

基于改进 FCOS 网络的光学遥感影像目标检测算法

杨玉春, 王腾军, 任会涛, 杨 耘

(长安大学 地质工程与测绘学院, 西安 710054)

摘要: 针对当前遥感影像背景复杂、目标尺度小等情况导致的检测精度偏低的问题, 基于 FCOS 网络提出了一种结合位置注意力和感受野增强的遥感影像目标检测算法 PARF-FCOS; 该算法构造了一种位置注意力模块, 并利用该模块对特征提取网络进行改进, 增强网络对目标信息的提取能力; 在特征融合阶段使用感受野模块 (RFB, receptive field block) 增强浅层特征图, 利用目标上下文信息进行辅助判断, 提升网络对小尺度目标的检测能力; 在训练过程中, 引入距离交并比损失 (DIOU loss, distance intersection over union loss) 进行边界框回归, 通过优化目标框与预测框中心点之间的距离, 使回归过程更加平稳和准确; 在公开数据集 DIOR 上评估了 PARF-FCOS 目标检测算法, 实验结果表明, 相较于原始 FCOS, 算法的平均精确度均值提高了 4.3%, 达到 70.4%, 检测速度达到 23.2 FPS。

关键词: 遥感影像处理; 目标检测; 卷积神经网络; 注意力机制; 无锚框

Object Detection Algorithm of Optical Remote Sensing Imagery Based on An Improved FCOS Network

YANG Yuchun, WANG Tengjun, REN Huitao, YANG Yun

(School of Geology Engineering and Geomatics, Chang'an University, Xi'an 710054, China)

Abstract: Aiming at the problem of low detection accuracy caused by complex background and small objects in remote sensing imageries, a remote sensing imagery object detection algorithm PARF-FCOS based on FCOS network is proposed. The algorithm constructs a position attention module, and uses the module to reconstruct the feature extraction network to enhance the ability of the network to extract target information; In the feature fusion stage, RFB (receptive field block) is used to enhance the shallow feature map, and the target context information is used for auxiliary judgment to improve the detection ability of the network for small-scale objects; During training, DIOU loss (distance intersection over union loss) is introduced for boundary box regression. By optimizing the distance between the center point of the target box and the prediction box, the regression process is more stable and accurate. Experiments are carried out with public dataset DIOR. Compared with the original FCOS, the mean Average Precision of the algorithm is improved by 4.3%, up to 70.4%, and the detection speed is 23.2 FPS.

Keywords: remote sensing imagery processing; object detection; convolutional neural network; attention mechanism; anchor-free

0 引言

遥感技术是 20 世纪 60 年代出现的一种空间信息探测技术, 能够在距离被探测目标十分遥远的位置, 通过影像传感器收集目标辐射和反射的电磁波信息, 根据物体的物理光谱特性对收集的电磁波信息进行分析和解译, 由此实现对目标地物的远距离探测。随着航空航天技术、图像传感器技术和计算机技术的发展, 遥感影像的空间分辨率越来越高, 高分辨率遥感影像包含着更为丰富、更为细致的地物信息, 推动了遥感影像处理技术的发展和进步。作为遥感影像处理领域的一个基本问题, 遥感影像目标检测是识别给定的航空或卫星影像是否包含一个或多个感兴趣类别的目标, 并确定目标在影像中的位置的过程, 其在环境监

测、地理信息系统更新、精准农业和城市规划等应用中发挥着重要作用^[1]。近年来, 随着遥感技术的迅猛发展, 高分专项工程的实施和商业化卫星遥感服务的兴起, 影像数量呈现爆炸式的增长趋势, 快速增长的影像数量对目标检测方法提出了更高的要求^[2]。

1 相关研究

传统的遥感影像目标检测多采用特征工程和分类器训练相结合的检测策略。例如姚远等^[3]通过提取影像的 Haar 和 HOG 特征并结合 AdaBoost 和 SVM 进行分类实现了飞机目标的检测。黎经元等^[4]将海陆分割与舰船目标检测相结合, 通过分析港口中舰船的分布规律和纹理特点, 提出了一种结合边缘线梯度定位和聚合通道特征的舰船检测方

收稿日期: 2022-03-03; 修回日期: 2022-03-29。

基金项目: 陕西省自然科学基金(2022JM-163)。

作者简介: 杨玉春(1997-), 男, 江西南昌人, 硕士研究生, 主要从事遥感影像分析与处理方面的研究。

引用格式: 杨玉春, 王腾军, 任会涛, 等. 基于改进 FCOS 网络的光学遥感影像目标检测算法[J]. 计算机测量与控制, 2022, 30(5): 6-12.

法。此类方法的检测效果十分依赖先验的特征设计, 且设计的特征通常只适用于特定目标, 当需要检测多类目标时, 效果往往较差。

深度学习特别是深度卷积神经网络 (DCNN, deep convolutional neural network)^[5] 凭借其优秀的特征提取能力, 目前已被广泛地应用于遥感影像目标检测中。当前基于 DCNN 的目标检测算法主要被分为两类, 以 R-CNN^[6] 为代表的双阶检测算法和以 YOLO^[7]、SSD^[8] 为代表的单阶检测算法。文献 [9] 等通过在 Faster R-CNN 的基础上引入 STN 结构, 引导网络自主学习旋转不变特征, 解决了目标旋转角度过大的问题。文献 [10] 在 Faster-RCNN 模型中引入注意力机制, 使用 SE 模块校准特征通道权重, 以较低的计算成本获得了显著的性能提升。文献 [11] 通过在 SSD 的基础上添加注意力分支, 有效的提升了网络获取目标位置信息的能力。文献 [12] 设计了一种基于 Faster R-CNN 的 Radar 目标检测模型, 显著的降低了目标的虚警率。以上算法均是基于锚框的检测算法, 通过人工预设的锚框来预测目标的位置, 依据待检测目标的形状、大小、密集程度等信息在图像上生成大量的锚框, 再通过这些锚框来进行目标边界框回归。然而预设的锚框会占用大量的存储和运算资源, 导致模型复杂度高, 计算量大。此外, 锚框引入的额外超参数需要手动设置, 大幅降低了网络的训练效率。

针对基于锚框目标检测算法的缺陷, 许多无锚框的目标检测方法被提出。文献 [13] 提出了 ConerNet, 该方法将目标检测问题转化关键点检测问题, 通过预测目标边界的左上角点和右下角点实现目标的检测。文献 [14] 提出了一种全卷积一阶 (FCOS, fully convolutional one-stage) 目标检测算法, 采用逐像素预测的方法对目标进行分类和回归, 并使用中心度预测分支来抑制低质量检测边框, 在 COCO 数据集上取得了不错的检测效果。与自然场景图像相比, 遥感影像中目标所处的背景信息复杂多变, 且更易受到环境噪声的干扰。此外, 遥感影像中小尺度目标居多, 更容易导致误检和漏检的发生。针对遥感影像目标检测中存在的难点, 本文以无锚框的 FCOS 检测网络为基础网络, 提出了一种结合位置注意力和感受野增强特征的目标检测算法。该算法首先设计了一种注意力模块, 并将其插入特征提取网络, 提升网络的特征提取能力; 其次,

使用感受野模块处理浅层特征图以增大感受野, 获取目标上下文信息, 提升小尺度目标的检测精度; 此外, 在训练过程中引入 DIoU loss, 使边框回归过程更加平稳和准确。本文将该算法简称为 PARF-FCOS。

2 PARF-FCOS 目标检测算法

2.1 PARF-FCOS 网络结构

PARF-FCOS 目标检测算法的网络结构如图 1 所示, 主要包括特征提取模块、特征融合模块和检测模块 3 个部分。特征提取模块采用融合位置注意力 (PA, position attention) 的残差网络 (ResNet)^[15] 作为骨架网络, 用于提取影像的多尺度特征信息, 其中 PA 模块引导网络关注有用特征, 抑制无用特征, 从而提升特征图的表达能力。骨架网络生成特征图 C1~C5, 大小分别为原图的 1/2、1/4、1/8、1/16 和 1/32。特征融合模块仅处理特征图 C3、C4 和 C5, 它先采用 RFB 结构处理浅层特征图 C3, 用来扩充浅层特征的感受野大小, 增强网络获取目标上下文信息的能力。然后进行特征融合, 使用 1×1 卷积对特征图 C5 进行降维处理得到特征图 P5, 经过 2 倍上采样再与特征图 C4 相加, 得到特征图 P4, 同理特征图 P4 经过 2 倍上采样再与特征图 C3 相加, 得到特征图 P3。此外, 模型还在特征图 P5 的基础上通过两个步长为 2 的 3×3 卷积得到两个深层特征图 P6、P7, 大小为原图的 1/64 和 1/128。将特征图 P3、P4、P5、P6、P7 输入检测模块进行预测, 预测信息包括目标类别得分、中心度预测值和边框回归参数。最后利用非极大值抑制算法滤除得分较低的预测框, 得到最终检测结果。

2.2 位置注意力模块

注意力机制来源于人类的视觉机制, 能够使网络专注

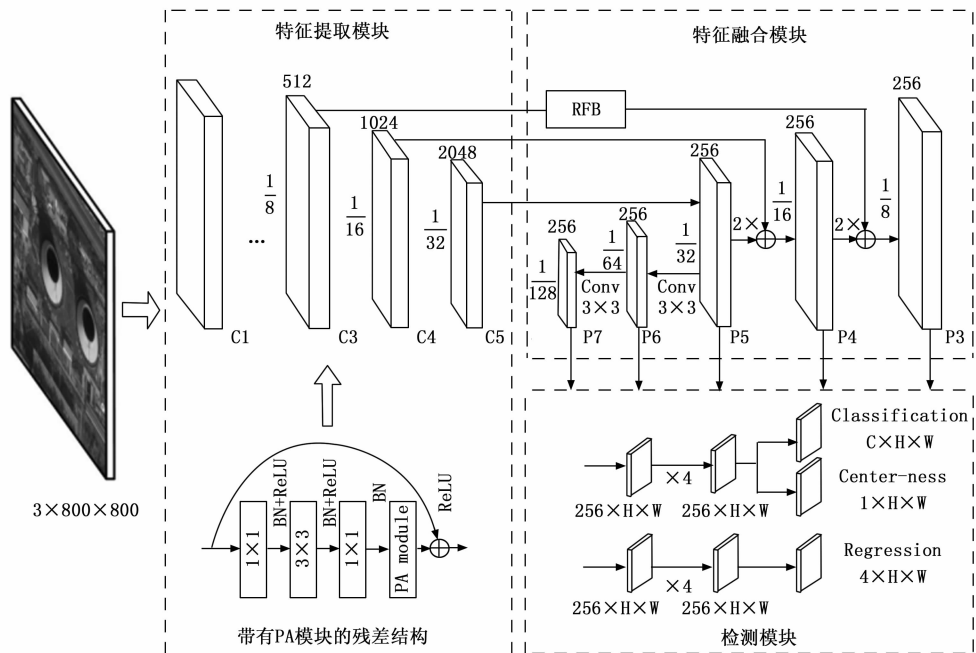


图 1 PARF-FCOS 网络结构图

于有效信息，过滤无用信息，从而提高信息的利用率。挤压激励 (SE, squeeze and excitation)^[16] 模块是一种常用的注意力模块，通过对特征通道的相互关系进行建模，自适应的校准特征权重，实现性能的提升。然而 SE 模块仅考虑了通道维度上的特征关系，忽视了空间信息对于网络性能的影响，导致注意力权重无法在空间维度上突出目标，限制了网络对于目标信息的提取。针对这个问题，本文提出了 PA 模块，其主体部分由两条注意力分支结构构成，分别编码了单个方向的位置信息，两个分支之间相互协作，共同构建特征的空间位置关系，其结构如图 2 所示。对于输入特征 $X \in \mathbb{R}^{C \times H \times W}$ ，首先经过池化核大小为 $H \times 1$ 和 $1 \times W$ 的一维平均池化，生成一对特征描述符 $Z_w \in \mathbb{R}^{C \times 1 \times W}$ 和 $Z_h \in \mathbb{R}^{C \times H \times 1}$ ，分别编码了水平和垂直位置信息，计算公式如下：

$$Z_w(c, w) = \frac{1}{H} \sum_{j=1}^H X(c, j, w) \quad (1)$$

$$Z_h(c, h) = \frac{1}{W} \sum_{i=1}^W X(c, h, i) \quad (2)$$

然后通过一个共享的卷积神经网络，并结合 Sigmoid 函数计算得到特征权重 $S_w \in \mathbb{R}^{C \times 1 \times W}$ 和 $S_h \in \mathbb{R}^{C \times H \times 1}$ ，其中共享参数可以降低网络复杂度，特征权重的计算公式如下：

$$S_w = \sigma(F_2(\delta(F_1(Z_w)))) \quad (3)$$

$$S_h = \sigma(F_2(\delta(F_1(Z_h)))) \quad (4)$$

式中， F_1, F_2 表示 1×1 卷积， δ 为非线性函数， σ 为 Sigmoid 函数。最后 S_h 和 S_w 之间相互协作，调整特征权重，计算公式为：

$$\hat{X} = X \times S_h \times S_w \quad (5)$$

2.3 结合位置注意力的特征提取网络

FCOS 使用 ResNet50 作为特征提取网络，ResNet50 由多个残差结构单元堆叠而成，通过卷积融合局部感受野内的空间和通道信息来提取图像特征，并且默认所有特征的权重相等。然而在特征提取的过程中，网络应当“重视”有用特征，“忽视”无用特征和作用小的特征。因此，本文在残差结构中添加 PA 模块来重新标定特征权重，提升网络的表达能力，使模型判断更加准确。添加 PA 模块后的残差结

构如图 3 所示，其中前 3 个卷积层用来提取特征信息，PA 模块用来调整特征权重，短路连接用来传递跨层信息。

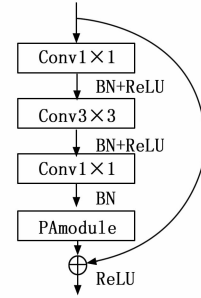


图 3 带有位置注意力的残差结构

基于上述的注意力残差结构，构造了位置注意力网络 PA-Net。ResNet50 和 PA-Net 的结构参数如表 1 所示。

表 1 特征提取网络参数

网络层名称	ResNet50	PA-Net
Conv1	$7 \times 7, 64, \text{stride}2$	$3 \times 3, \text{max pooling}, \text{stride}2$
Conv2	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \\ 1 \times 1, [16, 256] \end{bmatrix} \times 3$
Conv3	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \\ 1 \times 1, [32, 512] \end{bmatrix} \times 4$
Conv4	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \\ 1 \times 1, [64, 1024] \end{bmatrix} \times 6$
Conv5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \\ 1 \times 1, [128, 2048] \end{bmatrix} \times 3$

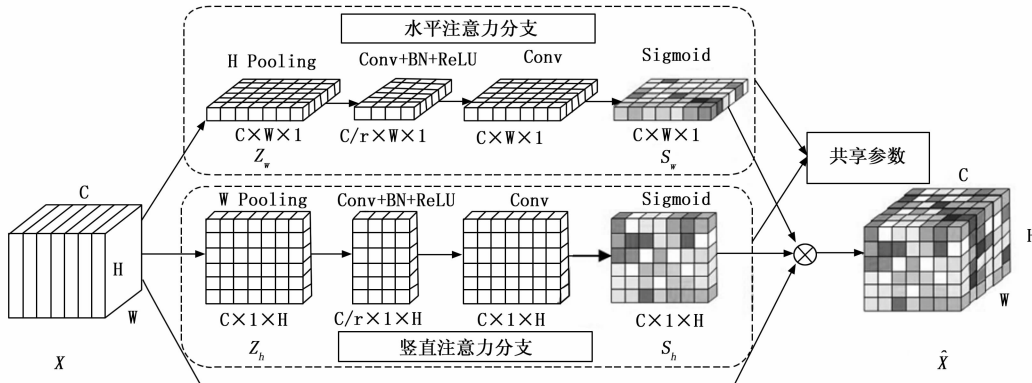


图 2 位置注意力模块结构图

2.4 感受野增强模块

FCOS 利用不同层级的特征对影像目标进行多尺度分层检测，浅层特征分辨率高纹理细节信息丰富，用于检测影像中的小尺度目标，但同时浅层特征的感受野小，导致网络无法利用目标周围的上下文信息进行辅助判断，容易造成小尺度目标的漏检和误检。为增大特征图感受野，本文在 FCOS 中引入感受野模块^[17] (RFB,

receptive field block)。RFB 通过构造不同的卷积分支来实现不同的感受野, 与 Inception^[18] 结构简单的堆叠卷积层不同, RFB 还考虑了感受野大小与偏心度的关系, 利用空洞卷积的空洞率 (r) 来实现不同的偏心度。RFB 的结构如图 4 所示, 包含并行的 3 条卷积分支, 卷积核参数分别为连续的 1×1 、 3×3 , 连续的 1×1 、 3×3 和 3×3 ($r=3$), 连续的 1×1 、 3×3 、 3×3 和 3×3 ($r=5$), 对应的感受野大小分别为 3×3 , 9×9 , 15×15 。将各卷积分支的输出特征进行拼接, 利用 1×1 卷积进行特征融合, 最后使用残差连接与原始特征进行相加。

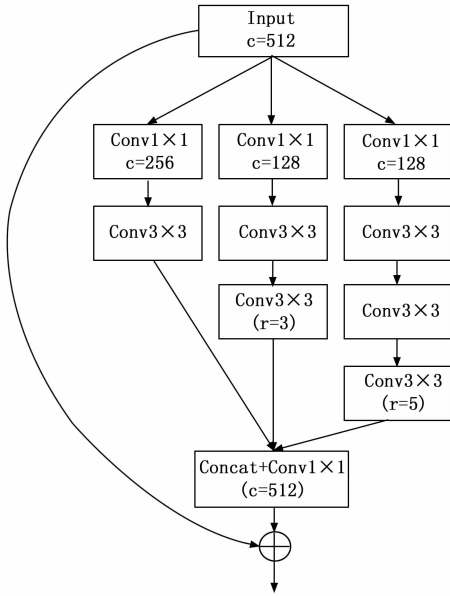


图 4 RFB 结构图

2.5 距离交并比损失

为了优化边界框的回归过程, 引入距离交并比损失 (DIoU loss)^[19] 作为边界框回归的损失函数, DIoU loss 的计算公式为:

$$L_{DIoU} = 1 - \frac{|B \cap B^{st}|}{|B \cup B^{st}|} + \frac{\rho^2(b, b^{st})}{c^2}, \quad (6)$$

式中, B 为目标真实框, B^{st} 为目标预测框, b 为预测框坐标中心点, b^{st} 为真实框坐标中心点, ρ 为两框中心点之间的欧式距离。 c 为能同时覆盖预测框和真实框的最小外接矩形的对角线长度。与原网络使用的 IOU loss 相比, DIOU loss 额外考虑了真实框与预测框之间的距离, 使边框回归更加准确, 同时还避免了两框之间距离过远而导致的梯度消失的问题。

3 实验与分析

3.1 数据集与评价指标

为了验证 PARF-FCOS 目标检测算法的性能, 本文在大型遥感影像目标检测数据集 DIOR^[20] 上进行了实验。DIOR 数据集由西北工业大学于 2020 年发布, 包含来自 Google Earth 的 23 463 幅光学遥感影像, 影像尺寸大小统

一至 800 像素 \times 800 像素, 空间分辨率范围从 0.5 m 到 30 m。数据集共包含 192 472 个实例对象, 每个实例对象均采用水平边框进行标注, 覆盖飞机、棒球场、汽车、大坝、立交桥、港口等 20 个目标类别。数据集随机地选取了 11 725 幅影像作为训练集, 11 738 幅影像作为测试集, 各占数据集总量的 50%。

使用平均精确度 (AP)、平均精确度均值 (mAP) 和每秒帧率 (FPS) 来评估网络的性能。AP 用于评价单类目标的检测精度, 由精确率 (P) 和召回率 (R) 计算得到。P 表示预测正确的正样本数占预测为正样本总数的比率, R 表示预测正确的正样本数正样本总数的比率, 是目标检测领域内的重要精度指标, 计算公式如下:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (7)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (8)$$

其中: N_{TP} 表示正样本被正确识别为正样本的数量, N_{FP} 表示负样本被错误识别为正样本的数量, N_{FN} 表示正样本被错误识别为负样本的数量。AP 值为 P 随 R 的变化曲线即 P-R 曲线与坐标轴围成的面积, 计算公式为:

$$A_p = \int_0^1 P(r) dr \quad (9)$$

mAP 用来评价模型的整体检测精度, 由各类目标的 AP 值取平均得到:

$$V_{mAP} = \frac{1}{n} \sum A_p, \quad (10)$$

其中: n 为数据集中目标类别数。

3.2 实验配置与模型训练

本文网络的训练和测试均在深度学习服务器上进行, 服务器的处理器 (CPU) 为 Intel (R) Xeon (R) Gold 6248, 显卡 (GPU) 为 Tesla V100, 操作系统为 CentOS7.5, 使用 Python3.7 编程语言和 Pytorch1.6 深度学习框架搭建网络, 并结合 CUDA10.0 和 CUDNN7.6 工具包进行加速, 具体的配置信息如表 2 所示。

表 2 实验配置

参数	配置
操作系统	CentOS7.5
CPU	Intel(R)Xeon(R)Gold 6248
GPU	Tesla V100
编程语言	Python3.7
深度学习框架	Pytorch1.6
GPU 加速工具	CUDN10.0、CUDNN7.6

训练 PARF-FCOS 目标检测算法的损失函数由分类损失函数、边界框回归损失函数和边框中心度损失函数三部分组成, 计算公式如下:

$$L(\{p_{x,y}\}, \{t_{x,y}\}, \{s_{x,y}\}) = \frac{1}{N_{pos}} \sum_{x,y} L_{cls}(p_{x,y}, c_{x,y}^*) +$$

$$\frac{\lambda}{N_{pos}} \sum_{x,y} 1_{\{c_{x,y} > 0\}} L_{reg}(t_{x,y}, t_{x,y}^*) + \frac{\mu}{N_{pos}} \sum_{x,y} 1_{\{c_{x,y} > 0\}} L_{cen}(s_{x,y}, s_{x,y}^*) \quad (11)$$

其中： $p_{x,y}$ 、 $t_{x,y}$ 和 $s_{x,y}$ 分别表示网络在 (x,y) 位置像素点预测的目标类别、边界框参数和中心度， $c_{x,y}^*$ 、 $t_{x,y}^*$ 和 $s_{x,y}^*$ 为相对应的真实值， N_{pos} 表示正样本的数量， $1_{\{ \cdot \}}$ 为指示函数， L_{cls} 、 L_{reg} 和 L_{cen} 分别为分类损失函数、边界框回归损失函数和边框中心度损失函数， λ 和 μ 为平衡系数。

本文在 DIOR 数据集上对 PARF-FCOS 进行训练，训练步骤如下：

- 1) 使用 Pytorch 深度学习框架搭建网络模型；
- 2) 选择随机梯度下降法 (SGD, stochastic gradient descent) 作为网络的优化器，其中动量和权重衰减系数分别设置为 0.9 和 0.0005；
- 3) 初始学习率设置为 0.005，并且在训练过程中使用余弦学习率策略将学习率降低至原来的 1/10，保证网络在训练过程中能够更快地收敛；
- 4) 每批次送入 8 张影像进行正向传播，计算网络的损失值；
- 5) 依据损失值反向计算参数的梯度，并根据梯度调整网络参数；
- 6) 重复步骤 4) ~5) 直到网络收敛或者达到最大迭代次数。

在训练了 90 个 epoch，迭代了 125 000 多次后，模型的损失趋于稳定，总损失大致稳定在 0.7 左右，损失变化情况如图 5 所示。

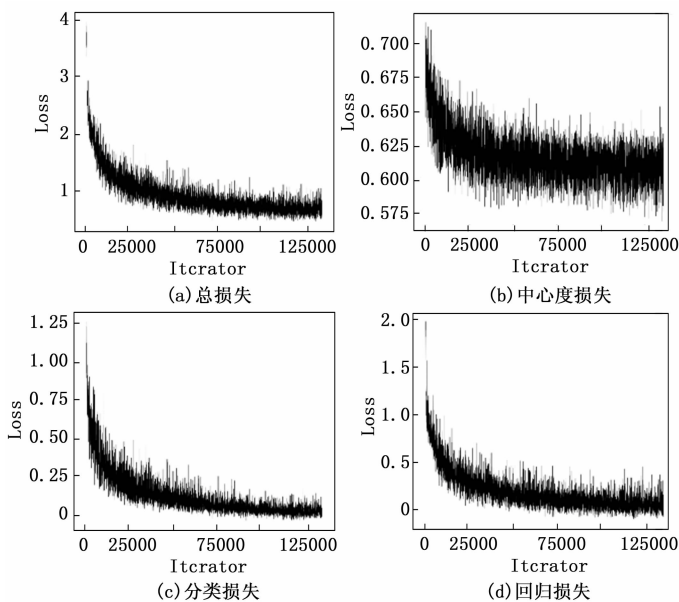


图 5 损失变化情况

3.3 PARF-FCOS 和一些典型算法的对比

为了验证本文算法在遥感影像目标检测任务上的有效性与可靠性，将其与当前一些典型检测算法^[21-22]在 DIOR

数据集上进行精度对比。实验结果如表 3 所示。可以看出，各网络模型在 DIOR 数据集上均取得了一定的检测效果。从单类别的 AP 值来看，PARF-FCOS 在篮球场、桥梁、烟囱等 8 个类别上取得了最高检测精度，与基准网络 FCOS 相比，PARF-FCOS 对火车站、高尔夫球场和机场的提升效果明显，分别提升了 9.5%、8.4% 和 7.1%，同时对于汽车、舰船和飞机这些小尺度目标也有提升，分别提升了 6.3%、5.4% 和 5.3%；从整体检测精度 mAP 来看，PARF-FCOS 领先于其它对比检测模型，其中，与 Faster R-CNN 相比提升了 16.3%，与采用了特征金字塔网络的 Faster R-CNN 相比提升了 7.3%，与 YOLOv3 相比提升了 13.3%，与 CornerNet 相比提升了 5.5%，与 FCOS 相比提升了 4.3%，同时本文网络也超过了高性能的 YOLOv5 检测模型。充分地证明了 PARF-FCOS 目标检测模型的有效性和可靠性。

表 3 不同算法的检测结果

目标类别	Faster R-CNN	Faster R-CNN with FPN	YOLO Ov3	YOLO Ov5	Corner Net	FCOS	PARF-FCOS
飞机	53.6	54.1	72.2	83.8	58.8	64.1	69.4
机场	49.3	71.4	29.2	75.5	84.2	75.4	82.5
棒球场	78.8	63.3	74.0	78.7	72.0	72.8	74.5
篮球场	66.2	81.0	78.6	84.6	80.8	86.7	88.4
桥梁	28.0	42.6	31.2	45.1	46.4	44.8	46.8
烟囱	70.9	72.5	69.7	76.2	75.3	74.8	76.5
大坝	62.3	57.5	26.9	53.7	64.3	52.4	53.5
高速服务区	69.0	68.7	48.6	62.0	81.6	73.0	77.2
高速收费站	55.2	62.1	54.4	60.3	76.3	63.4	69.3
高尔夫球场	68.0	73.1	31.1	73.1	79.5	76.4	84.8
田径场	56.9	76.5	61.1	70.2	79.5	74.7	80.5
港口	50.2	42.8	44.9	58.4	26.1	47.6	52.8
立交	50.1	56.0	49.7	56.8	60.6	55.1	58.7
舰船	27.7	71.8	87.4	89.9	37.6	71.3	76.7
体育场	73.0	57.0	70.6	70.9	70.7	66.3	72.0
储罐	39.8	53.5	68.7	77.2	45.2	58.6	60.5
网球场	75.2	81.2	87.3	86.5	84.0	84.0	85.9
火车站	38.6	53.0	29.4	49.0	57.1	54.2	63.7
汽车	23.6	43.1	48.3	54.7	43.0	45.0	51.3
风力机	45.4	80.9	78.7	77.4	75.9	81.8	83.3
mAP/%	54.1	63.1	57.1	69.2	64.9	66.1	70.4

3.4 各模块有效性验证

为了分析 PARF-FCOS 中各个改进模块对于模型的影响，本文在 FCOS 模型的基础上依次对 PA-Net、RFB 和 Diou loss 的有效性进行了实验验证。经过实验发现，以上方法均能提升网络性能，实验结果如表 4 所示。从检测精度上看，原始 FCOS 算法在 DIOR 数据集上取得了 66.1% 的 mAP 值，而采用 PA-Net 提取特征后，取得了 69.3% 的 mAP 值，提升了 3.2%，这说明添加 PA 模块有效的提升

了网络的特征提取能力, 这是因为在 PA 模块的作用下, 特征中的背景噪音等无用信息得到抑制, 更有利于网络提取目标的有用信息, 提高目标信息的利用率。在加入 RFB 后, 网络取得了 69.9% 的 mAP 值, 提升了 0.6%, 说明 RFB 能够增强浅层特征的目标响应, 其原因是 RFB 的多分支结构具有更大的感受野, 能够提取更加全面和丰富的上下文信息, 弥补了浅层特征在感受野上的不足。在引入 DIoU loss 训练网络后, mAP 值提升 0.5% 达到了 70.4%, 验证了 DIoU loss 能使边界框的回归过程更加准确, 有效地提升了边界框坐标的预测精度。从检测速度上看, PA 模块和 RFB 在一定程度上增加了模型的计算量, 检测速度分别下降了 4.6FPS 和 0.9FPS, 而引入 DIoU loss 训练并不会影响检测速度, 最终网络的检测速度达到 23.2FPS, 仍然能满足实时检测的要求。

表 4 各模块有效性验证结果

实验序号	PA-Net	RFB	DIoU loss	mAP / %	FPS / (frame · s ⁻¹)
1	×	×	×	66.1	28.7
2	√	×	×	69.3	24.1
3	√	√	×	69.9	23.2
4	√	√	√	70.4	23.2

3.5 注意力网络对比

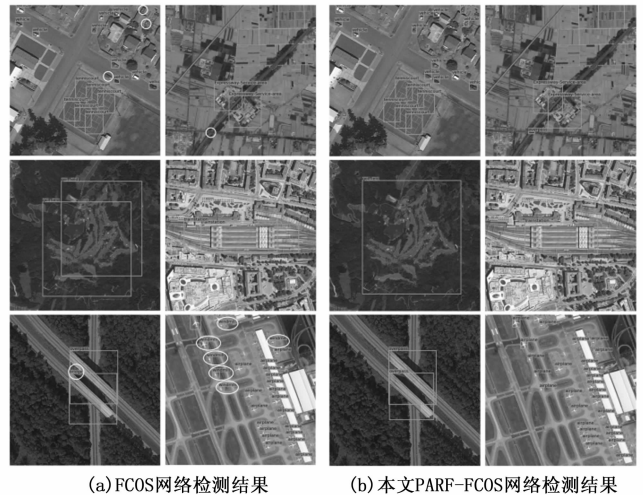
为了进一步验证 PA-Net 的优势, 将其与经典注意力网络 SE-Net 和 CBAM-Net^[23] 进行了对比, 其中 SE-Net 和 CBAM-Net 表示将 PA-Net 中的 PA 模块替换成相应的注意力模块, 其余网络参数均保持一致, 实验结果如表 5 所示。原始网络取得了 66.1% 的 mAP 值和 28.7 FPS, 采用 SE-Net、CBAM-Net 和 PA-Net 后分别取得了 67.9%、68.4% 和 69.3% 的 mAP 值, 分别提升了 1.8%、2.3% 和 3.2%, 相应的检测速度分别为 27.6 FPS、21.7 FPS 和 24.1 FPS。与 SE-Net 相比, 采用 PA-Net 的 mAP 提高了 1.4%, 检测速度下降了 3.5 FPS, 说明空间位置信息有利于目标检测精度的提升, PA 模块通过双注意力分支结构有效的利用了空间位置信息, 增强了网络的性能, 但是 PA 模块的计算量较 SE 模块有所增加, 导致检测速度有所降低。与 CBAM-Net 相比, 采用 PA-Net 后的 mAP 提高了 0.9%, 检测速度提高了 2.4 FPS, 可以看出, 相较于 CBAM 使用卷积获得的局部空间位置关系, PA 模块通过一维平均池化能够捕获到更具代表性的全局位置信息, 更有利于网络性能的提升, 并且 PA 模块的计算量更少, 检测速度更快。

表 5 注意力网络对比结果

实验序号	骨架网络	mAP / %	FPS / (frame · s ⁻¹)
1	ResNet50	66.1	28.7
2	SE-Net	67.9	27.6
3	CBAM-Net	68.4	21.7
4	PA-Net	69.3	24.1

3.6 检测效果可视化

为了直观地展示本文算法的检测效果, 本文对多类场景下的多种目标的检测结果进行了可视化, 如图 6 所示, 其中 6 (a) 为 FCOS 的可视化结果, 6 (b) 为改进模型 PARF-FCOS 的可视化结果。从图中可以看出, PARF-FCOS 的检测结果较 FCOS 的检测结果更为准确, 有效地改善了漏检、错检等现象, 如第 1 行对比图所示, FCOS 漏检了影像中的小目标“汽车”和“立交桥”等目标(圆圈标出), 而 PARF-FCOS 能有效的检测出来; 在第 2 行对比图中, FCOS 产生了低质量的冗余框, 而 PARF-FCOS 的边框更加准确; 在第 3 行结果图中, FCOS 将影像中的“立交桥”识别为“桥梁”, 将“飞机”识别为“风力机”(圆圈标出), 而 PARF-FCOS 均能正确的识别出这些目标的类别。可以看出, 本文方法的通用性更强, 对于不同的场景下的多种结构、尺度差异较大的目标依然能取得优秀的检测效果。



(a) FCOS 网络检测结果

(b) 本文 PARF-FCOS 网络检测结果

图 6 检测结果对比图

4 结束语

针对当前遥感影像背景复杂、小目标等情况导致的检测效果不佳的问题, 提出了一种基于 FCOS 改进的目标检测模型 PARF-FCOS。通过构造位置注意力模块, 差异化的调整特征图各像元的通道权值, 突出目标特征, 抑制背景特征; 使用 RFB 处理浅层特征图以增大感受野, 获取目标上下文信息, 提升小尺度目标的检测精度; 训练时采用 DIoU loss 作为边框回归损失函数, 使边框回归过程更加平稳和准确。通过 DIOR 数据集上的实验分析表明, 本文算法有效的提高了目标的检测精度, 在多类目标上取得了较高的检测精度, 算法对于小目标检测精度也有明显的提升, 验证了本文算法的有效性和鲁棒性。本文提出的网络模型虽然提高了检测精度, 但 PA 模块和 RFB 给网络增加了时间代价。今后, 将从网络推理速度出发, 尝试利用卷积通道裁剪、网络剪枝等轻量化技术进一步优化模型。

参考文献:

[1] CHENG G, HAN J. A Survey on Object Detection in Optical Remote Sensing Images [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11 - 28.

[2] ZHANG J, XIE C, XU X, et al. A contextual bidirectional enhancement method for remote sensing image object detection [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 4518 - 4531.

[3] 姚 远, 姜志国, 张浩鹏. 基于层次化分类器的遥感图像飞机目标检测 [J]. 航天返回与遥感, 2014, 35 (5): 88 - 94.

[4] 黎经元, 厉小润, 赵辽英. 基于边缘线分析与聚合通道特征的港口舰船检测 [J]. 光学学报, 2019, 39 (8): 217 - 226.

[5] LECUN Y, BOTTOU L. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278 - 2324.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2014: 580 - 587.

[7] REDMON J, FARHADI A. YOLOv3: an incremental improvement [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018, 87 (8): 101 - 104.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Proceedings of the European Conference on computer Vision, 2016: 21 - 37.

[9] 欧 攀, 张 正, 路 奎, 等. 基于卷积神经网络的遥感图像目标检测 [J]. 激光与光电子学进展, 2019, 56 (5): 74 - 80.

[10] 周雪珂, 刘 畅, 周 滨. 多尺度特征融合与特征通道关系校准的 SAR 图像船舶检测 [J]. 雷达学报, 2021, 10 (4): 531 - 543.

[11] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型 [J]. 激光与光电子学进展, 2021, 58 (2): 363 - 371.

[12] 张 暄, 高跃清. 基于卷积神经网络的雷达目标检测方法 [J]. 计算机测量与控制, 2021, 29 (2): 49 - 52.

[13] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [M] // FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 765 - 781.

[14] 刘 宏, 王天正, 晋 涛, 等. 基于气象灾害监测雷达基数据质量控制方法探讨 [J]. 科技通报, 2019, 35 (8): 1 - 5, 12.

[15] 刘丽丽, 李英华, 王雪莲, 等. 基于观测系统模拟试验的海表气象观测站点布局方案研究 [J]. 气候与环境研究, 2019, 24 (6): 711 - 722.

[16] 官元红, 任 杰, 鲍艳松, 等. 基于一维变分算法的红外高光谱 (IASI) 卫星遥感大气温湿廓线研究 [J]. 大气科学学报, 2019, 42 (4): 602 - 611.

[17] 宋广南, 卢海梁, 李 浩, 等. 复杂天气及海风对地基被动干涉微波辐射无源探测系统性能的影响 [J]. 电子与信息学报,

[14] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully Convolutional One-Stage Object Detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627 - 9636.

[15] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2016: 770 - 778.

[16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2018: 7132 - 7141.

[17] LIU S, HUANG D. Receptive field block net for accurate and fast object detection [C] // Proceedings of the European Conference on Computer Vision. 2018: 385 - 400.

[18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1 - 9.

[19] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C] // Proceedings of AAAI Conference on Artificial Intelligence, 2020, 34 (7): 12993 - 13000.

[20] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296 - 307.

[21] 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.

[22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2017: 936 - 944.

[23] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [M] // FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3 - 19.

[18] 李 芬, 杨 勇, 胡 超, 等. 基于天气类型识别的跟踪式光伏发电系统能量评估及优化 [J]. 太阳能学报, 2020, 41 (11): 112 - 119.

[19] 刘文雯, 段菁春, 胡京南, 等. 基于环境监测数据的大气重污染应急减排措施效果评估 [J]. 环境科学研究, 2019, 32 (5): 734 - 741.

[20] 闫文辉, 黄兴友, 李盈盈, 等. 基于多普勒天气雷达的低空多普勒速度的切变识别算法研究 [J]. 热带气象学报, 2019, 35 (2): 253 - 261.

~~~~~  
 (上接第 5 页)