

基于改进 YOLOv5 的光学遥感图像水坝检测研究

薛继伟, 孙宇锐

(东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163000)

摘要:目标检测是计算机视觉领域的一个重要应用,针对光学遥感影像的目标检测任务也是当下的研究热点之一。现阶段科技进步的同时带来了一系列环境问题,环境保护已经成为当下值得关注的重点问题。水坝的建设是影响全球环境保护以及资源利用的一个重要因素,对水坝进行监测可以为环境保护工作提供参考依据。为了环境保护后续工作的开展,分析水坝在图像中的位置,该文针对高分辨率光学遥感影像中的水坝目标检测方法进行研究,对比了深度学习三个阶段较为典型的目标检测模型,根据实验结果选用精度较高的 YOLOv5 通用目标检测模型,并根据遥感图像背景复杂的特性结合 CBAM 注意力机制提高网络对图像中水坝目标的重点关注。在 DIOR 光学遥感目标检测数据集中提取含有水坝目标的图像并验证模型精度,实验表明 YOLOv5-CBAM 在并不显著增加模型大小的情况下比 YOLOv5 运算能力强,并且 AP50 可以达到 86.4%,比仅使用 YOLOv5 的模型 AP50 提高了 3.2 百分点。

关键词:遥感图像;目标检测;水坝;注意力机制;YOLO

中图分类号:TP751;TP391.4

文献标识码:A

文章编号:1673-629X(2023)05-0069-06

doi:10.3969/j.issn.1673-629X.2023.05.011

Research on Dam Detection of Optical Remote Sensing Image Based on Improved YOLOv5

XUE Ji-wei, SUN Yu-rui

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163000, China)

Abstract: Object detection is an important application in the field of computer vision, and the task of object detection for optical remote sensing images is also one of the current research hotspots. At this stage, the progress of science and technology has brought a series of environmental problems at the same time, so environmental protection has gradually become a key issue worthy of attention. The construction of dams has become an important factor affecting global environmental protection and resource utilization. Monitoring of dams can provide reference for environmental protection work. In order to carry out the follow-up work of environmental protection and analyze the position of the dam in the image, we study the dam target detection method in high-resolution optical remote sensing images, and compare the typical target detection models in three stages of deep learning. According to the experimental results the YOLOv5 general target detection model with higher accuracy is selected, and according to the complex characteristics of the remote sensing image background combined with the CBAM attention mechanism, the network's focus on the dam target in the image is improved. Extract images containing dam targets in the DIOR optical remote sensing target detection dataset and verify the model accuracy. Experiments show that YOLOv5-CBAM has stronger computing power than YOLOv5 without significantly increasing the model size, and the AP50 of YOLOv5 can reach 86.4%, which is 3.2 percentage points higher than that of the model using only YOLOv5.

Key words: remote sensing image; object detection; dam; attention mechanism; YOLO

0 引言

当今世界科学技术发展的同时也带来了一系列环境问题,生态环境的破坏将会给人们的日常生活带来严重影响,保护环境并及时采取措施已经刻不容缓^[1]。

水坝是影响全球碳循环和水循环的一个重要因素,对其进行监测是非常有必要的,想要监测水坝首先需要了解水坝在图像中的位置。对高分辨率遥感影像的目标分类识别进行研究,是对地观测系统进行图像分析

收稿日期:2022-06-22

修回日期:2022-10-26

基金项目:黑龙江省高等教育教学改革研究项目(SJGZ20200036, SJGY20200108);东北石油大学青年科学基金项目(2018QNL-56);东北石油大学引导性创新资金项目(2020YDL-15)

作者简介:薛继伟(1973-),女,教授,博士,通信作者,研究方向为人工智能与大数据;孙宇锐(1999-),女,硕士研究生,研究方向为计算机视觉。

的一种重要手段^[2],目标检测是将图像数据转化为应用成果的关键一环^[3]。目前大部分基于深度学习的目标检测算法只针对普通图像,对于遥感图像的一些特征提取能力有待提升^[4]。为了环境保护后续工作的开展,该文针对遥感图像中的水坝目标检测方法进行研究,使用检测通用目标的 YOLOv5 算法进行实验并结合 CBAM 注意力机制进一步提升水坝目标检测的精度。

1 相关工作

目标检测技术如今已经逐渐趋于成熟,从最初的 SIFT^[5]、HOG^[6]、DPM^[7]等传统目标检测算法发展到如今基于深度学习的目标检测算法,目标检测的精度和速度都在不断提高。基于深度学习的目标检测方法的发展主要分为三个阶段,分别是双阶段、单阶段和无锚框阶段^[8]。双阶段方法主要有 R-CNN 系列^[9]、SPP-Net 等,单阶段方法主要有 YOLO^[10]系列、SSD^[11]、EfficientDet^[12]等,无锚框方法主要有 CenterNet^[13]、FCOS^[14]等。

以上方法主要针对普通图像,而光学遥感图像具有背景复杂、目标形状尺度变化大等特征^[15],一些学者根据遥感图像的特征在通用的目标检测算法基础上进行改进。R2CNN^[16]方法改进了 Faster R-CNN 算法以适应遥感图像中宽较大的目标,YOLT^[17]方法将 YOLOv2 模型引入遥感图像检测领域,ROI Transformer 方法对水平框采用可学习参数训练方式转为旋转框,再对旋转框内部区域特征池化,R3Det^[18]方法对特征图重构解决特征不对齐的问题,SCRDet^[19]方法针对遥感图像背景复杂的问题引入并行的像素级注意力机制和通道注意力机制弱化背景信息,提高了检测性能。

目前在光学遥感目标检测领域大部分是针对舰船、运动场以及飞机和机场等目标的研究^[20],针对水坝目标的检测方法较少。在深度学习成为研究热点之前,对水坝的检测基本上使用传统的目标检测算法。

2020 年 Zou Caigang 等学者^[21]使用深度学习的方法提出了基于双阈值的检测网络模型,该模型在基本的单级目标检测网络基础上增加双阈值结构,对检测结果进行二次决策,提高了网络检测的准确性,并在水坝遥感数据集中进行测试证明了该网络的有效性。

针对水坝目标的特性,该文使用深度学习目标检测方法 YOLOv5 模型结合 CBAM 注意力机制,对水坝目标进行训练及验证,可以使得检测精度在只使用 YOLOv5 模型的基础上有所提高。

2 目标检测算法研究

2.1 YOLOv5 算法

YOLOv5 算法是单阶段目标检测方法,网络结构主要分为四个部分,分别是输入端、骨干网络(Backbone)、Neck 结构和 Head 输出层。

输入端为用户输入的图像,YOLOv5 在输入端对数据进行缩放、归一化、Mosaic 数据增强等预处理。Backbone 结构的作用是提取目标的通用特征,最新版本 YOLOv5 的 Backbone 包含 CSPDarkNet53 结构。Neck 部分位于 Backbone 结构和 Head 结构之间,可以进一步提取特征,提升特征的多样性和鲁棒性。最新版本 YOLOv5 的 Neck 结构使用了 SPPF 和 C3 模块。SPPF 模块通过多个大小为 5×5 的最大池化层串行传递输入特征大小。

YOLOv5 最新版本中的激活函数使用 SiLU,该损失函数没有上界,有下界,平滑且不单调,如公式(1)所示。

$$f(x) = x \cdot \text{sigmoid}(x) \quad (1)$$

Head 结构也是网络的输出层,包含三个预测分支,三个预测分支的总损失乘以 batchsize 可以得到用于更新梯度的损失。

YOLOv5 的损失函数包括定位损失、置信度损失以及分类损失三个部分,总的损失为三种损失之和。其中置信度损失采用二分类交叉熵损失函数(BCELoss),如公式(2)所示。

$$L = - \sum_{i=1}^N y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \quad (2)$$

其中, $\hat{y}^{(i)}$ 为预测标签, $y^{(i)}$ 为真实标签, N 为样本数量。

边界框(Boundingbox)回归损失使用 CIoU Loss^[22]。CIoU Loss 考虑了重叠面积、中心点距离、长宽比三种几何参数的计算,如公式(3)~公式(5)所示。

$$\text{CIoU} = \text{IoU} - \left(\frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v \right) \quad (3)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (4)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (5)$$

其中, α 和 v 为边界框的长宽比, w 和 h 表示预测框的宽和高, w^{gt} 和 h^{gt} 分别表示 GroundTruth 的宽和高。 b 表示预测框中心坐标的参数, b^{gt} 表示 GroundTruth 中心坐标参数。 ρ^2 表示两个中心点之间距离的平方。

2.2 CBAM 注意力机制

在计算机视觉领域中适当应用注意力机制可以在一定程度上增强对图像信息的处理。注意力模块能够在大量的信息中提取出重要信息从而增强对重点目标

的关注,这种关注是根据不同的分配权重获得。基于深度学习的目标检测中常用的注意力机制模块有SE^[23]、ECA^[24]、CA^[25]、CBAM^[26]等。该文在YOLOv5模型的网路结构中添加通道注意力模块与空间注意力模块相结合的注意力机制Convolutional Block Attention Module(CBAM)模块,可以进一步提高网络特征的表达能力。

CBAM 模块在 SE 注意力机制的基础上对通道注意力模块进行了改进,并且增加了空间注意力模块。通道注意力在前,空间注意力在后,这种排列方式增强了模块的特征提取能力。通道注意力机制将输入的特征图经过全局最大池化(Global Max Pooling)和全局平均池化(Global Average Pooling)后,分别经过多层感知机(MLP)得到通道注意力的权重,再经过一个激活函数 Sigmoid 将注意力权重进行归一化,最后通过乘法加权到原始特征图上,生成空间注意力需要的输入特征。AveragePooling 和 MaxPooling 共用一个 MLP 可以减少学习参数。

通道注意力模型如图 1 所示。

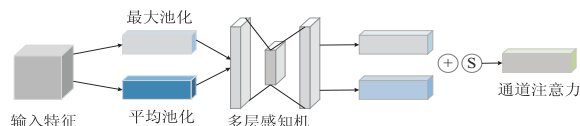


图 1 通道注意力模型

空间注意力模块在得到的通道注意力权重基础上进一步操作,与通道注意力部分相似,首先同样需要经过 MaxPooling 和 AveragePooling,将结果进行 concat 连接,再经过一个卷积层将特征图降维为一个通道。最后同样使用 Sigmoid 函数乘法加权相应特征,获得最终的空间注意力特征。

空间注意力模型如图 2 所示。

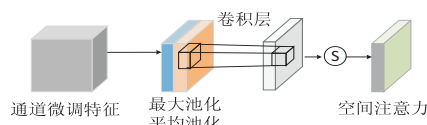


图2 空间注意力模型

文中将 CBAM 注意力模块插入 YOLOv5 模型 backbone 结构中最后一个 C3 模块与 SPPF 模块之间, 结合 CBAM 后的模型如图 3 所示。

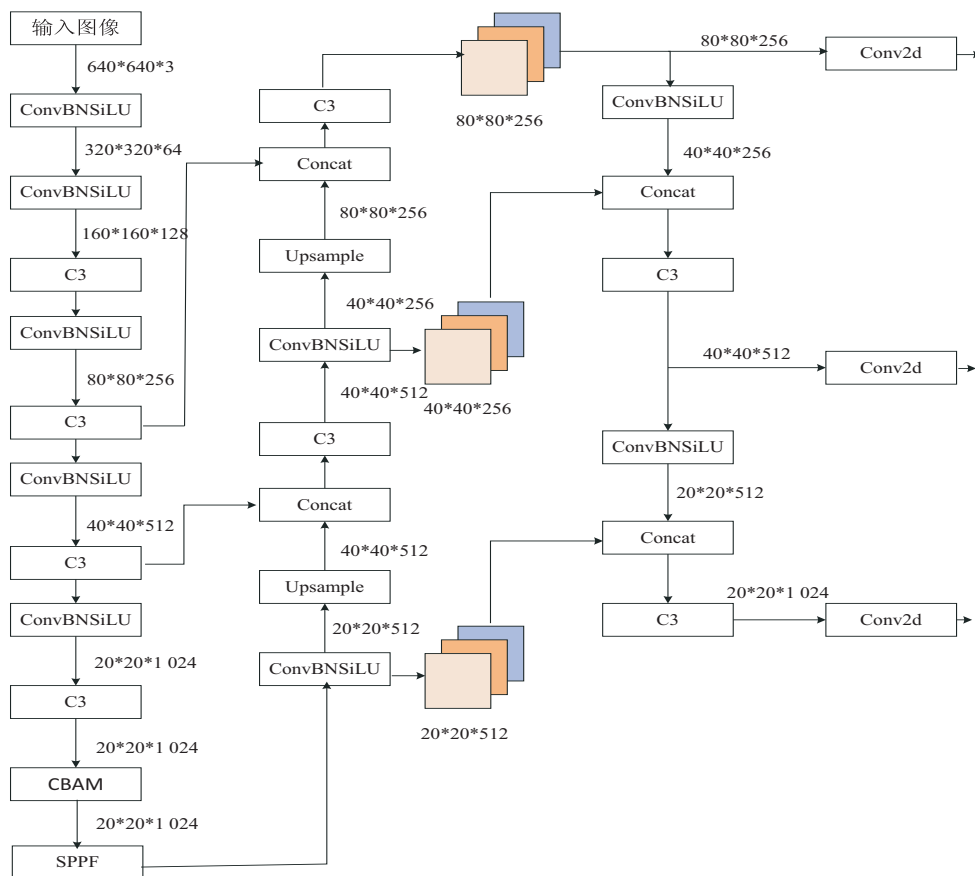


图3 YOLOv5+CBAM 网络结构

3 实验与分析

针对目标检测三个阶段中的经典算法在水坝遥感目标检测数据集上进行对比实验,并选用其中精度最

高的 YOLOv5 方法进行改进,结合 CBAM 注意力机制模块进一步提高检测精度。该文使用包含水坝目标的高分辨率遥感图像目标检测数据集 DIOR,并对数据集进行预处理,提取需要的图像,舍弃冗余图像,最后

对算法进行训练和验证,得到目标检测模型。

3.1 数据集

现有的常用光学遥感目标检测数据集中仅有 DIOR^[27] 数据集含有水坝目标,该数据集使用 LabelMe 工具标注,共有 23 463 张图像和 190 288 个实例,包含了 20 个目标类,每个类别包含约 1 200 张图像,其中包含水坝目标的图像 987 幅。DIOR 数据集具有四个特征,图像和实例数量规模较大,目标的尺寸变化范围在不同空间分辨率以及物体类间和类内都较大。以水坝目标为例,在仅含有一个水坝目标的图像中,水坝目标尺寸为大目标,但在含有其他目标的图像中,如在含有高尔夫球场目标和水坝目标共存的图像中,水坝相对来讲是一个小目标,如图 4。对数据集中目标大小进行分析,散点图分布如图 5,其中左图表示数据集中所有目标标注框中心点的 xy 坐标,右图表示标注框的高 (height) 和宽 (width),越接近右上角说明目标越大,越接近左下角说明目标越小。



图 4 DIOR 数据集水坝目标图像示例

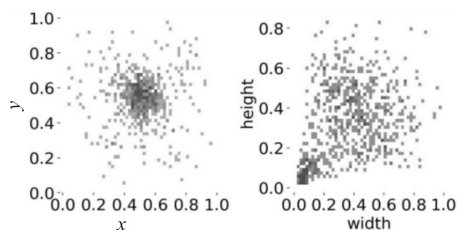


图 5 DIOR 数据集水坝目标分布散点图

3.2 数据预处理

DIOR 数据集的标注格式为 VOC 格式,为了避免 DIOR 数据集中其他类别目标对水坝目标检测精度的影响,需要将 images 文件夹中含有水坝标注信息的图像筛选出来。

由于依旧存在部分图像中含有多个类别的标注信息,需要对标注信息进行二次处理。删除初筛集多类别标注图像中除水坝类别外的其他类别标注,最终得到仅含有水坝类别标注的 xml 文件。

将数据集进行划分,20% 作为验证集,80% 作为训练集,网络的输入层将图像大小由 800 px * 800 px Resize 为 640 px * 640 px 并进行一系列缩放、翻转等操作。

在将图像输入到网络之前,需要对数据进行归一

化处理,使得数据经过处理后的限制区间为 $[0, 1]$ 。在对比实验中,除了 YOLOv5,其他方法在归一化的基础上进行了标准化,使得处理后的数据限制区间为 $[-1, 1]$ 。基于数据的均值和方差进行标准化,即在 R、G、B 三个维度减去均值并除以方差。在对比实验中除 YOLOv5 方法外,其他几种方法使用的均值在 R、G、B 上分别为 0.485、0.456、0.406,方差分别为 0.229、0.224、0.225。YOLOv5 在文中只进行归一化,不进行标准化。

3.3 设备及环境

CPU: Intel Core i7-11800H;

GPU: NVIDIA GeForce RTX3070;

语言: Python3.7;

深度学习框架: Pytorch1.11.0;

CUDA: 11.3.0;

编译器: Pycharm2021。

3.4 评价指标

单类别目标检测中的评价指标主要为 AP、AP₅₀。

目标检测的精确率 (Precision) 是指一组图像中模型检测出真正例占有所有目标的比例,召回率 (Recall) 表示所有真实目标中被模型检测出的目标比例。根据精确率与召回率的值可以绘制 PR 曲线,AP 即为 PR 曲线下的面积值,可以用积分计算,如公式 (6),其中 p 表示 PR 曲线中作为纵坐标的精确率值, r 表示作为横坐标的召回率值。对目标检测模型性能的评估通常需要与 IoU 阈值联系起来,而 AP₅₀ 表示的是 IoU 阈值为 0.5 时,AP 的测量值。

$$AP = \int_0^1 p(r) dr \quad (6)$$

3.5 训练和验证

该文分别使用双阶段模型 FasterR-CNN、单阶段模型 YOLOv5、EfficientNet 以及 AnchorFree 模型 CenterNet、FCOS 进行训练,并选择 AP50 相对较高的 YOLOv5 模型进行改进实验。

训练设置 batchsize 为 8, epoch 为 300, 学习率为 0.01, YOLOv5 模型使用 Backbone 为 DarkNet53, FasterR-CNN、FCOS 和 CenterNet 使用 ResNet50 作为 Backbone, EfficientNet 使用 RetinaNet 作为 Backbone。对比实验结果如表 1。

在 YOLOv5 的基础上结合 CBAM 注意力机制对水坝单目标进行训练,训练设置同上。结果表明添加 CBAM 注意力机制模块可以提高对水坝目标的检测精度。YOLOv5 和 YOLOv5-CBAM 的 PR 曲线对比如图 6 所示,结合 CBAM 训练得到的 best.pt 最佳模型在验证集中的 AP₅₀ 可以达到 86.4%, 未结合 CBAM 的模型 AP₅₀ 为 83.2%。

表1 水坝目标检测算法实验结果对比

算法	Backbone	AP ₅₀ /%	GFLOPs	Parameters	模型大小/MB
FasterRCNN	ResNet-50+FPN	78.2	—	—	315
CenterNet	ResNet-50	61.1	—	—	113.8
FCOS	ResNet-50	61.6	—	—	244
EfficientNet	RetinaNet	62.9	—	—	212
YOLOv5	DarkNet53	83.2	15.8	7 012 822	13.7
YOLOv5-CBAM	DarkNet53	86.4	16.1	7 218 744	14.1

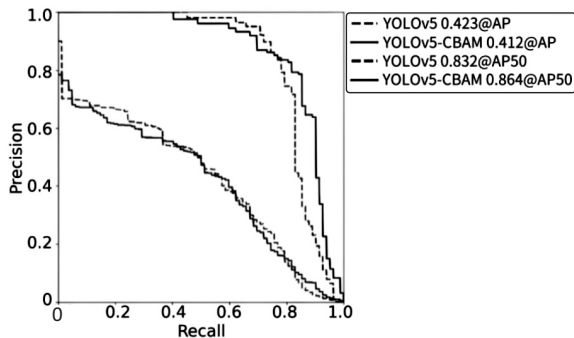


图6 PR曲线对比

在验证过程中发现,有部分图像产生漏检情况如图7左,与其他图像相比水坝在图中目标较小,在下一步的改进中考虑提高算法对小目标的检测能力。图7右中产生了误检情况,检测器将与水坝特征相似的湖泊边沿误检为水坝进行了标记,说明未来算法改进应考虑到与水坝具有相似特征的目标影响,如桥梁等。

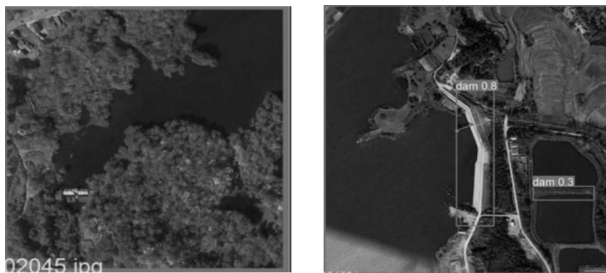


图7 漏检和误检情况

3.6 测试

测试集为根据全球水坝地理位置信息数据集GOODD在谷歌地图中靶向搜索国内水坝图像并进行标注,共100张,将上述训练得到的模型应用于测试集中,测试结果如图8所示,其中上为标签,下为检测结果。

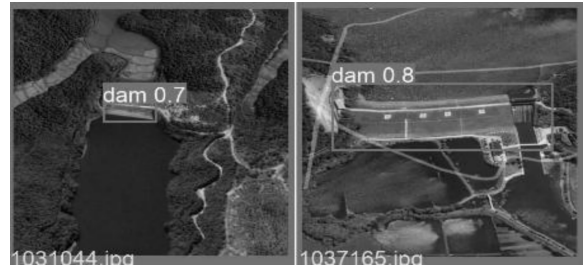
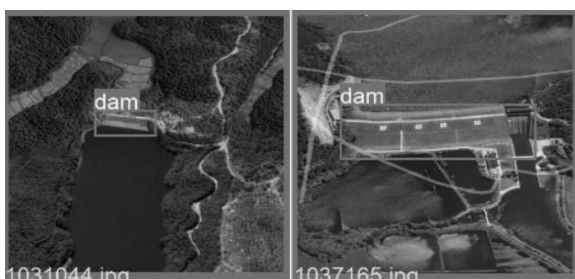


图8 YOLOv5-CBAM检测结果

4 结束语

该文主要针对光学遥感图像中的水坝目标进行检测,在单阶段目标检测模型YOLOv5的基础上在Backbone阶段引入了CBAM注意力模型,该方法结合了通道注意力和空间注意力,增强模型对水坝目标的关注进而提高对水坝目标的检测精度。在验证集上,可以将YOLOv5的AP₅₀从83.2%提升至86.4%,有针对性地单独训练水坝目标可以降低误检率。后续可以根据此模型建立水坝目标检测系统,扩展该模型在遥感图像目标检测范围内的应用。

参考文献:

- [1] 张学礼. 水坝建设对溱潼河流域生态环境的影响探讨[J]. 生态经济, 2017, 33(7): 171-174.
- [2] 廖育荣, 王海宁, 林存宝, 等. 基于深度学习的光学遥感图像目标检测研究进展[J]. 通信学报, 2022, 43(05): 190-203.
- [3] 曹家乐, 李亚利, 孙汉卿, 等. 基于深度学习的视觉目标检测技术综述[J]. 中国图象图形学报, 2022, 27(6): 1697-1722.
- [4] 饶宸. 基于深度学习的光学遥感目标检测技术研究[D]. 南昌: 东华理工大学, 2020.
- [5] 王洪森, 杨健, 殷君君. 基于Wishart梯度和SIFT方法的极化SAR图像配准[J]. 中国科学: 信息科学, 2022, 52(4): 687-699.
- [6] 韩松来, 王钰婕, 王星, 等. 多尺度PCA-HOG遥感异源图像匹配算法[J]. 国防科技大学学报, 2022, 44(1): 146-155.
- [7] 崔艺涵, 陈涛, 陈宝刚. 基于DPM和KCF的十字靶标检测与跟踪[J]. 液晶与显示, 2018, 33(12): 1026-1032.

- [8] 赵立新,邢润哲,白银光,等.深度学习在目标检测的研究综述[J].科学技术与工程,2021,21(30):12787-12795.
- [9] REN Shaoqing, HE Kaiming, 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.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of European conference on computer vision. Berlin: Springer, 2016: 21-37.
- [12] TAN Mingxing, PANG Ruoming, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 10781-10790.
- [13] DUAN K W, BAI S, XIE L X, et al. CenterNet: keypoint triplets for object detection[C]//Proceedings of 2019 IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 6568-6577.
- [14] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of 2019 IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 9626-9635.
- [15] 刘涛. 基于单阶段检测网络的遥感图像目标检测研究[D]. 西安: 陕西师范大学, 2020.
- [16] JIANG Y, ZHU X, WANG X, et al. R2CNN: rotational region CNN for orientation robust scene text detection[J]. IEEE Access, 2017(7): 126-129.
- [17] VAN ET TEN A. You only look twice: rapid multi-scale object detection in satellite imagery[J]. arXiv: 1805.09512, 2018.
- [18] YANG X, LIU Q, YA J, et al. R3Det: refined single-stage detector with feature refinement for rotating object[C]//Proceedings of the 35th AAAI conference on artificial intelligence. Menlo Park: AAAI, 2021: 3163-3171.
- [19] YANG X, YANG J C, YAN X K, et al. SCRDet: towards more robust detection for small cluttered and rotated objects[C]//Proceedings of 2019 IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 8231-8240.
- [20] 王浩桐, 郭中华. 改进 SSD 的飞机遥感图像目标检测[J]. 液晶与显示, 2022, 37(1): 116-127.
- [21] ZOU Caigang, ZHU Lei, LV Suidong. Remote sensing image dam detection based on dual threshold network[C]//2020 Chinese control and decision conference (CCDC). Hefei: IEEE, 2020: 1451-1456.
- [22] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the 34th AAAI conference on artificial intelligence. Menlo Park: AAAI, 2020: 12993-13000.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2020: 2011-2023.
- [24] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 11531-11539.
- [25] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Los Alamitos: IEEE, 2021: 13713-13722.
- [26] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Berlin: Springer, 2018: 3-19.
- [27] KE Li. Object detection in optical remote sensing images: a survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159(C): 296-307.