

基于多尺度特征融合网络的遥感影像目标检测

田婷婷^{1,2,3}, 杨军^{1,2,3*}¹兰州交通大学测绘与地理信息学院, 甘肃 兰州 730070;²地理国情监测技术应用国家地方联合工程研究中心, 甘肃 兰州 730070;³甘肃省地理国情监测工程实验室, 甘肃 兰州 730070

摘要 在图像分析与解译中, 遥感影像目标检测是一项基础性的工作。针对遥感影像目标尺度多样和背景复杂等问题, 提出了一种多尺度空洞卷积特征融合检测器(MDCF²Det)来实现遥感目标的精确检测。首先, 改进原始特征金字塔网络(FPN), 用空洞卷积代替普通卷积, 增大感受野; 其次, 增加从输入节点到输出节点的跳跃连接操作以充分地利用不同层级的语义和位置信息; 最后, 为了抑制噪声并突出前景, 在区域候选网络前增加多维注意力机制模块, 从而实现更精确的遥感影像目标检测。在DOTA和RSOD数据集上进行了实验, 所提算法的mean average precision (mAP)分别达到了92.95%和73.39%。实验结果表明, 所提算法能够有效提升遥感影像目标检测精度。

关键词 遥感; 目标检测; 多尺度; 卷积神经网络; 特征融合

中图分类号 TP751.1

文献标志码 A

DOI: 10.3788/LOP202259.1628003

Object Detection For Remote Sensing Image Based on Multiscale Feature Fusion Network

Tian Tingting^{1,2,3}, Yang Jun^{1,2,3*}¹Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China;²National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, Gansu, China;³Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, Gansu, China

Abstract Object detection in remote sensing images is a fundamental task in image analysis and interpretation. We proposed a Multiscale Dilated Convolution Feature Fusion Detector (MDCF²Det) to achieve precise object detection in remote sensing by addressing the problems of multiscale objects and the complexity of the background. To begin, we improve the original feature pyramid network by replacing the general convolution with the dilated convolution to increase the receptive field. Second, to take full advantage of different levels of semantic and location information, we add a skip connection operation from the input node to the output node. Finally, to suppress the noise and highlight the foreground, we add the multi-dimensional attention model before the regional proposal network, to achieve more accurate object detection in remote sensing images. Experiments are carried out on the DOTA and RSOD datasets, and the proposed algorithm's mean average precision reaches 92.95% and 73.39% respectively. The results show that the proposed algorithm can significantly improve the object detection accuracy of remote sensing images.

Key words remote sensing; object detection; multiscale; convolutional neural network; feature fusion

1 引言

遥感影像目标检测是航空和卫星影像分析领域的

基础性研究工作, 在资源探测、环境监测等方面有重要的应用。近年来, 随着卫星和遥感技术的快速发展, 大量高分辨率遥感影像得以获取, 大大提高了对自动分

收稿日期: 2021-07-01; 修回日期: 2021-07-07; 录用日期: 2021-07-09

基金项目: 国家自然科学基金(61862039)、甘肃省科技计划(20JR5RA429)、兰州市人才创新创业项目(2020-RC-22)、兰州交通大学天佑创新团队(TY202002)

通信作者: *yangj@mail.lzjtu.cn

析和解译遥感影像的需求^[1-2]。自 2012 年 AlexNet 获得 ILSVRC 2012 挑战赛冠军后,神经网络开始受到广泛关注,传统目标检测算法逐渐被基于卷积神经网络的检测算法所取代。近年来,基于卷积神经网络的算法在自然场景图像的分类、语义分割和目标检测等方面取得了较好的成果,然而由于遥感影像的特殊性,目标检测在遥感影像领域还有待进一步研究。相比于自然图像,遥感影像具有以下特性^[3]:1)目标地物的尺度多样性,同类目标在同一幅影像中的尺度大小不一;2)信息量大且复杂,目标所处的背景包含的信息十分丰富且复杂;3)小目标多且密集排列,遥感影像中小目标种类和数量较多,且普遍具有排列密集的特点。因为遥感影像是俯瞰拍摄的,其包含的空间场景更大更复杂,且包含的目标种类和数量更多,所以无法直接将针对自然图像设计的深度学习网络直接迁移到遥感影像中,遥感影像的目标检测是一项更具挑战性的任务。此外,由于卷积神经网络提取特征时,深层的特征感受野比较大,语义信息丰富但缺乏空间几何特征信息;浅层特征具有丰富的几何细节信息,但由于经过的卷积次数少,其感受野小、语义信息少且噪声较多,因此如何更好地利用浅层与深层特征信息也是亟待解决的难点问题。

在目标检测任务中,目标尺度变化是影响算法检测效果的一个主要问题。在计算机视觉发展初期,传统方法均使用图像金字塔处理多尺度目标,在图像金字塔中,直接对图像进行不同尺度的缩放,然后将这些图像输入检测器中进行检测,取得了不错的效果,但其缺点在于增加了时间成本。为了解决这个问题,文献[4]提出特征金字塔网络(FPN)。FPN加入了新的top-down分支用来弥补浅层的语义信息不足的缺陷,使最终输出的特征能够更好地表示输入图像各个维度的信息。PANet^[5]在 FPN 的基础上进行了改进,引入bottom-up path augmentation结构,通过从最底层到最顶层的信息传输路径,充分利用特征融合,加强了特征金字塔。随着神经架构搜索(NAS)的发展,文献[6]在一个覆盖所有交叉尺度连接的可扩展搜索空间中,采用神经网络结构搜索,发现了一种新的特征金字塔结构:NAS-FPN,它由自顶向下和自下而上的连接组合而成,可以跨范围地融合特征。文献[7]采用对称的编码器-解码器(Encoder-Decoder)结构,融合深度网络中的高维和低维特征,使输出的多尺度信息更为丰富,并对顶层和底层两个不同尺度输出预测结果进行双重约束,进一步提升了建筑物检测精度。Deng等^[8]重新设计特征提取器,采用Concatenated ReLU和Inception模块设计了多尺度目标候选网络(MMS-OPN)和精确目标检测网络(AODN),该网络有效提高了多模态遥感影像多尺度目标检测的精度。Azimi等^[9]针对复杂背景下目标规模、方向、类别的变化问题,提出了一种基于多尺度卷积核的图像级联和特征

金字塔网络,该网络能够在各种尺度上提取信息并显著改善检测结果。Zhou等^[10]提出了Rotated Feature Network (RFN),该网络的编码器单元分配权重,解码器单元通过对旋转和重加权特征映射分别执行恢复操作来提取特征映射,解决了特征映射的分类和回归问题。姚群力等^[11]设计了一种多尺度特征金字塔网络(EPFN),该设计增强了网络对多尺度目标的特征提取能力,并引入聚焦分类损失加强了网络对难样本的学习能力,解决了遥感影像复杂背景下多尺度目标检测精度低以及网络泛化能力差的问题。

此外,由于遥感影像的特殊性,相比自然图像会包含更多的背景噪声,而过多的噪声会导致目标之间的界限变得模糊,导致漏检或虚警增加等问题^[12]。文献[13]首次提出了适用于目标检测的注意力模型,该模型利用图像中目标之间的相互关系来优化检测效果,将注意力机制引入Non-Maximum Suppression (NMS)操作中,实现了端到端训练,且提升了检测网络的性能。文献[14]结合了空间注意力机制和通道注意力机制,提出Convolutional Block Attention Module (CBAM),与SENet^[15]相比,两个注意模块的结合能更有效地突出特征并抑制背景,且CBAM具有轻量化的特点,可集成到任何卷积神经网络(CNN)架构中,额外开销可忽略不计。文献[12]为了更有效地捕捉遥感影像复杂背景下小物体的特征,设计了一种有监督的多维注意力网络(MDA-Net),该网络在减少噪声的同时增强了前景,虽然没有完全消除非目标信息,但有助于保留上下文信息并提高鲁棒性。Wang等^[16]针对高分辨率遥感影像中目标方向和尺度的变化以及背景杂乱的问题,提出了一种多尺度视觉注意力网络(MS-VANs),该网络对每个尺度的特征都使用一个注意力网络学习,在突出目标区域的同时抑制背景噪声。

综上所述,目前大多数目标检测算法是针对自然场景设计的,而对具有背景信息复杂、数据量大、尺度变化多样等特点的遥感影像,不能有效提取影像中地物的细节特征,检测效果和整体精度均有待提高。因此本文针对遥感影像数据集的特点,提出了一种结合注意力机制的多尺度空洞卷积特征融合目标检测框架(MDCF²Det)。首先,使用空洞卷积代替FPN中的卷积操作以增大局部感受野;其次,增加从原始输入到输出节点的边,在不增加过多计算复杂度的同时更好地融合特征;然后,添加空间和通道注意力模块,抑制遥感影像中固有的噪声,使网络在特征尺度上聚焦于前景信息;最后,经过全连接层(FC),通过分类和回归两个分支,实现目标类别的识别和位置的确定。

2 多尺度目标检测框架

2.1 网络结构

所提结合注意力机制的多尺度特征融合的目标检测算法的网络框架如图1所示。整个网络由骨干网络

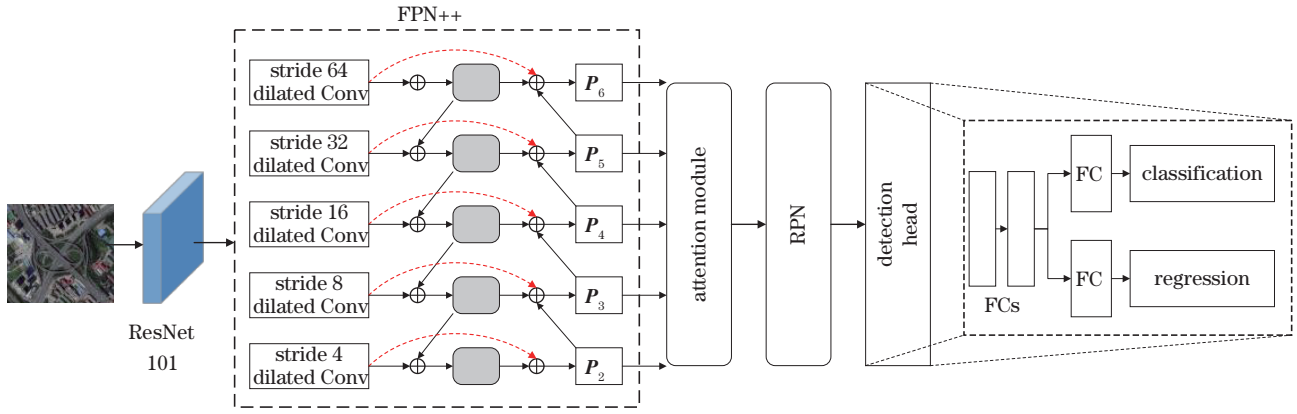


图 1 所提目标检测模型框架
Fig. 1 Architecture of proposed object detection model

ResNet101、多尺度空洞卷积特征融合网络 FPN++、注意力模块、候选区域提取和分类回归 5 部分组成。FPN++是在原始 FPN 的基础上进行改进的,使用空洞卷积代替普通卷积,增大局部感受野,并增加了从输入节点到输出节点的跳跃连接,以融合信息更丰富的多尺度特征。

2.2 特征融合

2.2.1 空洞卷积

空洞卷积^[17]又名扩张卷积,不同于普通的卷积,空洞卷积层引入空洞率来定义卷积操作时卷积核处理数据值的间距。在相同的计算复杂度下,空洞卷积可以提供比普通卷积更大的感受野,因此空洞卷积被广泛应用于语义分割等领域。本实验组对原始 FPN 进行改进,使用空洞卷积代替卷积,以增大感受野,同时使特征图保留更多目标信息。以 3×3 卷积核为例,对比普通卷积与空洞卷积的区别,结果如图 2 所示,其中 [图 2(a)] 为扩张率为 1 的 3×3 普通卷积, [图 2(b)] 为扩张率为 2 的 3×3 空洞卷积。很明显可以看出,空洞卷积在计算量保持不变的情况下完成了感受野从 5×5 到 7×7 的变化。

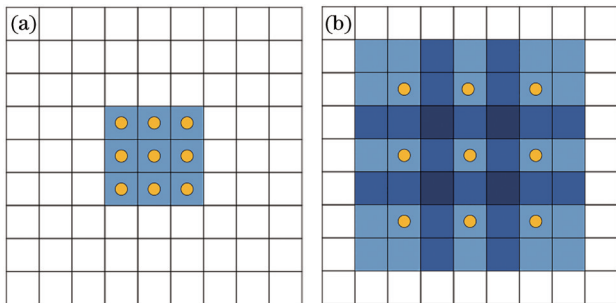


图 2 普通卷积与空洞卷积对比。(a)普通卷积;(b)空洞卷积
Fig. 2 Comparison between convolution and dilated convolution.
(a) General convolution; (b) dilated convolution

2.2.2 FPN++

在卷积神经网络中,浅层特征图的空间分辨率高,几何细节信息表征能力强,但其感受野较小,语义信息

比较匮乏,适合检测小目标;深层的特征图分辨率低,空间几何特征细节匮乏,但其感受野大,语义信息丰富。多尺度特征融合网络将浅层特征与深层特征相互融合,构建具有细粒度特征和丰富语义信息的特征表示,融合后的特征具有更强的描述性,更有利于目标的检测。

FPN 通过对深层特征进行上采样和对浅层特征进行自顶向下的连接,将深层的特征信息自上而下传递到浅层,并进行特征融合,再分别对每一层进行预测,如 [图 3(a)] 所示。但 FPN 只能从上向下进行融合, PANet 针对该缺陷自下而上进行二次融合,进一步提高了小目标的检测精度,证明了双向融合的有效性,如 [图 3(b)] 所示。因此,为了提高检测精度,本实验组对特征融合网络进行改进,提出了 FPN++。首先使用空洞卷积代替普通卷积,其次增加了输入节点到输出节点的跳跃连接来更好地进行特征融合,如图 1 黑色虚线框所示。利用改进的多尺度空洞卷积特征融合网络来融合不同的深层和浅层特征图之间的信息:自上而下的路径可以用深层的语义信息细化浅层的特征,自下而上的路径可以将浅层的位置信息插值到深层,同一尺度的输入到输出的跳跃连接可以改善上采样无法完全恢复已经丢失的信息的问题,使最后得到的特征的每一层都可以同时捕获丰富的语义表示和位置信息,以实现遥感影像复杂地物目标的高精度检测。

原始 FPN 以自顶向下的方式聚合多尺度特征,即

$$\begin{cases} P_5 = \text{Conv}(C_5) \\ P_4 = \text{Conv}[C_4 + \text{Resize}(P_5)] \\ \dots \\ P_2 = \text{Conv}[C_2 + \text{Resize}(P_3)] \end{cases}, \quad (1)$$

式中:Resize 是用于分辨率匹配的上采样或下采样操作;Conv 是用于特征处理的卷积操作; C_i 表示第 i 层的特征图; P_i 为融合后第 i 层的输出特征。本实验组以基础网络提取的第 2~6 层特征图 $\{C_2, C_3, C_4, C_5, C_6\}$ 作

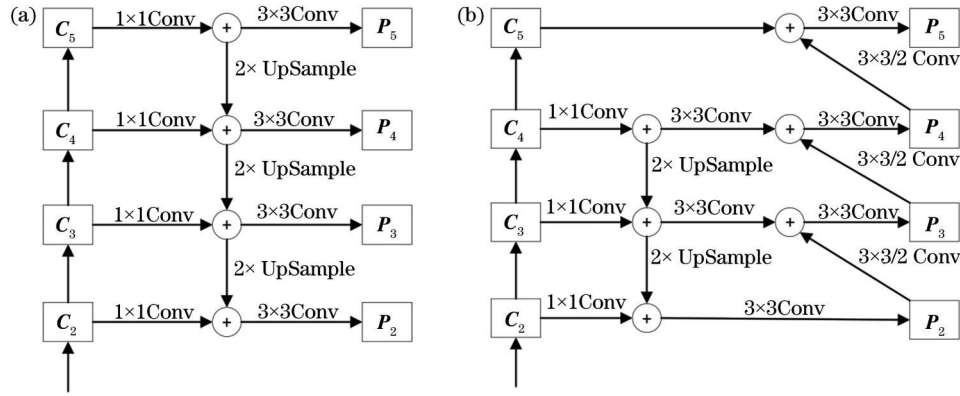


图 3 特征融合网络。(a) FPN; (b) PANet

Fig. 3 Feature fusion network. (a) FPN; (b) PANet

为输入,每一层特征图的大小为输入图像的 $1/2^i$,在特征金字塔网络中添加同一尺度的从输入到输出的跳跃连接来融合深层与浅层特征图之间的信息。特征图 P_i 首先经过卷积核尺寸为 3×3 ,步长为 2 的卷积层,此时特征图尺寸缩减为原来的一半,然后和 P_{i+1} 以及经过 3×3 卷积之后的原始输入 C_{i+1} 相加,得到的结果再经过一个卷积核尺寸为 3×3 ,步长为 1 的卷积层,最后进行融合得到表达能力更强的 P_{i+1} 。以 P_5 为例:

$$\begin{cases} P_6^{ms} = \text{Conv}(C_6) \\ P_5^{ms} = \text{Conv}[C_5 + \text{Resize}(P_6^{ms})] \\ P_5 = \text{Conv}[C_5 + P_5^{ms} + \text{Resize}(P_4)] \end{cases}, \quad (2)$$

式中: P_i^{ms} 是第 i 层的中间状态值; P_6^{ms} 由 C_6 经过 1×1 卷积得到; P_5^{ms} 由 C_5 经过 1×1 卷积,融合 P_6^{ms} 的 $2 \times$ Upsample 特征,再经过 3×3 卷积得到。 C_5 与 P_5^{ms} 经过 1×1 卷积,融合 P_4 进行 $2 \times$ Upsample 的特征,再经过 3×3 卷积得到 P_5 。

2.3 多维注意力模块

为了更有效地捕捉复杂背景下的小目标特征,本

实验组设计了一种多维注意力模块,针对遥感影像中背景信息复杂且存在大量噪声的特点,注意力机制能很好地引导网络去学习遥感影像中的目标特征,抑制背景信息,从而提高检测精度,如图 4 所示,其中 $U \in \mathbf{R}^{H \times W \times C}$, H 表示特征的高度 (Height), W 表示宽度 (Width), C 表示通道数 (Channel)。该模块由通道注意力机制和空间注意力机制组成。首先将特征输入通道注意力模块,经过全局平均池化 (GAP) 后输出 C 个特征通道的数值分布,然后将特征的维度降低到输入的 $1/r$, r 为缩放参数,经过 ReLU 激活后再通过一个全连接层将维度恢复至原始维度。然后通过 Sigmoid 获得归一化的权重。将通道注意力模块生成的特征输入空间注意力模块中,经过全局平均池化和全局最大池化得到两个不同的特征描述 $F_{\max} \in \mathbf{R}^{H \times W \times 1}$ 和 $F_{\text{avg}} \in \mathbf{R}^{H \times W \times 1}$,将两个特征描述按通道合并后经过卷积层生成大小为 $H \times W \times C$ 的注意力特征图。通道注意力模块通过建模通道间的相关性,达到关注感兴趣的通道特征的同时抑制非必要通道特征的目的,而空

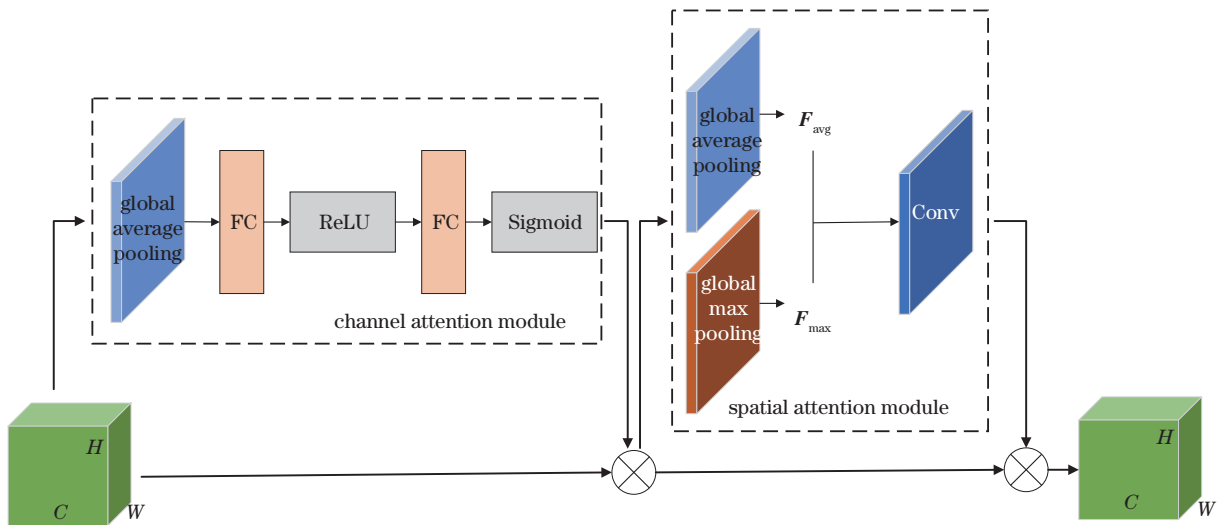


图 4 多维注意力模块

Fig. 4 Multi-dimensional attention module

间注意力模块可以有效地捕捉特征的空间位置信息,增强特征图的表现力。多维度注意力机制在通道注意力和空间注意力两个方面对特征图进行前景突出和噪声抑制,在不增加过多参数的同时有效增强了特征表示,可更好地实现精确检测。

2.4 损失函数设计

为了改善样本不平衡问题,本实验组设计了多任务损失函数,其表达式为

$$L(p_j, p_j^*, \mathbf{b}_j, \mathbf{b}_j^*) = \frac{1}{N_{\text{cls}}} \sum_j L_{\text{cls}}(p_j, p_j^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_j p_j^* L_{\text{reg}}(\mathbf{b}_j, \mathbf{b}_j^*), \quad (3)$$

式中: j 表示每个batch size中候选框的索引; p_j 表示预测为目标的概率, p_j^* 表示目标的真实类别标签,当候选框中的目标为正样本时, p_j^* 为1,否则 p_j^* 为0; \mathbf{b}_j 表示预测的bounding box的4个参数化坐标向量, \mathbf{b}_j^* 表示真实框的坐标向量; N_{cls} 和 N_{reg} 分别为特征提取网络和预测网络中正样本包围盒的数量; λ 为平衡超参数; L_{cls} 和 L_{reg} 分别为分类损失和回归损失函数。为加强网络的学习能力,设计 L_{cls} 为二元交叉熵损失函数, L_{reg} 为smooth_{L1}损失函数,其表达式分别为

$$L_{\text{cls}}(p_j, p_j^*) = -\log[p_j p_j^* + (1 - p_j^*)(1 - p_j)], \quad (4)$$

$$L_{\text{reg}}(\mathbf{b}_j, \mathbf{b}_j^*) = \sum_j \text{smooth}_{L1}(\mathbf{b}_j - \mathbf{b}_j^*), \quad (5)$$

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases} \quad (6)$$

3 实验结果与分析

3.1 数据集

为了验证所提算法的有效性,在RSOD^[18-19]与DOTA^[20]数据集上分别进行了训练、测试和评估。RSOD是包含飞机、操场、立交桥和油罐四类目标的遥感影像目标检测数据集。RSOD主要来源于Google Earth和Tianditu,包含946张影像,6950个实例,像素大小范围为1024×768~1044×915,空间分辨率为0.3~3 m。实验时,随机选取该数据集中70%的影像作为训练集,30%的影像作为测试集与评估集。DOTA是高分遥感影像目标检测数据集,包含飞机、棒球场、桥梁、田径场、小型车辆、大型车辆、船舶、网球场、篮球场、储油罐、足球场、环形车道、港口、游泳池和直升机等15个种类的目标地物。DOTA包含2806幅航拍影像,图像大小范围为800×800~4000×4000,包含各种尺度、方向和形状的目标,完整注释的DOTA数据集基准包含188282个实例。实验时,随机选取该数据集的1/2原始影像作为训练集,1/6作为验证集,1/3作为测试集。

3.2 实验环境及参数设置

实验的硬件环境为Intel Core i9-9900k+NVIDIA RTX2080Ti(11 GB显存),实验程序在Linux Ubuntu

18.04+Tensorflow框架下实现。

因为DOTA数据集中影像像素过大,无法直接输入网络模型进行训练,需要对其裁剪,本实验组首先对DOTA数据集中的影像进行预处理,将其分割成为1024 pixel×1024 pixel的子图像,重叠像素为200 pixel。分别在RSOD和DOTA数据集上通过K-means方法进行聚类,得到合适的锚框尺寸。采用随机梯度下降法更新权值,动量参数为0.9,权重衰减为0.0005,设置最大迭代次数为300000,初始学习率为0.001,批归一化大小为1。

3.3 评价指标

为了定量评价实验结果,采用平均检测精度(AP)和类别平均检测精度(mAP)两个评价指标评估模型的性能。AP是根据召回率(R)和精确率(P)计算得到的,即

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (7)$$

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (8)$$

$$P_{\text{AP}} = \int_0^1 P(R) dR, \quad (9)$$

式中: N_{TP} 为真正例的数量,表示预测值为1,真实值为1; N_{FP} 为假正例的数量,表示预测值为1,真实值为0; N_{FN} 为假反例的数量,表示预测值为0,真实值为1; P 为预测正确的正例数占预测为正例总量的比率,即查准率; R 为预测正确的正例数占真正的正例数的比率,即查全率; P 是一个以 R 为参数的函数。

mAP为所有类别的平均AP值,其表达式为

$$P_{\text{mAP}} = \frac{1}{K} \sum_{k=1}^K R_{\text{AP}_k}, \quad (10)$$

式中: K 为目标类别个数。

3.4 实验结果

图5、图6分别为所提算法在RSOD和DOTA数据集上测试得到的结果。从图中可以看出,不管是田径场等大型目标,还是飞机、车辆等小目标,所提算法均取得了较为理想的检测结果。主要原因在于改进后的特征融合网络将浅层特征与深层特征进行了充分融合,得到了含有丰富信息的融合特征,同时注意力模块抑制了影像的背景噪声并突出了目标信息,从而提升了多尺度目标的检测精度。

所提算法整体上取得了比较理想的检测结果,但也存在漏检及错检的现象,如图7矩形框标注部分。因为遥感影像中的目标具有密集排列且杂乱无序的特点,所以在检测时会出现漏检的现象,如图7(a)所示;且像素较小的地物具有相似的纹理和轮廓特征,如图7(b)将船舶错检为小型车辆。

表1、表2分别为RSOD和DOTA数据集各类别的检测精度。RSOD数据集中的各类目标均达到了较高的检测结果,尤其是立交桥和操场的检测精度达到

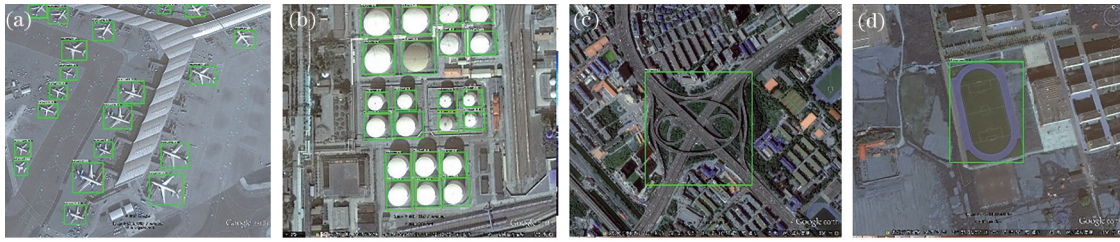


图5 RSOD数据集检测结果。(a)飞机;(b)油罐;(c)立交桥;(d)操场
Fig. 5 Detection results on RSOD dataset. (a) Aircraft; (b) oiltank; (c) overpass; (d) playground



图6 DOTA数据集检测结果。(a)飞机;(b)田径场和足球场;(c)桥梁;(d)棒球场;(e)大型车辆和小型车辆;(f)环形车道;(g)港口;
(h)储油罐;(i)船舶;(j)篮球场和网球场;(k)游泳池;(l)直升机
Fig. 6 Detection results on DOTA dataset. (a) PL; (b) GTF and SBF; (c) BR; (d) BD; (e) LV and SV; (f) RA; (g) HA; (h) ST; (i) SH;
(j) BC and TC; (k) SP; (l) HC



图7 漏检和错检示例。(a)漏检;(b)错检
Fig. 7 Examples of miss detection and error detection. (a) Miss detection; (b) error detection

表1 RSOD数据集的检测结果

Table 1 Detection results on RSOD dataset unit:%

Dataset	Aircraft	Oiltank	Overpass	Playground	mAP
RSOD	81.02	90.77	100	100	92.95

了100%。而DOTA数据集中的目标大小不一、方向各异,背景环境更复杂并且尺度变化更大,因此检测难度相对较高。从表2可以看出,网球场、棒球场、飞机和储油罐等地物取得了比较高的检测精度,其精度均在84%以上,这是由于这几类目标的特征较为明显且背景较为单一,而形状、纹理等特征相似的车辆与船舶等尺度较小的目标以及数据集中数量较少且背景复杂

表 2 DOTA 数据集的检测结果
Table 2 Detection results on DOTA dataset

unit: %

Dataset	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
DOTA	89.68	84.39	52.12	72.71	64.49	67.07	77.45	90.11	83.98	86.01	64.03	63.33	74.45	67.74	62.87	73.39

的桥梁等目标检测难度更大,因此检测精度略低。

改进的 FPN++ 在提高目标检测精度方面起到了主要作用,为了验证 FPN++ 的有效性,本实验组设计了一组多尺度特征融合对比试验,分别应用于 RSOD 与 DOTA 数据集。为了控制变量,实验仅改变 FPN 部分,其余网络框架保持不变,检测结果如表 3、表 4 所示。在 RSOD 数据集中飞机、油罐、立交桥和操

场的检测精度均有不同程度的提升,尤其是飞机相较于原始 FPN 提高了 2.16 个百分点。DOTA 数据集中绝大多数地物的检测精度有一定程度的提高,飞机、棒球场、网球场、游泳池等目标的检测精度提升较明显,车辆、船舶等尺寸较小、分布较密集的地物仅有略微提高。这是由于特征提取与融合时使用了空洞卷积代替普通卷积,增大了局部感受野,捕获了更丰富的多尺度信息,同时,增加的同级跳跃连接可以有效增强输出特征包含的位置与语义信息。因此,在多尺度遥感目标检测中,使用 FPN++ 进行特征融合的算法的检测精度优于使用原始 FPN 的算法,这有效验证了 FPN++ 能够将浅层语义信息与深层语义信息更好地融合,从而提高目标检测精度。

表 3 RSOD 数据集上不同 FPN 实验对比

Table 3 Experimental comparison of different FPN on

RSOD dataset unit: %

Type	Aircraft	Oiltank	Overpass	Playground	mAP
FPN	78.86	90.48	99.73	100.00	92.26
FPN++	81.02	90.77	100.00	100.00	92.95

表 4 DOTA 数据集上不同 FPN 实验对比

Table 4 Experimental comparison of different FPN on DOTA dataset

unit: %

Type	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
FPN	88.81	84.33	50.98	72.45	62.71	65.35	75.34	89.96	81.16	84.07	53.26	63.33	72.24	66.29	60.11	71.36
FPN++	89.68	84.39	52.12	72.71	64.49	67.07	77.45	90.11	83.98	86.01	55.03	63.33	74.45	67.74	62.87	72.86

为了验证所提算法的有效性,与其他经典检测网络模型在 RSOD 与 DOTA 数据集上进行了实验对比,结果如表 5、表 6 所示。表 5 为所提算法 MDCF²Det 与 R-FCN^[21]、Deformable R-FCN^[22]、Faster R-CNN^[23] 和 RFN^[10] 在 RSOD 数据集上的检测精度对比。从表 5 可

以看出,所提算法的检测精度优于其他 4 种算法。因为立交桥和操场目标尺寸大、形状纹理等特征明显,易于检测,所以所提算法与 Faster R-CNN 均取得了比较理想的检测结果;飞机的检测精度较 R-FCN、Deformable R-FCN、Faster R-CNN 三种算法提高了近 10 个百分点。表 6 为所提算法 MDCF²Det 与 YOLO v2^[24]、R-FCN^[21]、CenterFPANet^[25]、FPN^[4]、ICN^[9]、FMSSD^[26] 在 DOTA 数据集上的检测精度对比。从表 6 可以看出,所提算法在棒球场、篮球场、储油罐和直升机的检测精度为所有算法中最高,分别为 84.39%、83.98%、86.01% 和 62.87%,而车辆与船舶等像素过小且密集排列的地物检测精度略低,其余地物如飞机、网球场和桥梁等目标检测精度大致与该类别的最高检测精度持平。从整体上看,所提算法优于其他算法,验证了该算法的有效性。

表 5 不同算法在 RSOD 数据集上的精度对比

Table 5 Accuracy comparison of different algorithms on RSOD

dataset unit: %

Algorithm	aircraft	oiltank	overpass	playground	mAP
R-FCN	71.48	90.23	76.84	97.70	84.07
Deformable R-FCN	71.50	90.26	81.48	99.53	85.70
Faster R-CNN	71.90	90.90	100.00	100.00	90.70
RFN	79.10	90.50	100.00	99.70	92.30
MDCF ² Det	81.02	90.77	100.00	100.00	92.95

表 6 不同算法在 DOTA 数据集上的精度对比

Table 6 Accuracy comparison of different algorithms on DOTA dataset

unit: %

Algorithm	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
YOLO v2	76.90	33.87	22.73	34.88	38.73	32.02	52.37	61.65	48.54	33.91	29.27	36.83	36.44	38.26	11.61	39.20
R-FCN	79.33	44.26	36.58	53.53	39.38	34.15	47.29	45.66	47.74	65.84	37.92	44.23	50.64	50.64	34.90	47.24
CenterFPANet	88.74	71.52	48.95	52.06	48.55	73.37	61.14	90.53	57.63	84.06	66.64	62.71	73.33	57.63	42.76	65.29
FPN	88.70	75.10	52.60	59.20	69.40	78.80	84.50	90.60	81.30	82.60	52.50	62.10	76.60	66.30	60.10	72.00
ICN	90.00	77.70	53.40	73.30	73.50	65.00	78.20	90.80	79.10	84.80	57.20	62.11	73.45	70.22	58.08	72.45
FMSSD	89.11	81.51	48.22	67.94	69.23	73.56	76.87	90.71	82.67	73.33	52.65	67.52	72.37	80.57	60.15	72.43
MDCF ² Det	89.68	84.39	52.12	72.71	64.49	67.07	77.45	90.11	83.98	86.01	55.03	63.33	74.45	67.74	62.87	72.86

4 结 论

针对背景复杂、尺度多样的高分辨率遥感影像目标检测的准确性问题,提出了一种多尺度特征融合的目标检测算法。该算法首先使用空洞卷积代替普通卷积以增大局部感受野,同时设计跳跃连接将原始输入与输出节点相连,能够更好地融合提取到的浅层特征与深层特征,构建兼具丰富位置信息与语义信息的特征表示;其次,在区域候选网络前加入通道和空间注意力机制,以抑制背景的噪声信息并突出前景信息,提高了模型对复杂地物的辨识能力。实验结果表明,与其他经典目标检测算法相比,所提算法提高了遥感影像目标检测的精度。然而,所提算法还存在一定局限性,例如并没有充分利用遥感影像中地物方向任意的特点设计旋转框,因此,在未来的研究中,会根据遥感影像中目标方向的特点设计更为合理的旋转包围框,获得更高的目标检测精度。

参 考 文 献

- [1] 戴媛, 易本顺, 肖进胜, 等. 基于改进旋转区域生成网络的遥感图像目标检测[J]. 光学学报, 2020, 40(1): 0111020.
Dai Y, Yi B S, Xiao J S, et al. Object detection of remote sensing image based on improved rotation region proposal network[J]. Acta Optica Sinica, 2020, 40(1): 0111020.
- [2] 李竺强, 朱瑞飞, 马经宇, 等. 联合连续学习的残差网络遥感影像机场目标检测方法[J]. 光学学报, 2020, 40(16): 1628005.
Li Z Q, Zhu R F, Ma J Y, et al. Airport detection method combined with continuous learning of residual-based network on remote sensing image[J]. Acta Optica Sinica, 2020, 40(16): 1628005.
- [3] 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): 0228003.
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.
- [4] 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.
- [5] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[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: 8759-8768.
- [6] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection [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: 7029-7038.
- [7] 伍广明, 陈奇, ShibasakiRyosuke, 等. 基于U型卷积神经网络的航空影像建筑物检测[J]. 测绘学报, 2018, 47(6): 864-872.
Wu G M, Chen Q, Shibasaki R, et al. High precision building detection from aerial imagery using a U-net like convolutional architecture[J]. Acta Geodaetica et Cartographica Sinica, 2018, 47(6): 864-872.
- [8] Deng Z P, Sun H, Zhou S L, et al. Multi-scale object detection in remote sensing imagery with convolutional neural networks[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 3-22.
- [9] Azimi S M, Vig E, Bahmanyar R, et al. Towards multi-class object detection in unconstrained remote sensing imagery[EB/OL]. (2018-07-07) [2021-03-02]. <https://arxiv.org/abs/1807.02700>.
- [10] Zhou K B, Zhang Z X, Gao C X, et al. Rotated feature network for multiorientation object detection of remote-sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(1): 33-37.
- [11] 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): 1128002.
Yao Q L, Hu X, Lei H. Object detection in remote sensing images using multiscale convolutional neural networks[J]. Acta Optica Sinica, 2019, 39(11): 1128002.
- [12] Yang X, Yang J R, Yan J C, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8231-8240.
- [13] Hu H, Gu J Y, Zhang Z, et al. Relation networks 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: 3588-3597.
- [14] 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.
- [15] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [16] Wang C, Bai X, Wang S, et al. Multiscale visual attention networks for object detection in VHR remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(2): 310-314.
- [17] Yu F, Koltun V. Multi-Scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2021-02-03]. <https://arxiv.org/abs/1511.07122>.
- [18] Long Y, Gong Y P, Xiao Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2486-2498.
- [19] Xiao Z F, Liu Q, Tang G F, et al. Elliptic Fourier

- transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. *International Journal of Remote Sensing*, 2015, 36(2): 618-644.
- [20] Xia G S, Bai X, Ding J, et al. DOTA: a large-scale dataset for object detection in aerial images[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: 3974-3983.
- [21] Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]//NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. [S.l.: s.n.], 2016: 379-387.
- [22] 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.
- [23] 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.
- [24] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [25] Gu X, Kong L B, Wang Z C, et al. A light-weight object detection framework with FPA module for optical remote sensing imagery[C]//Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence, July 3, 2020, Qingdao, China. New York: ACM Press, 2020: 109-115.
- [26] Wang P J, Sun X, Diao W H, et al. FMSSD: feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(5): 3377-3390.