

# 基于多尺度特征增强的遥感图像目标检测方法

宋智超<sup>1</sup>, 李 筠<sup>1</sup>, 杨海马<sup>1\*</sup>, 刘 瑾<sup>2</sup>, 金 焱<sup>1</sup>

<sup>1</sup>上海理工大学光电信息与计算机工程学院, 上海

<sup>2</sup>上海工程技术大学电子电气工程学院, 上海

收稿日期: 2023年3月9日; 录用日期: 2023年4月13日; 发布日期: 2023年4月23日

## 摘 要

针对遥感图像目标检测中存在的背景复杂、目标像素数少以及目标尺度变化大等问题, 本文提出一种基于多尺度特征增强的遥感图像目标检测方法。首先, 使用具有高分辨率输出的HRNet网络替换ResNet作为主干网络, 强化对遥感目标位置信息的提取; 其次, 在HRNet中引入注意力机制, 抑制复杂背景噪声的干扰; 最后, 设计多尺度特征增强金字塔网络, 进一步增强网络的多尺度特征信息表达。实验结果表明, 相较于原始Cascade R-CNN目标检测方法, 所提方法的目标检测均值平均精度提高了5.32%; 在与经典目标检测方法的对比实验中, 所提方法也表现出较好的检测性能。

## 关键词

目标检测, 多尺度特征增强金字塔, 注意力机制, 遥感图像, HRNet

# Remote Sensing Image Object Detection Algorithm Based on Multi-Scale Feature Enhancement

Zhichao Song<sup>1</sup>, Jun Li<sup>1</sup>, Haima Yang<sup>1\*</sup>, Jin Liu<sup>2</sup>, Yan Jin<sup>1</sup>

<sup>1</sup>School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

<sup>2</sup>School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai

Received: Mar. 9<sup>th</sup>, 2023; accepted: Apr. 13<sup>th</sup>, 2023; published: Apr. 23<sup>rd</sup>, 2023

## Abstract

To address the problems of complex image background, small number of object pixels and large

\*通讯作者。

文章引用: 宋智超, 李筠, 杨海马, 刘瑾, 金焱. 基于多尺度特征增强的遥感图像目标检测方法[J]. 软件工程与应用, 2023, 12(2): 309-317. DOI: 10.12677/sea.2023.122031

variation of object scale in remote sensing image object detection, we propose a remote sensing image object detection method based on multi-scale feature enhancement. First, the HRNet network with high-resolution output is used to replace ResNet to strengthen the backbone network to obtain the location of remote sensing objects; second, the attention mechanism is introduced into HRNet to suppress the interference of complex background noise; finally, the multi-scale feature-enhanced pyramid network is designed to further enhance the multi-scale information representation of the pyramid network. The results of the experiment show that compared with the Cascade R-CNN object detection method, the mean accuracy of the proposed method is improved by 5.32%, and the proposed method also shows better detection performance in comparison with the classical object detection method.

## Keywords

Object Detection, Multi-Scale Feature-Enhanced Pyramid, Attention Mechanism, Remote Sensing Image, HRNet

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

遥感图像目标检测是指使用特定的方法从遥感图像中搜索并标记感兴趣的目标，其在城市规划、智慧农业、路网监测、军事探测等方面均有广泛应用[1]。例如，基于遥感图像数据研究推测城市体系空间结构的演化趋势，设计、制定高效的土地使用战略；基于多源遥感数据构建综合干旱监测模型，为地区防旱抗旱实践提供科学依据；基于高分辨率遥感图像信息，搭建军事阵地动态监测平台等。随着遥感卫星技术的发展，遥感图像目标检测具有更多的应用前景和科研价值。

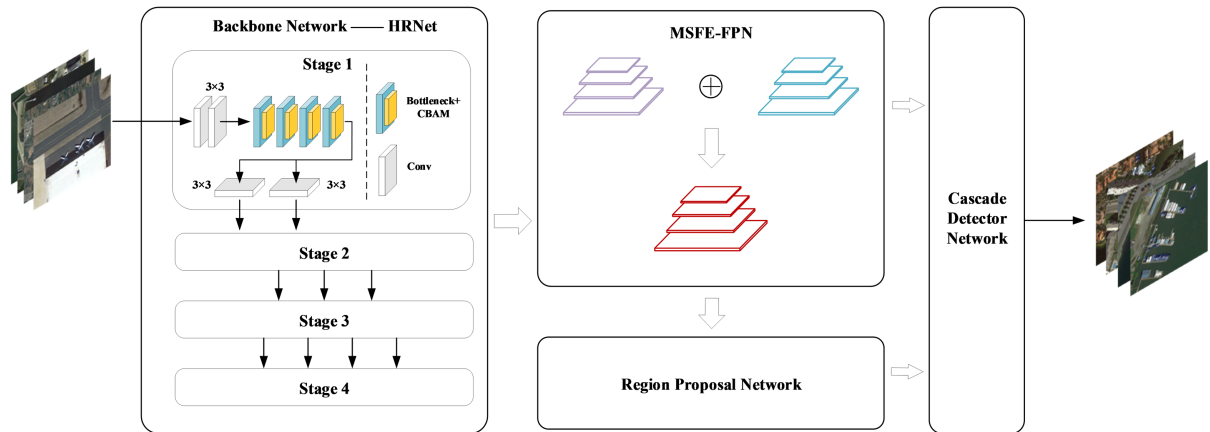
由于成像平台和成像方式不同，遥感图像相较于自然图像具有背景复杂、目标像素数少以及尺度变化大等特点，这导致直接将自然图像的目标检测方法应用于遥感图像目标检测时效果不佳。对此，许多研究学者提出、改进了适用于遥感目标检测的深度学习网络。Wang 等[2]利用 K-means 对大小不一的飞机目标进行聚类分析获得代表性的飞机尺寸，以此提高对多尺度飞机目标的检测精度。Ding 等[3]使用密集卷积网络和多尺度表示等改进、增强基础 VGG16 网络。姚群力等[4]提出了一种多尺度卷积神经网络遥感目标检测框架，通过改进特征金字塔结构增强多尺度目标的检测能力。刘楠等[5]改进 Faster R-CNN 网络，设计通道数可调模块和感受野可调模块以分别提升检测遥感目标检测的速度和精度。

为进一步提高复杂背景干扰下多尺度遥感目标的检测精度，本文提出一种基于多尺度特征增强的遥感图像目标检测方法，使网络关注遥感图片中丰富的位置信息和多尺度特征信息，提高遥感图像目标检测精度。首先使用引入注意力机制的 HRNet 网络替换原 ResNet 网络，引导主干网络关注有用的特征区域，抑制复杂遥感图像背景的干扰；然后设计多尺度特征增强金字塔网络，强化多尺度特征信息，增强对多尺度目标的检测能力。本文从公共数据集中选择五类遥感目标搭建实验数据集，通过消融实验验证所改进部分的有效性，与近年的经典目标检测方法进行对比实验验证本文方法的检测性能。

## 2. 网络结构

基于多尺度特征增强的遥感图像目标检测方法网络结构如图 1 所示，主要包含四个部分：主干网络 (Backbone Network)、多尺度特征增强金字塔网络 (Multi-Scale Feature-Enhanced Pyramid Network,

MSFE-FPN)、区域提议网络(Region Proposal Network)和级联检测器网络(Cascade Detector Network)。

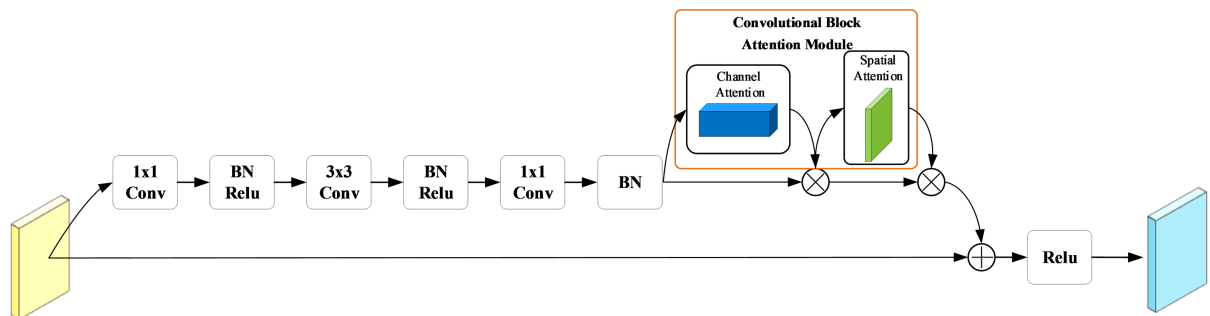


**Figure 1.** Overall network structure  
**图 1.** 整体网络结构

主干网络使用 HRNet [6] 替换 ResNet 来提取遥感图像的特征信息, 图片输入到主干网络后依次经历四个阶段提取特征信息。以输入三通道遥感图片为例, 在第一阶段, 首先经过卷积层将图像的特征维数从 3 扩展到 64; 然后经过四个添加注意力机制的瓶颈模块引导网络关注图像中的重要特征信息; 最后分别经过两个卷积块, 产生高低两个分辨率的并行子网输入到第二阶段。后续三个阶段与第一阶段类似, 逐步增加一条并行多分辨率子网, 在主干网络的第四阶段获得四个并行多分辨率子网输出。

## 2.1. 注意力机制

注意力机制在深度学习领域被广泛应用, 并且在目标检测任务中表现出优秀的性能。本文在 HRNet 第一阶段加入 CBAM (Convolutional Block Attention Module, CBAM) [7], 使主干网络能够自适应细化输出特征, 关注图像的重要特征区域, 增强其对复杂图像背景的抗干扰能力。加入 CBAM 的瓶颈块如图 2 所示。



**Figure 2.** Add CBAM to the bottleneck block  
**图 2.** 瓶颈块中添加 CBAM

在加入 CBAM 前使用  $1 \times 1$  卷积和  $3 \times 3$  卷积代替单独的  $3 \times 3$  卷积, 以降低图像维数、减少网络中的参数量。CBAM 包括注入通道注意力和空间注意力两部分。首先对特征图注入通道注意力, 通过均值池化和最大池化操作对特征图的位置信息进行聚合, 然后通过共享 MLP 生成通道注意特征图; 随后对特征图注入空间注意力, 在经平均池化和最大池化操作后将它们连接起来产生一个特征描述符, 通过卷积层生成空间注意特征图。

### 2.2. 多尺度特征增强金字塔网络

特征金字塔网络(Feature Pyramidal Network, FPN) [8]通过自下而上的路径和横向连接获得特征图中的语义信息和位置信息以提高多尺度目标检测精度。众多研究者提出促进高低层间的信息交流等策略以增强各层次的特征表达。例如, PANet [9], Balanced FPN [10], BiFPN [11]等。受上述研究启发, 本文构建一个多尺度特征增强金字塔网络(MSFE-FPN), 通过组合 HRNet 的多分辨率输出构建金字塔 A, 增强已有的特征金字塔 B, 然后两者相加输出多尺度特征增强金字塔。MSFE-FPN 结构如图 3 所示。

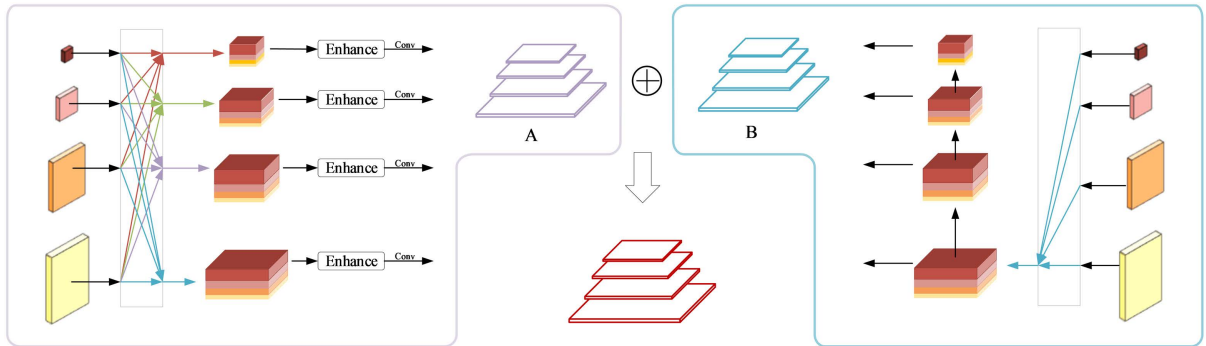


Figure 3. Multi-scale feature-enhanced pyramid network (MSFE-FPN)  
图 3. 多尺度特征增强金字塔网络(MSFE-FPN)

多尺度特征增强金字塔网络的构建主要有两个步骤: 组合 HRNet 各分辨率子网的特征输出; 注入自注意力权重, 强化各层特征信息。组合 HRNet 各分辨率子网的特征输出首先需要调整各分辨率输出到同一分辨率, 再对其进行连接。调整过程包括对高层特征图进行双线性插值上采样以及对底层特征图进行平均池化下采样操作, 调整过程公式表达如下所示。

$$P_{out} = \begin{cases} f_{up}(P_N), & N > out \\ P_N, & N = out \\ f_{down}(P_N), & N < out \end{cases} \quad (1)$$

其中,  $P_N$  为 HRNet 并行输出的特征图,  $f_{up}$  为双线性插值上采样操作,  $f_{down}$  为平均池化下采样操作。

强化各层特征信息是对特征图注入自注意力权重, 使网络根据全局的特征信息去加强有用的特征, 淡化无用的特征。输入特征图经全局平均池化操作获得自注意力权重, 将其转换到 0~1 范围后获得到特征图中, 具体过程及公式定义如下所示。

$$G' = G \oplus \text{Sigmoid} \left\{ f^{1 \times 1} \left\{ \text{Relu} \left\{ f^{1 \times 1} \left[ \text{AvgPool}(G) \right] \right\} \right\} \right\} \quad (2)$$

其中,  $G$  是输入特征图,  $G'$  是增强后的特征图输出, AvgPool 是平均池化操作,  $f^{1 \times 1}$  是卷积核尺寸为  $1 \times 1$  的卷积操作, Relu 和 Sigmoid 是激活函数(图4)。

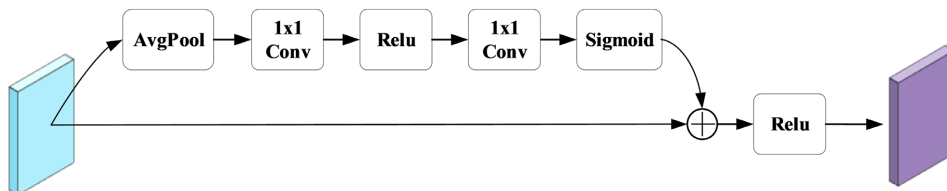


Figure 4. Feature information enhancement structure  
图 4. 特征信息增强结构

### 3. 实验结果与分析

#### 3.1. 实验数据集与评价指标

实验数据选自公开遥感图像数据集 DIOR [12]。由于 DIOR 数据集规模大、种类多，本文从中选择了五类尺度变化明显及有较多小目标的类别，包括飞机、船舶、储罐、烟囱和风力发电设备五类。各类别数量及训练、测试、验证集划分情况统计如表 1 所示。

**Table 1.** Experimental data set

**表 1.** 实验数据集

Object	Training	Validation	Test
airplane	344	169	168
ship	390	191	191
storage tank	649	326	327
chimney	202	102	102
windmill	404	201	202
Total	1989	989	990

使用平均精度(Average Precision, AP)和均值平均精度(mean Average Precision, mAP)作为评价指标。AP 是精确率 - 召回率(Precision-Recall)曲线下面的面积，mAP 是所有目标的平均精度的平均值。理想的目标检测网络可以在召回率增长的同时保持较高的精确率。AP 和 mAP 的公式定义如下。

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (4)$$

#### 3.2. 对比实验

将本文方法与近年的经典目标检测方法进行比较实验，选择对比的经典目标检测方法包括 YOLOv5、Faster R-CNN [13]、RetinaNet [14]、和 PANet [9]。使用 AP 和 mAP 指标对各目标检测网络的目标检测性能进行评估，检测结果如表 2 所示。从表中可以看出，本文方法具有最好的 mAP，并且在飞机、储罐和烟囱三个类别中具有很大的优势；与 RetinaNet、Faster R-CNN 和 PANet 相比，mAP 分别提高了 18.96%、18.16%和 13.82%。

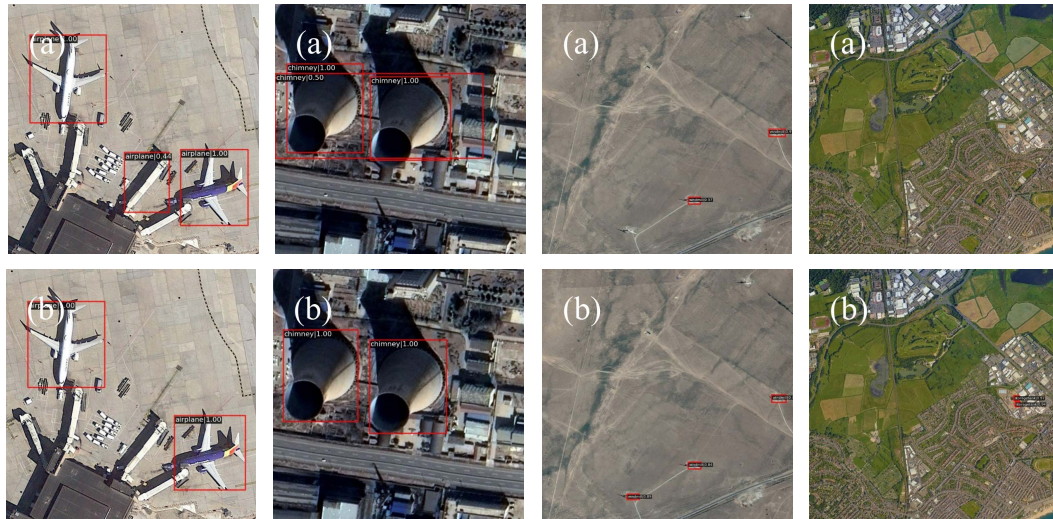
**Table 2.** Comparative experimental results

**表 2.** 对比实验结果

Method	AP/%					mAP/%
	airplane	ship	storage tank	chimney	windmill	
YOLOv5	72.2	<b>87.4</b>	68.7	69.7	78.7	75.34
Faster R-CNN	54.1	71.8	53.8	72.5	81.2	66.68
RetinaNet	53.7	71.2	45.8	73.2	85.5	65.88
PANet	61.9	71.7	62.3	72.5	<b>86.7</b>	71.02
<b>Proposed</b>	<b>91.1</b>	80.5	<b>83.8</b>	<b>89.7</b>	79.1	<b>84.84</b>



在上述对比实验中，YOLOv5 是经典目标检测方法中表现较好的一个，从数据集中选择部分遥感图像对 YOLOv5 和本文方法做可视化结果对比，对比结果如图 5 所示。同一遥感图片使用 YOLOv5 时会出现漏检、误检的情况，而本文方法能够正确地检出，结合表 2 实验结果可以得出：本文方法能够较好的完成遥感图像目标检测任务。



**Figure 5.** Comparison of object detection results in remote sensing images. (a) Use YOLOv5; (b) Use our method  
**图 5.** 遥感图像目标检测结果对比。(a) 使用 YOLOv5; (b) 使用本文方法

### 3.3. 消融实验

本文所提方法中主要改进两个部分：使用加入 CBAM 的 HRNet 作为主干网络；使用 MSFE-FPN 网络结构。为了证明这两部分的有效性，本节进行消融实验分析讨论。对比的基线模型为使用 HRNet 作为主干网络的 Cascade R-CNN 网络，使用 AP 和 mAP 作为评价指标，所得实验结果如表 3 所示。

**Table 3.** Ablation experiment results  
**表 3.** 消融实验结果

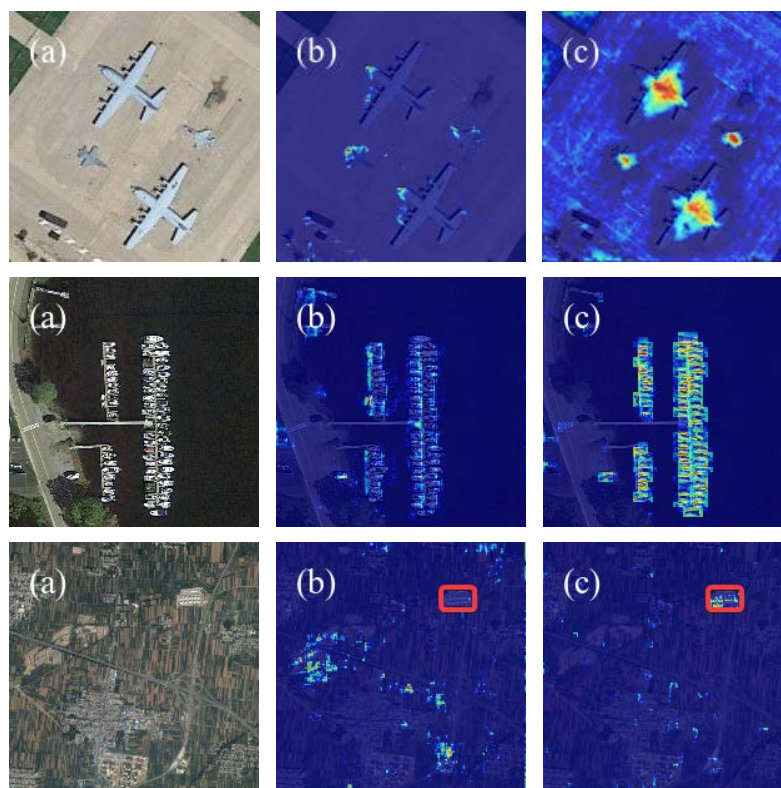
Model	AP/%					mAP/%
	airplane	ship	storage tank	chimney	windmill	
Cascade R-CNN	90.1	79.5	80.6	85.7	75.2	82.22
HRNet + CBAM	91.1	79.8	82.9	89.8	79.0	84.52
HRNet + MSFE-FPN	91.9	80.7	83.7	89.7	77.0	84.60
Proposed	91.1	80.5	83.8	89.7	79.1	84.84

从表 3 可以看出，引入 CBAM 后，各类别遥感目标检测平均精度均有提升，mAP 提高了 2.30 个百分点；使用 MSFE-FPN 后，除风力发电设备类别有所下降外，其余类别平均精度均有提升，mAP 也提升了 2.38 个百分点。为了证明整体结构的有效性，将 CBAM 与 MSFE-FPN 同时引入到网络中进行实验。在两个部分的共同作用下，前部分实验检测精度有所下降的风力发电设备类别提高了 3.9 个百分点，mAP 提高了 2.62 个百分点。由此可见，CBAM 一定程度上可以弥补 MSFE-FPN 的不足，这两个部分在提高

网络性能过程中具有互补性。这表明着本文方法所使用的加入 CBAM 的 HRNet 和 MSFE-FPN 可以有效提高网络的检测性能。

### 3.4. 检测结果可视化

特征热力图反映了网络更侧重关注图像中的哪块区域。本文方法和原始 Cascade R-CNN 的特征热力图响应对比如图 6 所示。从图中可以看出,相较于原始 Cascade R-CNN 方法,本文方法对于多尺度目标、密集小目标有较好的特征响应;对于复杂背景下的储罐目标检测(红色矩形框选中区域), Cascade R-CNN 方法并没有检测到目标区域,而本文方法在目标区域有正确的特征响应。



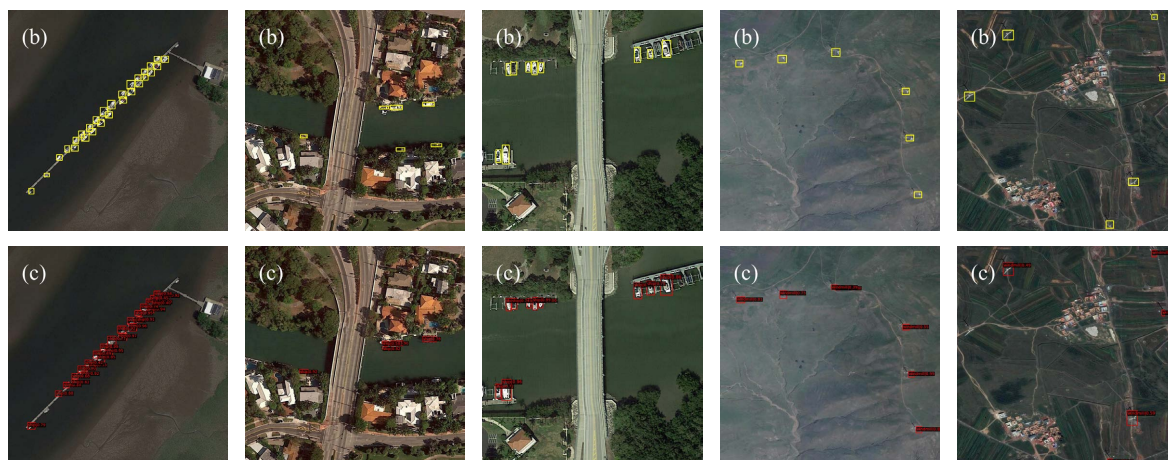
**Figure 6.** Comparison of feature heat maps. (a) Input images; (b) Feature heat maps of the Cascade R-CNN; (c) Feature heat maps of our method

**图 6.** 特征热力图对比。(a) 输入图片; (b) Cascade R-CNN 特征热力图; (c) 本文方法特征热力图

遥感小目标可用于目标检测的像素数较少,大大增加了遥感目标检测的难度。对此,筛选出验证数据集中存在遥感目标标注框小于  $30 \times 30$  的遥感图像进行检测,图片示例如图 7(a)所示。本文方法的检测结果如图 7(c)所示。在同标注图 7(b)的对比中可以看出,本文方法可以成功检出大部分遥感小目标,并且置信度都在 80%以上。







**Figure 7.** Remote sensing small object ( $<30 \times 30 \text{ Px}</math>) detection results. (a) Image to be tested; (b) Original labeling images; (c) Our method detection results$

**图 7.** 遥感小目标( $<30 \times 30 \text{ 像素}</math>)检测结果。(a) 待检测图; (b) 原始标注图; (c) 本文方法检测结果$

#### 4. 结束语

本文针对遥感图像目标检测中背景复杂、目标像素数少以及目标尺度变化大等问题,提出一种基于多尺度特征增强的遥感图像目标检测方法。使用 HRNet 网络作为主干网络并设计使用多尺度特征增强金字塔网络,增强对遥感目标位置信息提取以及多尺度特征表示;在主干网络 HRNet 中加入注意力机制,使网络关注有用图像区域,增强网络整体抗干扰能力。实验结果验证了本文方法的有效性,能够有效提高遥感图像目标检测精度。本文方法也有不足之处,所使用网络框架在检测速度方面检测推理较慢,后续工作将在轻量化网络模型的方向继续开展,进一步提升该方法的性能。

#### 基金项目

国家自然科学基金天文联合基金(U1831133)、上海市科委科技创新行动计划(21S31904200, 22S31903700)、中科院空间主动光电技术重点实验室开放基金(2021ZDKF4)。

#### 参考文献

- [1] 聂光涛, 黄华. 光学遥感图像目标检测方法综述[J]. 自动化学报, 2021, 47(8): 1749-1768.
- [2] Wang, B., Zhou, Y., Zhang, H. and Wang, N. (2019) An Aircraft Target Detection Method Based on Regional Convolutional Neural Network for Remote Sensing Images. 2019 *IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Beijing, China, 12-14 July 2019, 474-478. <https://doi.org/10.1109/ICEIEC.2019.8784637>
- [3] Ding, P., Zhang, Y., Deng, W.-J. Jia, P. and Kuijper, A. (2018) A Light and Faster Regional Convolutional Neural Network for Object Detection in Optical Remote Sensing Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **141**, 208-218. <https://doi.org/10.1016/j.isprsjprs.2018.05.005>
- [4] 姚群力, 胡显, 雷宏. 基于多尺度卷积神经网络的遥感目标检测研究[J]. 光学学报, 2019, 39(11): 346-353.
- [5] 刘楠, 毛昭勇, 王亦晨, 沈钧戈. 基于参数量和感受野可调的遥感目标检测方法[J]. 光子学报, 2021, 50(11): 302-313.
- [6] Ke, S., Yang, Z., Borui, J., *et al.* (2019) High-Resolution Representations for Labeling Pixels and Regions. <https://arxiv.org/abs/1904.04514>
- [7] Woo, S., Park, J., Lee, J.-Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., eds., *Computer Vision—ECCV 2018*, Springer, Cham, 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)



- 
- [8] Lin, T.-Y., Dollár, P., Girshick, R., *et al.* (2017) Feature Pyramid Networks for Object Detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21-26 July 2017, 2117-2125. <https://doi.org/10.1109/CVPR.2017.106>
- [9] Liu, S., Qi, L., Qin, H.F., Shi, J.P. and Jia, J.Y. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18-23 June 2018, 8759-8768. <https://doi.org/10.1109/CVPR.2018.00913>
- [10] Pang, J.M., Chen, K., Shi, J.P., Feng, H.J. and Ouyang, W.L. (2019) Libra R-CNN: Towards Balanced Learning for Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 15-20 June 2019. <https://doi.org/10.1109/CVPR.2019.00091>
- [11] Tan, M.X., Pang, R.M. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 13-19 June 2020, 10781-10790. <https://doi.org/10.1109/CVPR42600.2020.01079>
- [12] Li, K., Wan, G., Cheng, G., *et al.* (2020) Object Detection in Optical Remote Sensing Images: A Survey and a New Benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, **159**, 296-307. <https://doi.org/10.1016/j.isprsjprs.2019.11.023>
- [13] Ren, S.Q., He, K.M., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. <https://arxiv.org/abs/1506.01497>
- [14] Lin, T.-Y., Goyal, P., Girshick, R., *et al.* (2020) Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 318-327. <https://doi.org/10.1109/TPAMI.2018.2858826>