对两次 FPN 的输出使用 sigmoid 函数分别赋上权重, 然后通过相加得到最后输出的特征。普通的

FPN 在最深层特征图生成时直接使用 1×1 卷积进行降维, 导致最深层特征通道数减少, 造成信息丢失^[15]。为了弥补通道数减少带来的语义损失, 本文引入了比例自适应模块, 在 ResNet 网络最后一层后生成不同比例的特征图, 通过比例自适应模块进行融合, 这样可以充分考虑到不同比例的目标, 融合结果包含丰富的上下文信息, 自适应模块更加适用于遥感图像。循环特征金字塔结构如图 3 所示。图 3 中 C2~C5 代表主干网络中 4 个不同阶段的输出; P2~P5 代表特征提取网络不同阶段的输出。由于本文使用的是循环特征金字塔结构, 左边的 C2~C5 和 P2~P5 代表第一次特征提取阶段的输出, 右边的 C2~C5 和 P2~P5 代表第二次特征提取阶段的输出。

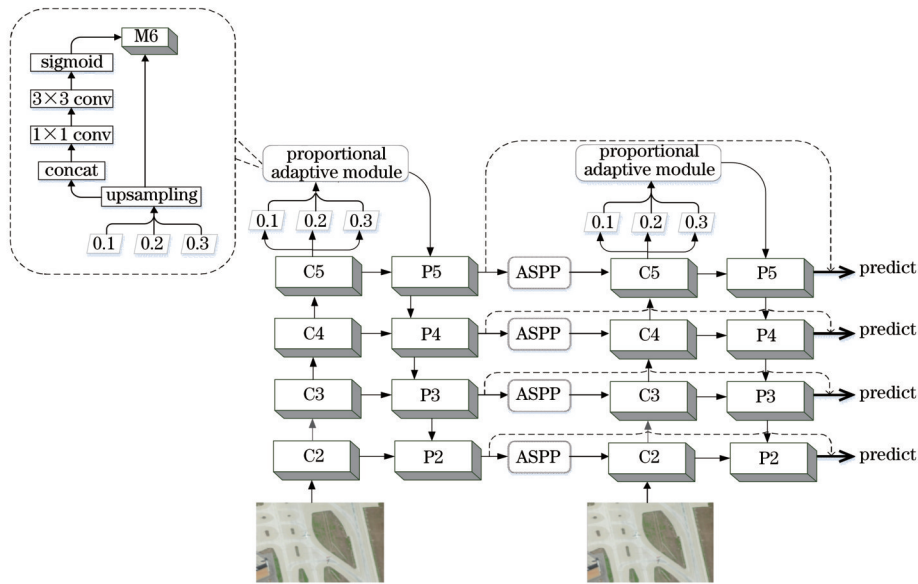


图 3 循环特征金字塔结构图
Fig. 3 Framework of R-FPN

3.2 基于可学习锚的建议区域生成模块

本文提出的算法是二阶段的目标检测算法, 将生成的特征图送入建议区域生成网络 (RPN) 中。而遥感图像的特点是目标很小, 背景占据较大部分, 使用普通的 RPN 网络时, 特征图上会先生成大量的锚点, 每

个锚点上生成设定的 9 个尺寸不同的锚框, 这样不仅需要大量的计算成本和时间成本, 而且遥感图像的特性导致大部分锚点所在的位置属于背景, 效率较低。本文使用一种基于可学习锚^[16]的建议区域生成网络 (LA-RPN), 该网络首先识别包含目标的子区域, 然后

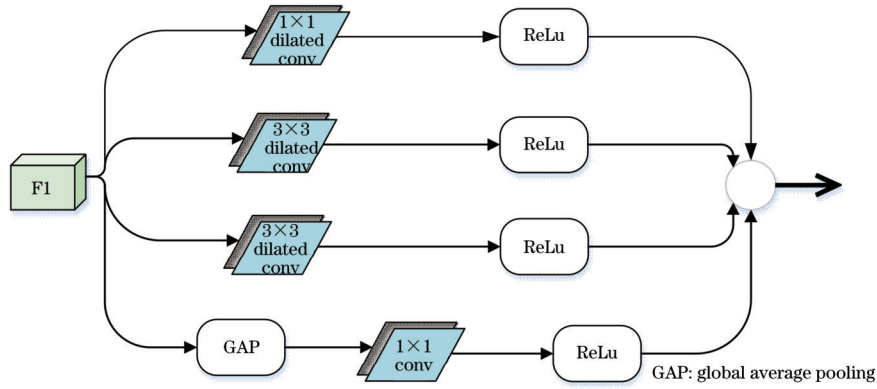


图 4 ASPP连接方式图

Fig. 4 ASPP connection mode

在这些子区域上生成不同尺寸的锚框。具体来说, LA-RPN是一个二分支结构,输入为R-FPN输出的特征图,一个分支预测目标所在的位置,采用 Focal loss^[17]作为损失函数 L_{loc} ,另一个分支预测目标的形状,采用 Bounded IoU loss^[18]作为损失函数 L_{shape} 。两个损失函数的表达式分别为

$$L_{loc}(p_i) = -\alpha_i(1-p_i)^\gamma \lg p_i, \quad (1)$$

$$L_{shape} = 2L_1[1 - \text{IoU}_B(i, b_i)], \quad (2)$$

式中: p_i 为预测为真实标签的概率; α_i, γ 为权重因子; L_1 为 Smooth L1 loss; $i \in \{x, y, w, h\}$; IoU_B 为 i 的 IoU 函数上限; b_i 为真实框的范围。通过将两个分支结果融合得到粗分类结果和建议区域,对遥感图像目标进

行精确定位。普通的RPN示意图和本文的基于可学习锚的建议区域生成网络示意图对比如图5所示。

普通的RPN只生成预测的锚框,并没有对特征图进行进一步优化,而遥感图像中存在目标尺度不一的问题,本文引入特征自适应模块,通过普通卷积和可变形卷积^[19]的结合使用让原本的特征变得可自适应,进而可以让不同位置的特征捕获到不同大小的内容。遥感图像不同目标的方向差异较大,本文还引入特征融合模块,赋予特征位置和形状信息,使特征图适应每个位置的锚框形状,回归结果更接近真实值。基于可学习锚的建议区域生成网络结构如图6所示, σ 表示经过 sigmoid 函数得到的结果。

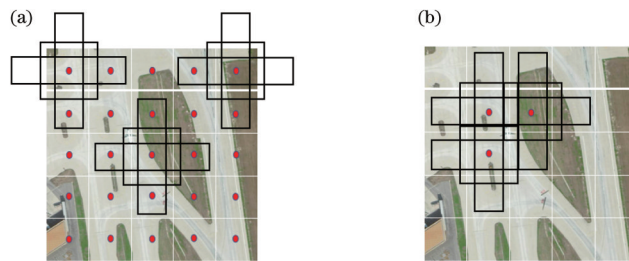


图 5 锚生成方式对比。(a) 普通 RPN 生成锚方式; (b) LA-RPN 生成锚方式

Fig. 5 Comparison of anchor generation methods. (a) Common RPN anchor generation method; (b) LA-RPN anchor generation method

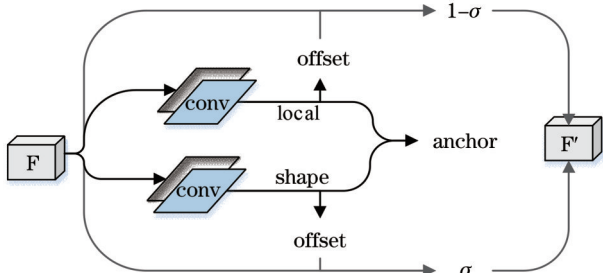


图 6 基于可学习锚的建议区域生成网络框架

Fig. 6 Framework of LA-RPN

3.3 基于级联的双分支检测头模块

本文提出的算法是基于级联检测头的,分为三个

阶段,每个阶段的输入由前一阶段的输出和特征图共同构成。级联模型在不同阶段设置了不同的IoU阈值以逐步提升检测性能。级联检测头使用的是和普通检测头一样的全连接层结构,如图7(a)所示,但 Wu等^[20]的研究结果显示全连接层检测头更适合用作分类任务,而卷积层检测头更适合定位,所以本文为级联的每一阶段都引入双分支结构,如图7(b)所示,一个分支为全连接层结构,另一个分支为卷积层结构,最后预测的分类结果融合了全连接分支和卷积分支的预测结果,位置预测结果取卷积分支的预测结果。本文为卷积分支引入了特征自适应模块(FAM),使预测框能进一步逼近真实框。使用交叉熵损失函数^[21]作为分类损失 L_{cls} ,使用 Smooth L1 损失函数^[22]作为回归损失 L_{reg} ,

则二者分别可表示为

$$L_{cls} = \sum_{i=1}^N \lg(1 + e^{-y_i s}), \quad (3)$$

$$L_{reg} = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}, \quad (4)$$

式中: y 为样本标签; s 为上一层的输出。算法整体的损失函数为

$$L = \lambda_1 L_{loc} + \lambda_2 L_{shape} + L_{cls} + L_{reg}, \quad (5)$$

式中: λ_1, λ_2 为权重因子。基于级联的双分支检测头 (db-head) 结构如图 8 所示。

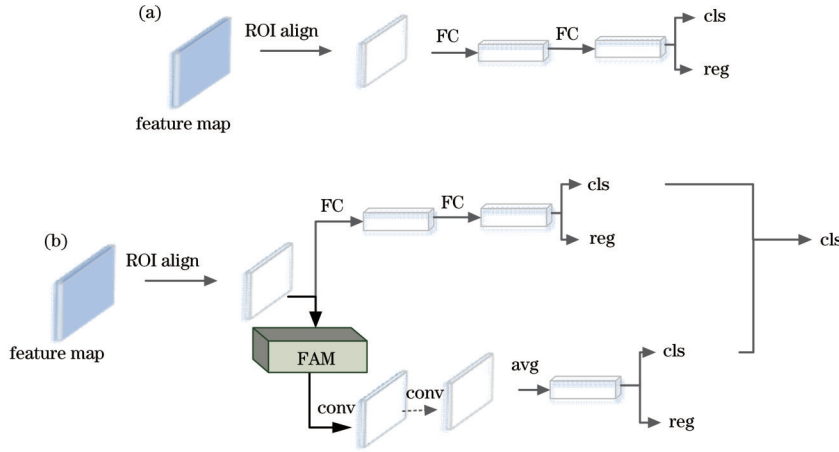


图 7 检测头结构对比图。(a)全连接层结构检测头;(b)双分支检测头

Fig. 7 Structure comparison diagram of detection head. (a) Full connection layer structure detection head; (b) db-head detection head

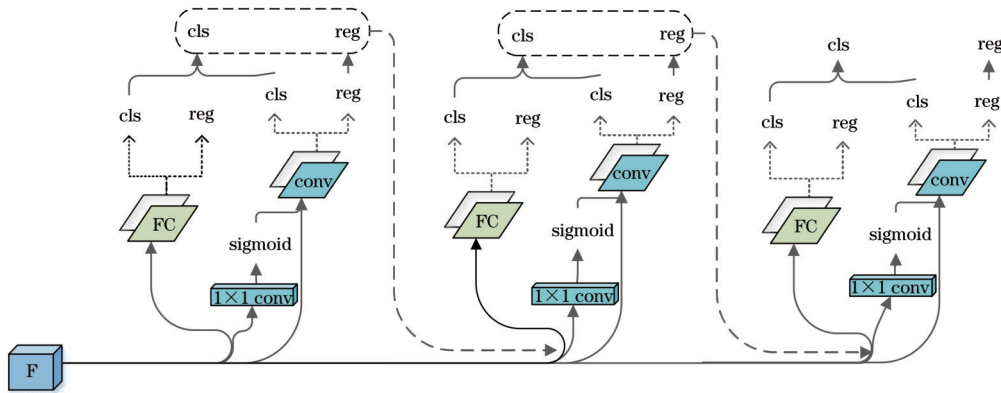


图 8 基于级联的双分支检测头模块结构图

Fig. 8 Framework of db-head

4 实验结果与分析

4.1 数据集

本文所用的数据集为中国科学院西安光学精密机械研究所光学影像分析与学习中心制作的用于研究高分辨率遥感图像目标检测的数据集 TGRS-HRRSD-Dataset^[23]。考虑到不同类别目标数量不平衡不利于目标检测模型的训练,所以该数据集保持不同类别样本量的平衡。TGRS-HRRSD-Dataset 包含 13 类目标: Airplane (APL)、Baseball diamond (BD)、Basketball court (BC)、Bridge (BR)、Crossroad (CR)、Ground track field (GTF)、Harbor (HA)、Parking lot (PL)、Ship (SH)、Storage tank (ST)、T junction (TJ)、Tennis court (TC)、Vehicle (VE), 包括从谷歌地图获取的空间分辨率为 0.15~1.2 m 的 21761 幅彩色图像和从百度地图获取的空间分辨率为 0.6~1.2 m 的 4961 幅彩色图像。此

外本文还使用由天津大学机器学习与数据挖掘实验室的 AISKYEYE 团队收集的数据集 VisDrone-DET dataset^[24]验证新算法在遥感图像数据集上的通用性。

4.2 实验配置与评价指标

本实验基于 Mmdetection V2 框架^[25], 实验环境为 Python 3.8, Pytorch 1.7.0, torchvision 0.7.0, batch size 为 2, 学习率初始值设置为 0.001, 最大训练 epoch 数为 12, 在迭代到第 9、11 个 epoch 后学习率分别下降到 10^{-4} 和 10^{-5} 。硬件设备为 Intel® Core™ i9-10900X CPU, NVIDIA RTX3080Ti 显卡。考虑到遥感图像较大, 将输入图像的尺寸调整为 800 pixel × 800 pixel。以数据集各类目标的精度 AP 和全类目标平均精度 mAP 作为实验的衡量指标。

4.3 消融实验

为了验证 SA-Cascade 算法的有效性, 设计了一组消融实验, 将 Cascade R-CNN 作为基准方法, 用本文

提出的 3 个模块与级联框架组合进行对比实验,消融实验在 TGRS-HRRSD-Dataset 上完成,实验结果如表 1 所示,√表示模型含有该模块。

由表 1 可以看出本文所提出的各模块在遥感图像数据集目标检测任务上的有效性。由于遥感图像中背景占据较大部分,所以针对这一问题提出的 LA-RPN 对精度提升效果最好。db-head 由于在卷积分支堆叠了残差块,增加了计算量,对速度影响较大,而三个模块组合使用时可以缓解速度问题,所以本文后续实验

采用三个模块组合完成。由表 1 可知本文提出的算法在级联框架上效果更好,这是因为级联框架逐级提高设定的 IoU 阈值,逐步提升检测精度,对小目标有较好的检测效果。

为了验证本文提出算法在不同主干网络上的通用性,设计了一组对比实验,将 Cascade R-CNN 与本文提出的 SA-Cascade 算法在不同的主干网络上进行对比,实验数据集为 TGRS-HRRSD-Dataset,实验结果如表 2 所示。

表 1 消融实验结果对比

Table 1 Comparison of ablation experimental results

Cascade framework	R-FPN	LA-RPN	db-head	mAP / %	Increased percentage	Recall / %	Detection rate / (frame·s ⁻¹)
				89.84		93.10	47.3
√				90.17	0.33	92.60	18.4
√	√			91.12	1.28	92.64	16.5
√		√		91.78	1.94	94.55	13.0
√			√	90.67	0.83	94.23	9.0
√	√	√		92.73	2.89	94.92	15.4
√	√		√	91.61	1.77	94.45	6.7
√		√	√	91.64	1.80	94.49	16.6
	√	√	√	92.10	2.26	94.76	13.0
√	√	√	√	93.11	3.27	94.82	10.4

表 2 不同主干网络的结果对比

Table 2 Comparison of results of different backbones

Model	Network	mAP / %	AP for IoU of 0.5 / %	Recall / %
Cascade R-CNN ^[10]	ResNet50	84.90		
Cascade R-CNN (ours)	ResNet50	90.17	90.20	92.17
	ResNet101	90.66	90.70	92.60
	ResNet152	91.31	91.30	92.97
SA-Cascade	ResNet50	93.11	93.10	94.82
	ResNet101	94.62	94.60	96.75
	ResNet152	94.98	94.96	96.92

对比实验中主干网络使用 ResNet50、ResNet101、ResNet152,由表 2 可以看出单纯地增加网络深度对于遥感图像目标检测效果的提升有限,这是因为遥感图像中小目标具有较大部分,而小目标特征多存在于浅层的语义信息中,所以一味地增加网络深度对处理遥感图像目标检测任务的帮助有限。本文提出的算法相较于基准网络 Cascade R-CNN 在主干网络为 ResNet50 的情况下精度提高 2.94%,在主干网络为 ResNet101 的情况下精度提高 3.96%,在主干网络为 ResNet152 的情况下精度提高 3.67%。

4.4 算法对比实验

为了验证 SA-Cascade 算法对遥感图像中各种尺寸目标的有效性,本文依据 COCO 数据集定义的大、中、小目标划分,将尺寸小于 32² 的目标定义为小目标,

将尺寸在 32²~96² 的目标定义为中目标,将尺寸大于 96² 的目标定义为大目标,并在数据集上与基准 Cascade R-CNN 进行对比。实验结果如表 3 所示。表 3 中 mAP_s 表示模型检测目标尺寸小于 32×32 的平均检测精度,mAP_m 表示模型检测目标尺寸从 32×32 到 96×96 的平均检测精度;mAP_l 表示模型检测目标尺寸大于 96×96 的平均检测精度。

表 3 不同尺寸目标的精度对比

Table 3 Accuracy comparison of different size targets

unit: %

Model	mAP	mAP_s	mAP_m	mAP_l
Cascade R-CNN	90.17	44.40	83.50	84.40
SA-Cascade	93.11	58.70	86.30	87.30

由表 3 可知,SA-Cascade 算法对于基准小目标精度提升 14.3%,对于中目标精度提升 2.8%,对于大目标精度提升 2.9%,这说明 SA-Cascade 算法对遥感图像中各类尺寸目标检测的有效性,尤其在小目标检测

中优势明显。

实验将本文提出的算法与 YOLO V3、Faster R-CNN、Cascade R-CNN 进行对比,数据集为 TGRS-HRRSD-Dataset,实验结果如表 4 所示。

表 4 TGRS-HRRSD-Dataset 上本文算法与其他算法的精度对比

Table 4 Comparison of accuracy of this algorithm and other algorithms on TGRS-HRRSD- Dataset

unit: %

Model	YOLO V3 ^[26]	Faster R-CNN ^{▲[11]}	Faster R-CNN ^[11]	MS HEMN ^[27]	Cascade R-CNN ^[10]	Cascade R-CNN (ours)	SA-Cascade	SA-Cascade (ms)
APL	90.7	90.8	90.9	90.9	90.9	98.1	99.2	99.2
BD	76.4	86.9	87.4	88.5	88.7	93.3	95.4	95.7
BC	66.3	47.9	65.0	74.0	68.7	68.9	72.4	80.4
BR	75.2	85.5	84.3	88.3	88.5	94.1	97.5	97.5
CR	75.0	88.6	88.7	88.7	88.8	92.9	96.8	97.2
GTF	89.4	90.6	90.7	90.8	90.8	98.0	99.0	99.1
HA	45.9	89.4	89.1	88.8	90.0	97.8	97.5	98.1
PL	42.3	69.3	64.1	68.1	63.0	69.4	76.7	81.8
SH	73.6	88.5	87.5	89.6	89.8	94.2	95.5	95.7
ST	77.0	88.7	89.8	89.9	90.0	95.2	96.5	97.9
TJ	59.4	75.1	70.8	78.7	74.3	80.3	89.7	91.9
TC	86.2	80.7	89.9	90.4	90.3	92.2	93.5	95.4
VE	73.3	84.0	89.4	90.2	90.3	97.8	98.2	98.1
mAP	71.59	81.50	83.40	85.90	84.90	90.17	93.11	94.46

表 4 中 ms 表示多尺度训练,默认主干网络为 ResNet50,*表示主干网络为 Darknet53,▲表示主干网络为 VGG16。基于回归的 YOLO V3 的精度比基于候选框的二阶段方法更低,即基于候选框的方法能够降低小目标的漏检率。Faster R-CNN 中主干网络使用 ResNet50 的效果要好于使用 VGG16,这是因为 ResNet50 的结构深度更深,而且使用了残差结构,可以避免出现深度越深则效果越差的退化问题。本文提出的 SA-Cascade 的精度较 Cascade R-CNN(ours)提升 2.94 个百分点,使用多尺度训练的精度较基准网络提升 4.29 个百分点。图 9 为本文算法 SA-Cascade 在 TGRS-HRRSD-Dataset 上的部分检测结果以及和其他方法检测结果的对比,可以看出本文所提出的算法相较于 Faster R-CNN 和 Cascade R-CNN 算法检测精度较高且漏检率较低,所有小目标都可以被检测到。结合表 4 和图 9 可以看出本文提出的算法相较于其他算法可以较好地检测出遥感图像中的目标,相比图 9(a)和图 9(b),图 9(c)中所有的小目标均被检出且没有误检框。

为了验证本文提出算法在遥感数据集上的通用性,本文将模型在 VisDrone-DET 数据集上进行实验,实验结果如表 5 所示。图 5 中 AP50 指 IoU 的值取 0.5 时 AP 的值,AP75 指 IoU 的值取 0.75 时 AP 的值,AR100 指在每张图上设置 100 个检测框得到的最大召回率。

由表 5 可以看出本文所提出算法的精度较基准

表 5 VisDrone-DET 数据集上本文算法与其他算法的结果对比

Table 5 Comparison between results of this algorithm and other algorithms on VisDrone-DET dataset

Model	mAP	AP50	AP75	AR100
CornerNet ^[28]	17.41	34.12	15.78	24.37
Light-RCNN ^[29]	16.53	32.78	15.13	23.09
FPN ^[30]	16.51	32.20	14.91	20.72
DetNet59 ^[31]	15.26	29.23	14.34	20.87
RefineDet ^[32]	14.90	28.76	14.08	18.13
RetinaNet ^[17]	11.81	21.37	11.62	5.31
Cascade R-CNN ^[10]	16.09	31.91	15.01	21.37
SA-Cascade	25.80	41.70	27.20	34.90

Cascade R-CNN 有了较大提升,mAP 提升了 9.71 个百分点,AP50 提升了 9.79 个百分点,AP75 提升了 12.19 个百分点,AR100 提升了 13.53 个百分点。这是因为本文使用的基于可学习锚的建议区域生成模块相比普通 RPN 有着更好的分类能力和更高的边框回归精度,同时使用的循环特征金字塔结构可以有效降低小目标的漏检率。结合表 4 和表 5 可以看出本文所提算法针对不同的遥感图像数据集都有较大的精度提升。

5 结 论

为了解决遥感图像中小目标规模大、尺度变化大、

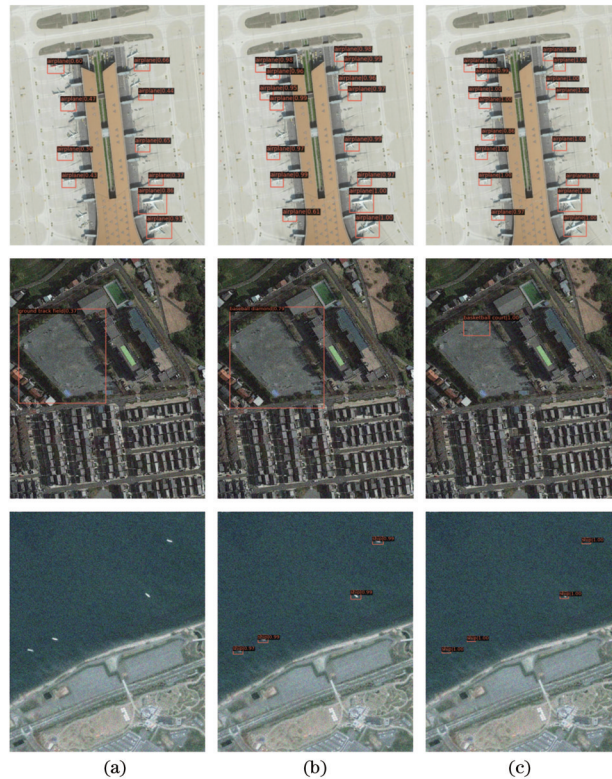


图 9 SA-Cascade 与其他方法的检测效果对比。(a) Faster R-CNN+ResNet50;(b) Cascade R-CNN+ResNet50;(c) SA-Cascade+ResNet50

Fig. 9 Comparison of SA-Cascade and other algorithms on detection effect. (a) Faster R-CNN+ResNet50;(b) Cascade R-CNN+ResNet50;(c) SA-Cascade+ResNet50

背景占据较大比重、背景混乱等问题,对传统 Cascade R-CNN 算法进行改进,提出了一种结合可学习锚和双分支检测头的级联算法 SA-Cascade。在传统级联算法的基础上引入循环特征金字塔模块和双分支检测头,用来增强算法对小目标的检测能力,使用基于可学习锚的建议区域生成网络在对遥感图像的目标进行精准定位的同时可以降低计算成本。实验结果表明该算法对于遥感图像的目标检测精度有较好的提升。由于使用了循环特征金字塔和双分支检测头模块,所提模型的结构较深,增加了模型的计算量,相较于原级联算法速度有所下降。下一步对模型进行轻量化改进,对网络结构进行剪枝优化,使模型在精度损失不大的基础上能够有较快的运行速度。

参 考 文 献

- [1] Yang F, Fan H, Chu P, et al. Clustered object detection in aerial images[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8310-8319.
- [2] Cui Z Y, Li Q, Cao Z J, et al. Dense attention pyramid networks for multi-scale ship detection in SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8983-8997.
- [3] Cheng G, Si Y J, Hong H L, et al. Cross-scale feature fusion for object detection in optical remote sensing

images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(3): 431-435.

- [4] Ma W P, Guo Q Q, Wu Y, et al. A novel multi-model decision fusion network for object detection in remote sensing images[J]. Remote Sensing, 2019, 11(7): 737.
- [5] Chen C Y, Gong W G, Chen Y L, et al. Object detection in remote sensing images based on a scene-contextual feature pyramid network[J]. Remote Sensing, 2019, 11(3): 339.
- [6] Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
- [7] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13): 1315002.
- [8] 邓世杰, 王海晏, 徐安, 等. 基于对抗生长的目标检测方法[J]. 光学学报, 2022, 42(2): 0210002.
- [9] 谢学立, 李传祥, 杨小冈, 等. 基于动态感受野的航拍图像目标检测算法[J]. 光学学报, 2020, 40(4): 0415001.
- [9] Xie X L, Li C X, Yang X G, et al. Dynamic receptive field-based object detection in aerial imaging[J]. Acta

- Optica Sinica, 2020, 40(4): 0415001.
- [10] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6154-6162.
- [11] 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.
- [12] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [13] Qiao S Y, Chen L C, Yuille A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 10208-10219.
- [14] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [15] Guo C X, Fan B, Zhang Q, et al. AugFPN: improving multi-scale feature learning for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 12592-12601.
- [16] Wang J Q, Chen K, Yang S, et al. Region proposal by guided anchoring[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2960-2969.
- [17] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [18] Tychsen-Smith L, Petersson L. Improving object localization with fitness NMS and bounded IoU loss[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6877-6885.
- [19] Xu Z Z, Xu X, Wang L, et al. Deformable ConvNet with aspect ratio constrained NMS for object detection in remote sensing imagery[J]. Remote Sensing, 2017, 9(12): 1312.
- [20] Wu Y, Chen Y P, Yuan L, et al. Rethinking classification and localization for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10183-10192.
- [21] 任进军, 王宁. 神经网络中损失函数的研究[J]. 甘肃高师学报, 2018, 23(2): 61-63.
- Ren J J, Wang N. Research on cost function in artificial neural network[J]. Journal of Gansu Normal Colleges, 2018, 23(2): 61-63.
- [22] Wu S K, Yang J R, Wang X G, et al. IoU-balanced loss functions for single-stage object detection[J]. Pattern Recognition Letters, 2022, 156: 96-103.
- [23] 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.
- [24] Zhu P F, Wen L Y, Du D W, et al. VisDrone-VDT2018: the vision meets drone video detection and tracking challenge results[M]//Leal-Taixé L, Roth S. Computer vision-ECCV 2018 workshops. Lecture notes in computer science. Cham: Springer, 2019, 11133: 496-518.
- [25] Chen K, Wang J Q, Pang J M, et al. MMDetection: open MMLab detection toolbox and benchmark[EB/OL]. (2019-06-17) [2021-05-08]. <https://arxiv.org/abs/1906.07155>.
- [26] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [201-08-09]. <https://arxiv.org/abs/1804.02767>.
- [27] Zhang L, Wang Y H, Huo Y. Object detection in high-resolution remote sensing images based on a hard-example-mining network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(10): 8768-8780.
- [28] Law H, Deng J. CornerNet: detecting objects as paired keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.
- [29] Li Z M, Peng C, Yu G, et al. Light-head R-CNN: in defense of two-stage object detector[EB/OL]. (2017-11-20)[2021-08-05]. <https://arxiv.org/abs/1711.07264>.
- [30] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [31] Li Z M, Peng C, Yu G, et al. DetNet: a backbone network for object detection[EB/OL]. (2018-04-17) [2021-07-08]. <https://arxiv.org/abs/1804.06215>.
- [32] Zhang S F, Wen L Y, Bian X, et al. Single-shot refinement neural network for object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4203-4212.