

DOI: 10.3785/j.issn.1008-973X.2022.11.012

联合多尺度与注意力机制的遥感图像目标检测

张云佐^{1,2}, 郭威¹, 蔡昭权³, 李文博¹

(1. 石家庄铁道大学 信息科学与技术学院, 河北 石家庄 050043; 2. 河北省电磁环境效应与信息处理重点实验室, 河北 石家庄 050043; 3. 汕尾职业技术学院, 广东 汕尾 516600)

摘 要: 遥感图像存在背景复杂、目标尺度差异大且密集分布等不足, 为提高现有算法的检测效果提出联合多尺度与注意力机制的遥感图像目标检测算法. 改进空洞空间金字塔池化模块, 增大不同尺寸图像的感受野; 提出注意力模块用于学习特征图通道信息和空间位置信息, 提升算法对复杂背景下遥感图像目标区域的特征提取能力; 引入加权双向特征金字塔网络结构与主干网结合来增进多层次特征的融合; 使用基于距离的非极大值抑制方法进行后处理, 改善检测框易重叠的问题. 在 DIOR 和 NWPUVHR-10 数据集上的实验结果表明: 所提算法的平均精度均值 mAP 分别达到 71.6% 和 91.6%, 相比于主流的 YOLOv5s 算法分别提升了 2.9% 和 1.5%. 所提算法对复杂遥感图像取得了更好的检测效果.

关键词: 遥感图像; 目标检测; YOLOv5s 算法; 多尺度特征; 注意力模块; 特征融合; 非极大值抑制
中图分类号: TP 751.1 **文献标志码:** A **文章编号:** 1008-973X(2022)11-2215-09

Remote sensing image target detection combining multi-scale and attention mechanism

ZHANG Yun-zuo^{1,2}, GUO Wei¹, CAI Zhao-quan³, LI Wen-bo¹

(1. School of Information Science and Technology, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; 2. Hebei Key Laboratory of Electromagnetic Environmental Effects and Information Processing, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; 3. Shanwei Institute of Technology, Shanwei 516600, China)

Abstract: Remote sensing images have deficiencies such as complex backgrounds, significant differences in target scales, and dense distribution, resulting in poor detection of existing algorithms. A remote sensing image object detection algorithm that combined multi-scale and attention mechanisms was proposed. The receptive field of images of different sizes improved the atrous spatial pyramid pooling module. An attention module was proposed to improve the feature extraction ability for target regions of remote sensing images under complex backgrounds by learning the feature map channel information and the spatial location information. A weighted bidirectional feature pyramid network structure was introduced to combine with the backbone network to improve the fusion of multi-level features. A distance-based non-maximum suppression method was used for postprocessing, which improved the problem of easy overlapping of detection frames. Experimental results on DIOR and NWPU VHR-10 datasets showed that the mean average precision (mAP) of the proposed algorithm reached 71.6% and 91.6%, which were 2.9% and 1.5% higher than those of the mainstream YOLOv5s algorithm respectively. The algorithm achieved good detection results for complex remote sensing images.

Key words: remote sensing image; target detection; YOLOv5s algorithm; multi-scale feature; attention module; feature fusion; non-maximum suppression

收稿日期: 2021-11-30. 网址: www.zjujournals.com/eng/article/2022/1008-973X/202211012.shtml

基金项目: 广东省重点领域研发计划资助项目 (2019B010137002); 国家自然科学基金资助项目 (61702347, 62027801); 河北省自然科学基金资助项目 (F2022210007, F2017210161); 河北省高等学校科学技术研究项目 (ZD2022100, QN2017132); 中央引导地方科技发展资金资助项目 (226Z0501G).

作者简介: 张云佐 (1984—), 男, 副教授, 博导, 从事图像处理、视频智能分析和大数据处理研究. orcid.org/0000-0001-7499-4835.
E-mail: zhangyunzuo888@sina.com

遥感图像目标检测在城市规划、资源调查和灾害监测等领域应用广泛,基于遥感图像的目标检测具有重要研究意义^[1]。然而遥感图像是在几百米到几千米范围内,通过高空俯视拍摄得到的,同一目标在不同图像中尺度大小不一并且差别较大。遥感图像覆盖范围广,包含多种背景信息,容易对检测目标产生干扰。传统目标检测方法大多是基于手工特征^[2]的构建,需要人工设计对应检测对象的特征提取算子,训练特定分类器,进而对提取的目标特征进行分类,实现对特定目标的检测。此方法的缺陷在于提取的特征信息大多是浅层,表达的信息量较少,泛化性较弱,并且对于背景复杂多变的遥感图像检测效果欠佳。近年来,随着深度学习理论的快速发展^[3],基于卷积神经网络的遥感图像目标检测方法迭出不穷。

Yang 等^[4]提出一种基于多任务旋转区域的卷积神经网络检测模型,通过构建稠密特征金字塔,利用稠密连接生成信息丰富的特征,用于提升稠密船舰的检测精度。Feng 等^[5]通过在区域卷积神经网络 R-CNN 算法中引入运动信息,并在损失函数中加入平滑因子,实现遥感图像小目标的有效检测。Guan 等^[6]设计一种胶囊特征金字塔网络,提取并集成多尺度的胶囊特征,提高了光学遥感图像中道路的检测精度。Courtrai 等^[7]改进超分辨率框架并将生成对抗网络 (generative adversarial networks, GAN) 集成到一个循环模型,提高超分辨率体系结构的质量用于遥感图像小目标的检测性能。上述方法均提升遥感图像中密集小目标的检测精度,但是这些方法只对遥感图像中的单类目标进行检测,缺少多类别目标检测。Zhang 等^[8]提出一种基于特征金字塔网络的遥感目标检测方法,结合强语义与弱语义特征,用于提高小而密集目标的检测效果。Li 等^[9]设计了 CPN 网络 (category proposal network) 和 F-RPN 网络 (fine-region proposal network),将生成的候选区域和目标数相结合,得到图像自适应候选框,进而实现精准的目标定位与检测。然而前 2 种模型较大,训练复杂度较高。Chen 等^[10]提出结合场景和上下文信息的遥感图像检测方法,用于加强目标与场景之间的关系,进而解决由目标尺寸变化引起的问题。He 等^[11]开发遥感图像目标检测模型 TF-YOLO,提出基于深度卷积神经网络的多尺度目标检测器,可以实现遥感图像目标的高效检测。Shamsolmoali 等^[12]利用多块特征金字塔网络检测

多尺度目标,将模块划分为多个关联子集,用来提高光学遥感图像目标的检测性能。这些方法对多尺度目标实现高效检测,但是没有考虑复杂背景对遥感图像目标检测的影响。

为实现遥感图像目标的精准检测,本研究提出联合多尺度与注意力机制的检测算法。首先,改进空洞空间金字塔池化模块,增加不同尺寸图像的感受野,以更好地捕获上下文信息;其次,在 backbone 中加入注意力模块,增强网络对遥感图像目标区域特征信息的关注,提升网络对目标和复杂背景的辨别能力;然后,引入加权双向特征金字塔网络结构,加强多尺度特征融合,丰富特征语义信息,进而提升算法检测能力;最后,采用基于距离的非极大值抑制方法来优化边界框的回归。

1 算法设计

为了保证遥感图像目标检测速度,采用体积最小的 YOLOv5s 网络模型为基础架构,并在此基础上进行改进,网络结构框图如图 1 所示。网络结构包括 Backbone 主干网络、Neck 瓶颈层及 Head 检测层 3 部分。Backbone 主干网络主要用于特征提取,改进空洞空间金字塔池化模块,在不增加计算量的前提下,捕获图像多尺度特征,并在 2 次卷积后嵌入提出的注意力模块,以抑制遥感图像中复杂背景的干扰。Neck 瓶颈层主要用于特征融合,引入加权双向特征金字塔网络结构,以融合信息更丰富的多尺度特征。Head 检测层主要用于预测图像特征,生成边界框并预测目标类别。其中,图 1 中 m 为输出特征图通道数,提出的算法采用基于距离的非极大值抑制方法剔除重复冗余的预测框。

1.1 空洞空间金字塔池化

YOLOv5s 中使用 SPP 模块提取不同感受野的信息,但是此模块不能充分整合局部信息与全局信息,容易造成信息的丢失。在借鉴 Deeplab^[13-15]下的空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 提出 ASPP+模块,ASPP+中主要引入空洞卷积,其核心是通过设置不同的膨胀系数,获得不同大小的感受野,进而提取多尺度上下文信息,提高网络模型识别不同尺度的目标能力。空洞卷积核和感受野大小的计算公式为

$$f_n = f_k + (f_k - 1) * (d - 1), \quad (1)$$

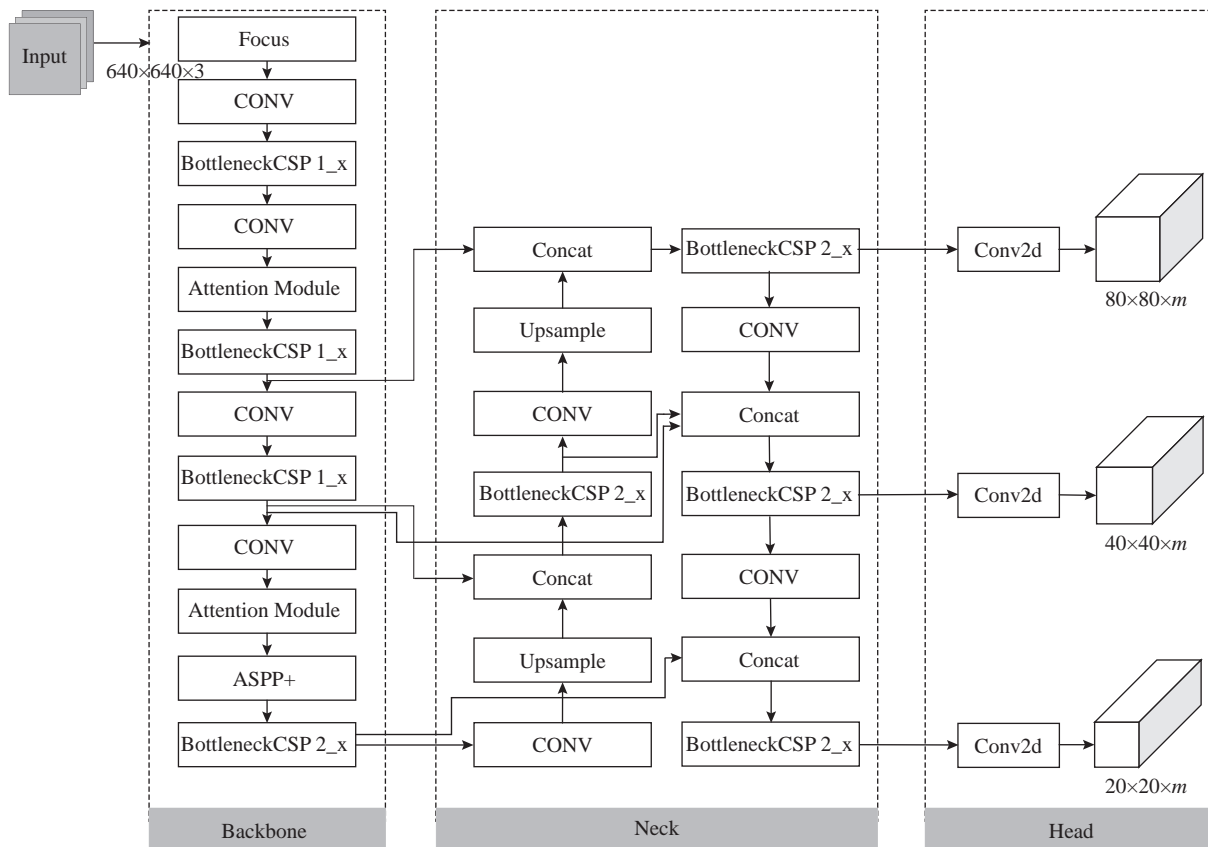


图 1 联合多尺度与注意力机制算法的网络结构框图

Fig.1 Network structure block diagram of joint multiscale and attention mechanism algorithm

$$R_m = R_{m-1} + (f_n - 1) * \prod_{i=1}^{m-1} s_i. \quad (2)$$

式中: f_n 为空洞卷积核大小, f_k 为原始卷积核大小, d 为卷积核的膨胀系数, R_{m-1} 为第 $m-1$ 层感受野大小, R_m 为经空洞卷积后第 m 层感受野大小, s_i 为第 i 层的步长。

ASPP+模块如图 2 所示,此模块分为 5 个阶段 aspp1~aspp5. 首先,删除 ASPP 中的 Dropout 层形成本模块中的 aspp5, Dropout 技术主要是在深度学习网络训练中随机丢弃一些节点,使得网络权重在更新过程中不依赖隐藏节点之间的固定关系,从而起到正则化效果,但是在遥感图像目标检测任务中,检测效果与类别信息和位置信息相关,因此随机丢弃节点信息并不适用于此类任务. 然后,采用 LeakyReLU 激活函数替换原始模块中的 ASPP ReLU 激活函数,当输入 ReLU 激活函数的值为零或负值时,函数的梯度变为零,此时网络无法执行反向传播,权重无法更新,LeakyReLU 激活函数解决了这一问题. 最后,将 aspp1~aspp5 这 5 个阶段操作的特征合并到模型中,并拼接在一起,通过一个 1×1 的卷积操作调整通道. 改进后的 ASPP+模块增大遥感图像的感受

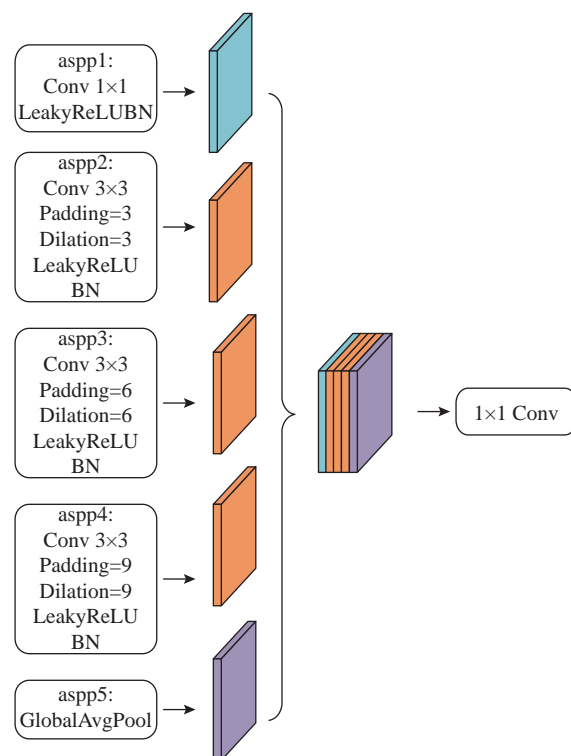


图 2 ASPP+模块

Fig.2 ASPP+ module

野,提升网络对于遥感图像目标尺度变化大的鲁棒性,并且提取到多尺度深度特征。

1.2 注意力机制

为了使网络更加关注重要信息,受压缩和激励网络(squeeze-and-excitation net, SENet)^[16]和卷积注意力模块(convolutional block attention module, CBAM)^[17]的启发,提出一种注意力模块,该模块包括通道注意力和空间注意力2个子模块,如图3所示.首先使用通道注意力模块重新校准每个通道的权重,使得网络关注重要特征,抑制不重要特征.然后使用空间注意力模块突出目标区域的空间位置信息,引导网络专注于遥感图像目标区域,抑制无关背景的干扰.在图3中,输入通道注意力模块的特征图为 F ,大小为 $C \times H \times W$, C 、 H 和 W 分别为特征图的通道数、长和宽.首先对输入特征图 F 进行平均池化(avg pooling)和最大池化(max pooling)的操作,有利于筛选辨识度高的特征^[18];然后经过2个全连接层 FC(fully connected layers)先降低特征图维度再升高维度,拟合通道之间的相关性;最后使用 Sigmoid 函数进行归一化处理,重新分配各个通道的特征权重,有利于学习目标区域对应的通道信息,进而使网络对遥感图像目标特征进行充分学习.空间注意力

模块增强对特征图的空间信息学习,以突出特征图中目标的相关区域.在第1层空间注意力模块中,为获取更多的语义信息,首先对特征图 F' 进行平均池化操作,提取特征图的空间信息,然后使用 7×7 的卷积层连接,并通过 Sigmoid 函数增加非线性特征,得到特征图 M_{s1} ,大小为 $1 \times H \times W$,计算公式为

$$M_{s1} = \sigma \{f_{\text{conv}}^{7 \times 7}[\text{avgpool}(F')]\}. \quad (3)$$

式中: σ 为 Sigmoid 激活函数, $f_{\text{conv}}^{7 \times 7}$ 为 7×7 卷积层,avgpool 为平均池化操作.

在第2层空间注意力模块中,为有效地利用上下文信息,首先使用感受野大小为 7×7 和 3×3 的空洞卷积增大感受野,融合上下文信息,然后使用 1×1 卷积层降维,得到特征图 M_{s2} ,大小为 $1 \times H \times W$,计算公式为

$$M_{s2} = \sigma \{f_{\text{conv}}^{1 \times 1} \{f_{\text{conv}}^{3 \times 3} [f_{\text{conv}}^{7 \times 7}(F')]\}. \quad (4)$$

式中: $f_{\text{conv}}^{1 \times 1}$ 、 $f_{\text{conv}}^{3 \times 3}$ 、 $f_{\text{conv}}^{7 \times 7}$ 分别为感受野大小为 1×1 、 3×3 、 7×7 的卷积层.最终特征图 F'' 的计算公式为

$$F'' = M_s \otimes F'. \quad (5)$$

式中:“ \otimes ”为特征图逐元素相乘, M_s 为 M_{s1} 与 M_{s2} 的和.

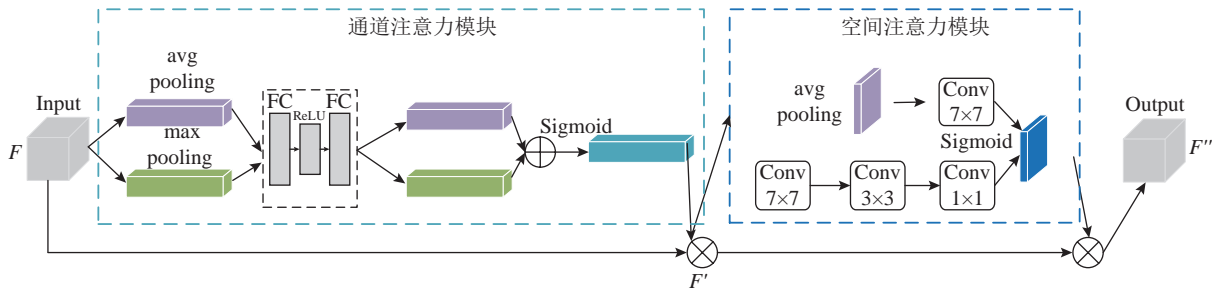


图3 注意力模块

Fig.3 Attention module

1.3 多尺度特征融合

随着网络层次不断加深,深层网络特征中包含较多的语义信息,但特征图的分辨率较低,对细节的感知能力较差^[19].浅层网络特征图的分辨率较高,包含较多的位置细节信息,但经过的卷积操作少,使得特征中包含的噪声更多,因此需要融合不同层级的特征以丰富特征语义信息.在特征融合阶段引入加权双向特征金字塔网络(bi-directional feature pyramid network, BiFPN),BiFPN^[20]是2020年由Google提出的一种双向特征融合网络,其结构如图4所示.BiFPN的功能是融合具有丰富位置信息的低层特征图和具有丰富语义信息的高层特征图,增强高层特征图位置信息的表达能力,以提高遥感图像目标检测的精度.

BiFPN在同一层的输入节点和输出节点间添加横向连接(如图4虚线所示),还用自上向下和自下向上的融合方法替换FPN仅进行自上向下融合的方法(如图4空心箭头所示).只移除一个输入的节点,原因在于只有一个输入的节点不包含特征融合,从而可以减少计算量,提高计算效率,具体公式为

$$P_4^{\text{td}} = \text{Conv} \left(\frac{w_1 \cdot P_4^{\text{in}} + w_2 \cdot \text{Resize}(P_5^{\text{in}})}{w_1 + w_2 + \varepsilon} \right), \quad (6)$$

$$P_4^{\text{out}} = \text{Conv} \left(\frac{w'_1 \cdot P_4^{\text{in}} + w'_2 \cdot P_4^{\text{td}} + w'_3 \cdot \text{Resize}(P_3^{\text{out}})}{w'_1 + w'_2 + w'_3 + \varepsilon} \right). \quad (7)$$

式中: P_4^{td} 为自上向下路径的中间特征; P_4^{out} 为自下

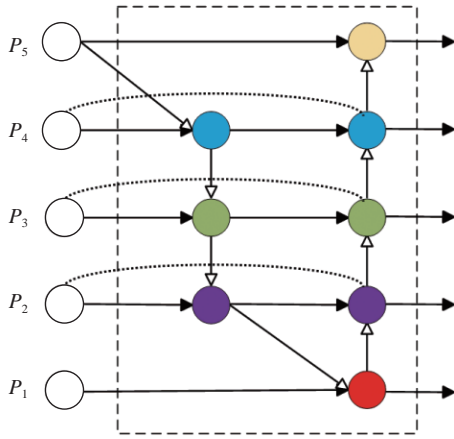


图 4 BiFPN 结构图

Fig.4 BiFPN structure diagram

向上路径的输出特征; Conv 为卷积操作; Resize 为上采样或下采样操作, 用来对齐分辨率; $\varepsilon=0.0001$, 采用一个很小的值, 用来避免数值的不稳定; “+”为特征在通道上的堆叠. BiFPN 有效地融合不同尺度的特征并增加同一尺度特征的信息融合, 对遥感图像目标检测效果有一定的提升作用.

1.4 非极大值抑制

在传统非极大值抑制 (non-maximum suppression, NMS) 中, 用交并比 (intersection over union, IoU) 抑制检测任务中的重复边界框为

$$\text{IoU} = \frac{|B \cap B^{\text{gt}}|}{|B \cup B^{\text{gt}}|}. \quad (8)$$

$$\text{NMS} = \begin{cases} s_i, & \text{IoU}(B^{\text{gt}}, B) < \varepsilon; \\ 0, & \text{IoU}(B^{\text{gt}}, B) \geq \varepsilon. \end{cases} \quad (9)$$

式中: B 为预测框, B^{gt} 为真实框, IoU 为预测框与真实框之间的重叠度, s_i 为分类得分, ε 为 NMS 的阈值. IoU 容易丢弃目标重叠率较高的边界框, 不利于密集分布的遥感目标检测. 针对此问题, 采用中心点距离非极大值抑制方法 (DIoU-NMS) 对检测结果进行后处理. DIoU-NMS 不仅考虑 IoU 的值, 而且考虑 2 个检测框中心点之间的距离. DIoU-NMS 定义为

$$s_i = \begin{cases} s_i, & \text{IoU} - R_{\text{DIoU}}(B^{\text{gt}}, B) < \varepsilon; \\ 0, & \text{IoU} - R_{\text{DIoU}}(B^{\text{gt}}, B) \geq \varepsilon. \end{cases} \quad (10)$$

$$R_{\text{DIoU}} = \rho^2(b, b^{\text{gt}})/c^2. \quad (11)$$

式中: b 和 b^{gt} 分别为锚框和真实框各自的中心点, ρ 为中心点之间的欧氏距离, c 为同时覆盖 b 、 b^{gt} 的最小矩形的对角线距离. DIoU-NMS 不仅能有效地对边界框进行抑制, 而且能解决密集目标中相邻目标可能被判断为同一目标的问题.

2 实验与分析

2.1 实验条件

实验的运行环境为: 操作系统 CentOS7, 显卡 NVIDIA Tesla V100S-PCIE-32GB, 处理器 Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz, 采用 Pytorch 深度学习框架. 在实验过程中, 所有模型均采用随机梯度下降法训练网络, 权重衰减系数为 0.0001, 动量因子为 0.9, 批量大小为 16.

2.2 DIOR 数据集实验结果

DIOR 数据集^[21]是一种大规模的光学遥感目标检测数据集, 由 23 463 幅像素为 800×800 的遥感图像组成, 总共分为 20 个类别. 整个数据集被分成训练集 (5 862 幅图像)、验证集 (5 863 幅图像) 和测试集 (11 738 幅图像), 数据集部分样例图片如图 5 所示.

分析实验数据集获得图 6 所示的可视化结果. 在图 6(a) 中, x 、 y 是确定目标中心点的位置, 图片中颜色越深代表目标框的中心点越集中于该点. 图 6(b) 中的 w 、 h 坐标分别为图片中目标的宽度和高度. 可以看出数据集中目标分布较均匀, 而且中小目标所占比例较大. 为了对所提算法进行充分评估, 验证其对遥感图像目标的检测能力, 本研究对常用目标检测算法进行实验对比分析, 并选取 mAP (平均精度均值) 作为 5 种不同检测算法的评估指标. 表 1 为 DIOR 数据集信息, C1~C20 为数据集的各个类别, 各算法模型在测试集上的 mAP 对比情况如表 2 所示, 加黑字体代表该类别的最优效果, 其中本研究算法用 ours 表示.



图 5 DIOR 数据集样例

Fig.5 Sample DIOR data set

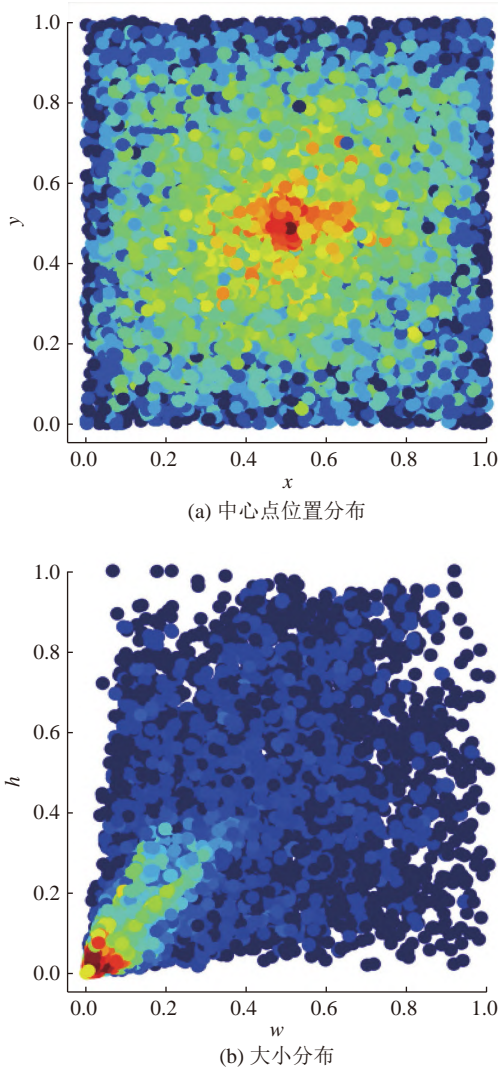


图 6 DIOR 数据集分析

Fig.6 DIOR data set analysis

所提检测算法在 DIOR 数据集上的 mAP 值和其他 4 种算法相比最高, 达到 71.6%. 实验表明, 所提算法有效提升了遥感图像目标检测的平均精度.

表 2 列出所提算法模型与其他 4 种算法在 DIOR 数据集上各类目标的 AP 值对比结果. 与 YOLOv5s 算法相比, airplane 类别的 AP 由 78.3% 提升至 85.8%, 提高了 7.5%, 提升效果较为明显;

表 2 不同算法模型在 DIOR 测试集上的对比

Tab.2 Comparison of different algorithm models on DIOR test set

算法模型	mAP	AP									
		C1/C11	C2/C12	C3/C13	C4/C14	C5/C15	C6/C16	C7/C17	C8/C18	C9/C19	C10/C20
RetinaNet ^[22]	65.7	53.7/74.2	77.3 /50.7	69.0/ 59.6	81.3/71.2	44.1/69.3	72.3/44.8	62.5/81.3	76.2 /54.2	66.0/45.1	77.7/83.4
PANet ^[23]	66.1	60.2/73.4	72.0/45.3	70.6/56.9	80.5/71.7	43.6/70.4	72.3/62.0	61.4/80.9	72.1/57.0	66.7 /47.2	72.0/ 84.5
CBD-E ^[24]	67.8	54.2/ 79.5	77.0/47.5	71.5/59.3	87.1/69.1	44.6/69.7	75.4/64.3	63.5 /84.5	76.2/ 59.4	65.3/44.7	79.3 /83.1
YOLOv5s	68.7	78.3/73.1	65/58.3	74.3/57.4	90.6 / 91.8	44.3/67.9	80.1 / 82.7	48.9/89.1	57.7/49.7	63.2/55.4	68.6/78.1
Ours	71.6	85.8 /75.7	74.2/ 59.9	78.9 /58.6	89.8/89.7	46.1 / 71.9	77.8/78.7	60.5/ 89.5	65.1/55.4	65.3/ 56.4	75.6/78.1

表 1 DIOR 数据集类别信息

Tab.1 DIOR data set category information

类别	名称	类别	名称
C1	airplane	C11	ground track field
C2	airport	C12	harbor
C3	baseball field	C13	overpass
C4	basketball court	C14	ship
C5	bridge	C15	stadium
C6	chimney	C16	storage tank
C7	dam	C17	tennis court
C8	expressway service area	C18	train station
C9	expressway toll station	C19	vehicle
C10	golf field	C20	wind mill

vehicle 类别的 AP 由 55.4% 提升至 56.4%, 提高了 1.0%. 需要说明的是, airplane 和 vehicle 这 2 个类别在遥感图像目标检测场景中具有尺度差异大且密集分布的特点. 在 dam 类别的检测中, AP 由 48.9% 提升至 60.5%, 提高了 11.6%, 提升效果较为明显; baseball field 类别的 AP 由 74.3% 提升至 78.9%, 提高了 4.6%. dam 和 baseball field 这 2 个类别在遥感图像目标检测场景中具有背景复杂的特点. 此外, 和 RetinaNet、PANet 和 CBD-E 算法模型相比, 所提算法在 7 类目标中取得了最好效果, 而这 7 类目标具有遥感图像背景复杂、目标尺度差异大或密集分布的特点. 因此, 所提算法在复杂的检测场景中更适用.

如图 7 所示, 给出数据集上的部分检测实例, 并将算法改进前后的检测效果进行对比. 图 19560.jpg 中包含 1 个 basketballcourt、1 个 baseball-field 和 1 个 vehicle, 而 YOLOv5s 算法检测到 2 个 vehicle; 图 22075.jpg 中包含 6 个 vehicle, 而 YOLOv5s 算法只检测到 5 个 vehicle; 图 09765.jpg 中只包含

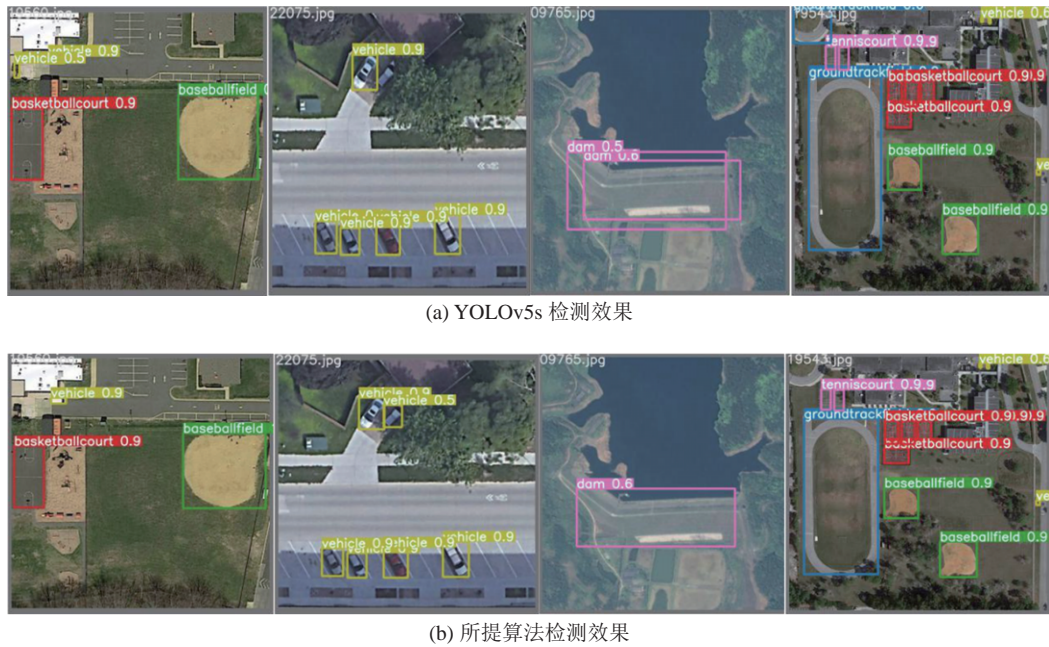


图 7 YOLOv5s 算法与所提算法检测效果对比

Fig.7 Comparison of detection effect between YOLOv5s algorithm and proposed algorithm

1 个 dam, 而 YOLOv5s 算法检测到 2 个 dam; 图 19543.jpg 中包含 4 个 basketballcourt、2 个 baseballfield、2 个 tenniscourt、4 个 vehicle 和 1 个 groundtrackfield, 而 YOLOv5s 算法漏检了 1 个 vehicle 且误检了 1 个 groundtrackfield, 所提算法漏检了 1 个 vehicle. 由此可见, 和 YOLOv5s 检测算法相比, 所提算法有效的减少了漏检和误检的现象.

2.3 NWPU VHR-10 数据集实验结果

NWPU VHR-10 数据集共有 650 幅图像, 包含 10 种类型的对象, 并使用 2 个不同图像数的数据集验证模型在不同数据集上的性能. 在实验过程中, 随机选取 90% 的图像作为训练集, 剩余 10% 的图像作为测试集. 不同算法模型在 NWPU VHR-10 测试集上的 mAP 对比情况如表 3 所示, 加黑字体代表该类别的最优效果. 由表 3 可知, 所提算法在 NWPU VHR-10 数据集上的 mAP 达到 91.6%, 相对其他 4 种算法取得了最优的检测效果. 此外, airplane 和 ship 类别中的目标具有尺度差异大、密

集分布的特点, tennis court 和 bridge 类别中的目标具有背景复杂的特点. 因此, 所提检测算法适合应用于复杂的遥感图像检测场景中.

2.4 消融实验

为了验证本研究中 ASPP+和所提注意力模块的性能, 以 YOLOv5s 作为 Baseline, 在 DIOR 数据集上进行消融实验的结果如表 4 所示. ASPP 为改进前的空洞空间金字塔池化模块, ASPP+为改进后的空洞空间金字塔池化模块, 括号中的数值为模块中扩张系数的设置, CBAM 为已有注意力模块, AM 为本研究所提注意力模块. 由表 4 可知, 当 ASPP+模块的扩张系数设置为 (3, 6, 9) 时 mAP 值最高, 因此将 ASPP+模块中的扩张系数设置为 (3, 6, 9).

为了验证各模块对遥感图像目标检测效果的影响, 选择精确率 P 、召回率 R 、mAP 和检测速度 FPS 作为实验的评价指标, 对所有模块的有效性进行验证, 各模块实验结果对比如表 5 所示.

表 3 不同算法模型在 NWPU VHR-10 测试集上的对比

Tab.3 Comparison of different algorithm models on NWPU VHR-10 test set

%

算法模型	mAP	AP									
		airplane	ship	storage tank	baseball diamond	tennis court	basketball court	ground track field	harbor	bridge	vehicle
RetinaNet ^[22]	84.3	91.2	82.8	88.5	93.8	83.0	85.9	79.4	73.5	78.8	86.0
文献[25]	83.8	90.2	86.2	90.1	96.7	89.8	68.5	91.0	81.4	63.9	79.2
文献[26]	84.8	93.0	84.5	87.1	92.8	82.0	89.0	78.0	76.0	81.0	84.5
YOLOv5s	90.1	94.6	90.3	81.8	92.2	90.5	88.7	99.5	93.1	82.1	88.2
Ours	91.6	95.3	91.9	88.7	95.8	91.2	88.5	99.5	92.4	85.1	87.6

表 4 ASPP+和注意力模块在精确度方面的性能对比

Tab.4 Performance comparison of ASPP+ and attention module in terms of mAP

Baseline	ASPP	ASPP+ (1,3,5)	ASPP+ (3,6,9)	ASPP+ (6,12,18)	CBAM	AM	mAP/%
√							68.7
√	√						68.8
√		√					69.1
√			√				70.3
√				√			69.8
√			√		√		69.2
√			√			√	70.9

表 5 各模块添加后的实验结果

Tab.5 Experimental results after adding each module

模型	P / %	R / %	mAP / %	FPS / (frame·s ⁻¹)
YOLOv5s	65.3	70.2	68.7	28.1
YOLOv5s-ASPP+	64.4	71.0	70.3	27.4
YOLOv5s-ASPP+-AM	63.7	72.2	70.9	25.9
YOLOv5s-ASPP+-AM-BiFPN	67.0	72.5	71.6	25.4

对于实验结果数据的具体分析如下: 1) 如表 5 所示, 在 YOLOv5s 中加入改进的 ASPP+模块后, 算法在 DIOR 数据集上的 mAP 达到 70.3%, 比 YOLOv5s 模型提升了 1.6%. 该实验数据证实: 改进的 ASPP+模块获得的多尺度特征可以进一步提高检测精度. 2) 在遥感图像目标检测过程中, 为了体现重要特征的关注以及不必要特征的抑制, 在网络中嵌入注意力模块, 嵌入注意力模块后的网络在 DIOR 数据集上的 mAP 为 70.9%, 比 YOLOv5s-ASPP+提高了 0.6%. 同时, 检测速度仅降低了 1.5 FPS, 仍能保证检测的实时性. 3) BiFPN 模块可以有效融合高层特征图和低层特征图, 加入 BiFPN 之后网络模型在 DIOR 数据集上的 mAP 为 71.6%, 其性能比 YOLOv5s 模型提升了 2.9%, 检测精度较高. 上述消融实验的结果证明, 本研究所提算法在不影响遥感图像目标检测实时性的前提下获得了更高的检测精度.

3 结 语

为实现遥感图像目标的精准检测, 提出一种联合多尺度与注意力机制的检测算法. 该算法在特征提取阶段, 通过 ASPP+模块提取多尺度上下

文信息, 同时嵌入注意力模块, 抑制遥感图像中复杂背景的干扰. 在特征融合阶段引入 BiFPN 加强网络特征的融合, 丰富语义信息. 在对检测框进行后处理过程中, 使用 DIoU-NMS 代替传统 NMS, 使检测框易重叠和错误抑制的问题得到改善. 在 DIOR 和 NWPU VHR-10 数据集上的实验结果表明, 在遥感图像目标检测中, 所提算法的 mAP 高于对比算法. 在未来的研究中, 一方面将针对遥感图像目标角度多样性的特点优化检测算法, 以进一步提高遥感图像目标检测精度, 另一方面从模型的轻量化方面展开研究, 最终目的在于提高检测速度.

参考文献 (References):

[1] 姜鑫, 陈武雄, 聂海涛, 等. 航空遥感图像的实时舰船目标检 [J]. 光学精密工程, 2020, 28(10): 2360–2369.
JIANG Xin, CHEN Wu-xiong, NIE Hai-tao, et al. Real-time ships target detection based on aerial remote sensing images [J]. **Optics and Precision Engineering**, 2020, 28(10): 2360–2369.

[2] 聂光涛, 黄华. 光学遥感图像目标检测算法综述 [J]. 自动化学报, 2021, 47(8): 1749–1768.
NIE Guang-tao, HUANG Hua. A survey of object detection in optical remote sensing images [J]. **Acta Automatica Sinica**, 2021, 47(8): 1749–1768.

[3] 王昶, 张永生, 王旭, 等. 基于深度学习的遥感影像变化检测方法 [J]. 浙江大学学报:工学版, 2020, 54(11): 2138–2148.
WANG Chang, ZHANG Yong-sheng, WANG Xu, et al. Remote sensing image change detection method based on deep neural networks [J]. **Journal of Zhejiang University: Engineering Science**, 2020, 54(11): 2138–2148.

[4] YANG X, SUN H, SUN X, et al. Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network [J]. **IEEE Access**, 2018, 6: 50839–50849.

[5] FENG J, LIANG Y P, YE Z W, et al. Small object detection in optical remote sensing video with motion guided R-CNN [C]// **IEEE International Geoscience and Remote Sensing Symposium**. Waikoloa: IEEE, 2020: 272–275.

[6] GUAN H Y, YU Y T, LI D L, et al. Road Caps FPN: capsule feature pyramid network for road extraction from VHR optical remote sensing imagery [J]. **IEEE Transactions on Intelligent Transportation Systems**, 2021: 1–11.

[7] COURTRAIL, PHAM M T, LEFEVRE S. Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks [J]. **Remote Sensing**, 2020, 12(19): 3152.

- [8] ZHANG X D, ZHU K, CHEN G Z, et al. Geospatial object detection on high resolution remote sensing imagery based on double multi-scale feature pyramid network [J]. **Remote Sensing**, 2019, 11(7): 755.
- [9] LI L L, CHENG L, GUO X H, et al. Deep adaptive proposal network in optical remote sensing images objective detection [C]// **IEEE International Geoscience and Remote Sensing Symposium**. Waikoloa: IEEE, 2020: 2651–2654.
- [10] 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.
- [11] HE W P, HUANG Z, WEI Z F, et al. TF-YOLO: an improved incremental network for real-time object detection [J]. **Applied Sciences**, 2019, 9(16): 3225.
- [12] SHAMSOLMOALI P, CHANUSSOT J, ZAREAPOOR M, et al. Multi-patch feature pyramid network for weakly supervised object detection in optical remote sensing images [J]. **IEEE Transactions on Geoscience and Remote Sensing**, 2021, 60: 1–13.
- [13] 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.
- [14] BERTASIUS G, TORRESANI L, YU S X, et al. Convolutional random walk networks for semantic image segmentation [C]// **2017 IEEE Conference on Computer Vision and Pattern Recognition**. Honolulu: IEEE, 2017: 858–866.
- [15] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. [2022-01-14]. <https://arxiv.53yu.com/abs/1706.05587v3>.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation network [C]// **IEEE Conference on Computer Vision and Pattern Recognition**. Salt Lake City: Computer Vision Foundation, 2018: 7132–7141.
- [17] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]// **European Conference on Computer Vision**. Berlin: Springer, 2018: 3–19.
- [18] 周勇, 陈思霖, 赵佳琦, 等. 基于弱语义注意力的遥感图像可解释目标检测 [J]. **电子学报**, 2021, 49(4): 679–689.
- [19] ZHOU Yong, CHEN Si-lin, ZHAO Jia-qi, et al. Weakly semantic based attention network for interpretable object detection in remote sensing imagery [J]. **Acta Electronica Sinica**, 2021, 49(4): 679–689.
- [20] ZHANG Y N, KONG J, QI M, et al. Object detection based on multiple information fusion net [J]. **Applied Sciences**, 2020, 10(1): 418.
- [21] TAN M X, PANG R M, LE Q V. Efficientdet: scalable and efficient object detection [C]// **IEEE Conference on Computer Vision and Pattern Recognition**. Seattle: Computer Vision Foundation, 2020: 10778–10787.
- [22] 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.
- [23] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]// **IEEE International Conference on Computer Vision**. Venice: IEEE, 2017: 2980–2988.
- [24] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]// **IEEE Conference on Computer Vision and Pattern Recognition**. Salt Lake City: Computer Vision Foundation, 2018: 8759–8768.
- [25] ZHANG J, XIE C M, 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.
- [26] WANG C, BAI X, WANG S A, et al. Multiscale visual attention networks for object detection in VHR remote sensing images [J]. **IEEE Geoscience and Remote Sensing Letters**, 2018, 16(2): 310–314.
- [27] JIANG S L, YAO W, WONG M S, et al. An optimized deep neural network detecting small and narrow rectangular objects in Google Earth Images [J]. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, 2020, 13: 1068–1081.