

基于多尺度卷积神经网络的遥感目标检测研究

姚群力^{1,2*}, 胡显^{1,2}, 雷宏¹

¹中国科学院电子学研究所航天微波遥感系统部, 北京 100190;

²中国科学院大学电子电气与通信工程学院, 北京 100049

摘要 针对现有遥感图像目标检测算法对于复杂场景下多尺度目标检测精度较低、泛化能力差的问题, 提出了一种多尺度卷积神经网络遥感目标检测框架——MSCNN。该方法通过构造一种深度特征金字塔, 增强了网络对多尺度目标特征的提取能力; 引入聚焦分类损失作为分类损失函数, 加强了网络对难样本的学习能力。所提方法在 NWPU VHR-10 公开数据集上取得了 0.960 的平均检测精度(mAP), 相较于 RetinaNet 检测框架, MSCNN 对小尺度以及中等尺度目标的平均检测精度分别提高了 1.5% 和 1.9%, 实现了对多尺度目标的高精度稳健检测。

关键词 遥感; 目标检测; 卷积神经网络; 多尺度目标

中图分类号 TP753

文献标识码 A

doi: 10.3788/AOS201939.1128002

Object Detection in Remote Sensing Images Using Multiscale Convolutional Neural Networks

Yao Qunli^{1,2*}, Hu Xian^{1,2}, Lei Hong²

¹Department of Space Microwave Remote Sensing Systems, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China;

²School of Electronics, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract A novel detection framework is proposed based on a multiscale convolutional neural network (MSCNN) to overcome low precision and insufficient generalization ability associated with existing object detection methods for multiscale objects with complex scenes. First, an essence feature pyramid network is constructed to enhance the extraction ability of multiscale features. Then, the focal classification loss is introduced as classification loss function to enhance the learning capability of the MSCNN over complex samples. The proposed method achieves a mean average precision(mAP) of 0.960 over the challenging NWPU VHR-10 dataset. In comparison with the RetinaNet detection method, the mAP of the proposed MSCNN on small- and medium-scale objects increases by 1.5% and 1.9%, respectively. The proposed method is found to be accurate and robust for multiscale objects.

Key words remote sensing; object detection; convolutional neural networks; multi-scale objects

OCIS codes 280.4788; 150.1135; 100.4996

1 引 言

遥感目标自动检测技术不仅是一种实现遥感目标自动分类和定位的智能化数据分析方法, 还是遥感图像解译领域的重要研究方向之一^[1-5]。遥感目标自动检测在城市规划、交通安全和环境监测等方面发挥着重要作用。然而, 在遥感目标检测中, 除了要考虑诸如光照、遮挡和几何形变等客观因素外, 目标实例的多尺度性、密集分布和形状差异等亦是需

要重点考虑的。因此, 复杂场景下多尺度的地理空间目标检测需要更具体的探测。

传统的遥感图像目标检测方法, 如尺度不变特征变换(SIFT)^[6]、梯度方向直方图(HOG)^[7], 以及可变形部件模型(DPM)^[8]等, 都是根据人工经验设计特征的, 虽然在特定的应用场景下能取得较好的检测效果, 但该类方法对先验知识的依赖性强, 导致检测模型的自适应性及泛化能力较差。

深度卷积神经网络(Deep ConvNets)能够从数

收稿日期: 2019-04-08; 修回日期: 2019-06-27; 录用日期: 2019-07-26

基金项目: 国家自然科学基金青年基金(61422113, 61601437)、国家重点研发计划(2017YFB0502700)

* E-mail: yaoqunli15@mails.ucas.ac.cn

据中主动学习特征,且不依赖于人工经验,为遥感图像目标检测研究提供了一种新的思路和框架。基于 Deep ConvNets 的目标检测模型在研究理念上可以分为两类:1) 基于回归的 Single-stage 目标检测框架(如 DSSD^[9]、RetinaNet^[10],以及 RefineDet^[11]等)。该类检测方法从回归的角度出发,直接在图像上回归出目标的边框位置和物体类别。如 Chen 等^[12]在 SSD(Single Short MultiBox Detector)^[13]检测框架的基础上提出了一种输入图像尺度可变的检测方法,Xia 等^[14]提出了基于 YOLOv2^[15]的遥感图像目标检测。但该类方法对于多尺度遥感目标的检测精度较差。2) 基于区域建议的 Two-stage 目标检测框架(如 Faster R-CNN^[16]、Mask RCNN^[17]、FPN^[18],以及 Cascade RCNN^[19]等)。该类方法基于原始图像生成一系列区域建议,并将区域建议和特征图输入感兴趣区域池化(ROI Pooling)层,最终实现目标的分类与定位。如 Han 等^[20]基于 Faster R-CNN^[16]体系结构提出的 R-P-Faster R-CNN,不仅实现了端到端的训练与检测,还提高了模型的检测精度和效率。Deng 等^[4]提出了一种基于多尺度目标建议网络(MS-OPN)和高精度目标检测网络(AODN)的多尺度遥感目标检测框架,进一步提高了多尺度目标的检测精度。任之俊等^[21]提出了一种基于改进特征金字塔的 MaskR-CNN 目标检测方法,该方法有效提高了中小尺度目标的检测精度。朱明明等^[22]也提出了基于级联区域建议网络的机

场目标检测算法。但该类算法的高计算成本限制了目标检测的速度。此外,在 FPN 体系结构中,较大的步长易造成小尺度目标在深层特征图中丢失的现象,进而导致上下文线索被削弱。基于此,本文先设计了一种多尺度特征金字塔网络 EFPN(Essence Feature Pyramid Networks)。通过控制基础网络的采样步长,引入膨胀瓶颈结构,增强了网络对多尺度特征的表达能力。引入了聚焦分类损失作为分类损失函数,加强了网络对遥感图像中难样本的学习能力,并通过极小化多任务损失函数端到端优化整个网络。本文基于 EFPN 构造了一种基于多尺度卷积神经网络的遥感目标检测框架(MSCNN),有效地提高了多尺度遥感目标的预测能力。

2 MSCNN 检测框架

2.1 MSCNN 网络结构

本文以 RetinaNet 目标检测网络作为基线结构,提出了一种多尺度目标检测框架——MSCNN,整体结构框架如图 1 所示,其中 A 为膨胀瓶颈结构,B 为带有 1×1 卷积的膨胀瓶颈结构, W 、 H 为卷积核尺度。首先以 ResNet-50 作为基础网络构造了一种新的特征金字塔网络 EFPN,生成了新的融合预测特征,提高了传统特征提取网络对多尺度特征的表达能力。之后通过目标检测子网络,分别给出多尺度目标的分类得分和边框位置。最后再采用非极大值抑制原理将相似结果进行合并,输出最终检测结果。

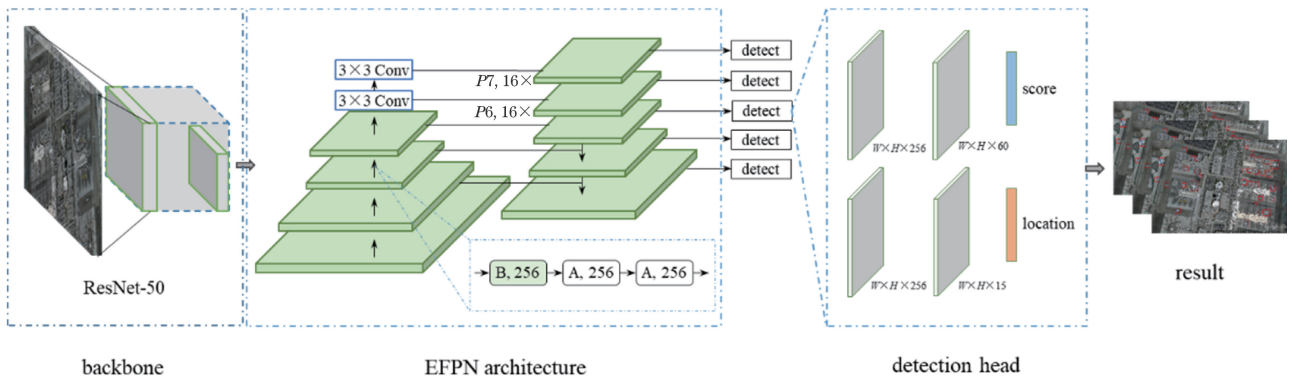


图 1 MSCNN 目标检测框架结构

Fig. 1 Target detection framework of MSCNN

2.2 EFPN

本文以 ResNet-50 作为基线,构建了 EFPN。为了构建自底向上的通路,从主干中选取卷积块 $\{C_3, C_4, C_5\}$ 作为基础层级结构。添加特征映射 C_6 和 C_7 ,以获得更精确的语义信息,特征层 C_6 和 C_7 的计算式为

$$C_6 = \text{Conv2D}[k = 256, s = (3, 3)](C_5), \quad (1)$$

$$C_7 = \text{Conv2D}[k = 256, s = (3, 3)][\text{ReLU}(C_6)], \quad (2)$$

式中:Conv2D 为二维卷积算子,它将给定的特征图与预定义的卷积核进行卷积; k 为卷积核的数量; s 为内核的尺度;ReLU 为激活函数。因此,通过自

底向上的路径生成了特征图 $\{C_3, C_4, C_5, C_6, C_7\}$ 。事实上, 为了进一步提高网络的特征表达能力, 如图 2(a)所示, 本文在 ResNet-50 的第 4 阶段以后, 首先将步长锁定为 $16 \times$, 然后将通道维度控制为 256, 接着在每个阶段之后均部署了一个低复杂度的膨胀瓶颈结构^[23-24], 如图 2(c)所示, 以保障网络能生成更深的特征图, 且能保持较高的分辨率。最后通过自上而下的通路, 构建出了特征金字塔网络层级 $\{P_3, P_4, P_5, P_6, P_7\}$ 。新的特征金字塔 EFPN 将包含 $P_3 \sim P_7$ 共 5 层预测层。这些特征层级的计算方法为

$$R_N = \text{Conv2D}[k = 256, s = (1, 1)](C_N), \quad (3)$$

$$T'_N = R_N + \sum_{j=N+1}^7 U(T_j, C_N), \quad (4)$$

$$P_N = \text{Conv2D}[k = 256, s = (3, 3)](T'_N). \quad (5)$$

式中: 序数 $N=3, 4, 5, 6, 7$; 序数 $N'=3, 4, 5, 6$, 且 j 为由 N 生成的索引; R_N 为自底向上路径经过卷积降维得到的特征层; T_N 为经过特征堆叠得到的新特征; $U(\cdot)$ 为将 T_j 的尺度调整到 C_N 大小的算子; P_N 为自顶向下构造的输出特征。所有的预测特征最后都将被送入检测网络进行多尺度目标预测。

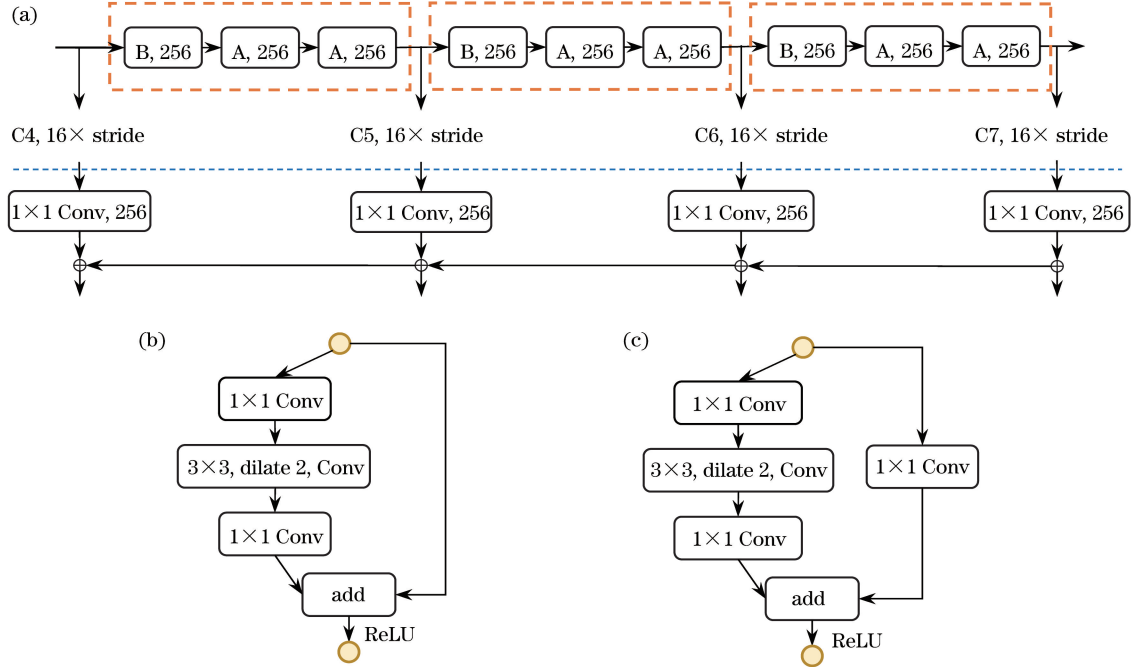


图 2 EFPN 网络构型和膨胀瓶颈结构。(a) EFPN 模块结构; (b) 膨胀瓶颈结构; (c) 带有 1×1 卷积的膨胀瓶颈结构
Fig. 2 Structures of EFPN and dilated bottleneck. (a) Structure of EFPN module; (b) dilated bottleneck structure; (c) dilated bottleneck structure with 1×1 Conv

2.3 损失函数设计

在实际的遥感影像数据中, 正、负样本的不平衡现象普遍存在。然而, 极端的正、负样本不平衡将会导致网络训练过程中正样本在数据中占比较少, 从而使网络训练效率降低, 使得一些难以学习的正样本得不到充分的学习, 严重制约了检测网络在遥感目标检测任务中的检测性能。为此, 本文引入聚焦分类损失^[10], 通过动态缩放交叉熵, 快速地将模型的训练集中在难样本稀疏集上, 用以加强网络对于遥感图像中难样本的学习和挖掘。本文所提出的多任务联合损失函数为

$$L(p_i, p_i^*, t_i, t_i^*) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{fl}}(p_i, p_i^*) + \alpha \frac{1}{N_{\text{loc}}} \sum_i p_i^* L_{\text{loc}}(t_i, t_i^*), \quad (6)$$

式中: N_{cls} 为批量尺度; N_{loc} 为锚点框数量; α 为平衡超参数; p_i^* 为目标的真实类别标签; p_i 为预测相应类别的概率; t_i 为预测的 4 个参数化坐标向量; t_i^* 为真实边框参数; i 为批量数据中锚点框的索引; L_{fl} 和 L_{loc} 分别为聚焦分类和边框回归损失, 其中 L_{fl} 可定义为

$$L_{\text{fl}}(p_T) = -\alpha_t (1 - p_t)^\gamma \log_2(p_t), \quad (7)$$

式中: $\alpha_t \in [0, 1]$, 为平衡超参数; $(1 - p_t)^\gamma$ 为调制因子, p_t 为前景目标对应的预测概率; $\gamma \in [0, 5]$, 为超参数, 当 $\gamma > 0$ 时意味着模型将更专注于难样本的训练。 L_{loc} 通常采用 Smooth L1 损失, 其表达式为

$$L_{\text{loc}}(t_i, t_i^*) = s_{\text{smoothL1}}(t_i - t_i^*), \quad (8)$$

$$s_{\text{smoothL1}}(x) = \begin{cases} 0.5x^2, & \text{if } |x| = 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (9)$$

3 实验结果与分析

3.1 数据集与评价指标

本文在 NWPU VHR-10^[25] 公开数据集上进行多尺度目标检测实验。NWPU VHR-10 数据集是一个用于多类多尺度目标检测的地理空间目标检测公开数据集。该数据集共包含 650 幅光学遥感图像,平均尺度约为 600×800 ,共标注了 757 架飞机、302 艘船只、655 个油罐、390 个棒球场、524 个网球场、159 个篮球场、163 个田径场、224 个港口、124 座桥梁,以及 477 辆车。图像分辨率在 $0.5 \sim 2.0$ m 之间,每一幅图像至少包含一个目标。

为了避免过拟合,本文采用旋转、翻转进行数据扩充。旋转变换采用的步长为 90° ,将图像从 0° 旋转到 270° 进行数据扩充。翻转变换是分别对原图进行水平翻转和垂直翻转。在训练过程中,本文将数据集随机分配 60% 用于训练,余下的 40% 用于测试。根据 NWPU VHR-10 数据分布信息统计,定义了相应的边界框尺度量 (small, medium, and large),如表 1 所示。

表 1 基于实例尺度分布的边界框面积定义方法

Table 1 Definition of bounding box areas based on distribution of instance scales

Small	Medium	Large
(0,60]	(60,120]	(120, +∞)

本文采用平均检测精度 (mAP) 作为目标检测的评价指标,该指标衡量了所有类别的检测精度的均值。事实上,AP 计算的是召回率 (AR) 在 0 到 1

区间内精度的平均值,即精度-召回率曲线所包围的面积,因此 AP 值越高代表检测性能越好。精度 p 和召回率 r 指标可以表示为

$$p = \frac{T_P}{T_P + F_P}, \quad (10)$$

$$r = \frac{T_P}{T_P + F_N}, \quad (11)$$

式中: T_P 、 F_P 、 F_N 分别为真正例、假正例,以及假反例。此外,本文还评估了在不同 IOU 阈值和不同边界框尺度 (small, medium, and large) 下目标的检测精度和召回率,分析了所提方法对多尺度目标的检测能力。IOU 代表了检测框与真值框的交并比,其定义可以表示为

$$I_{OU} = \frac{G_T \cap D_R}{G_T \cup D_R}, \quad (12)$$

式中: G_T 为真值框; D_R 为检测结果。

3.2 参数设置

本文训练和测试采用的硬件平台为 NVIDIA Titan Xp GPUs,利用 Pytorch 开源深度学习框架完成实验的构建。实验过程采用端到端训练方式,初始学习率设置为 0.001,优化方法为随机梯度下降,动量设置为 0.9,正则化系数设置为 0.0005,批处理大小设置为 1。

3.3 主要结果

本文所有检测实验均在 NWPU VHR-10 公开数据集上进行训练和测试,结果如表 2 所示,表中给出了本文所提方法与当前一些典型目标检测网络的对比结果。

表 2 不同算法在 NWPU VHR-10 数据集上的检测精度对比

Table 2 Comparison of the detection precision of different algorithms on NWPU VHR-10 dataset

Method	RICNN ^[25]	FRCN-VGG-16 ^[16]	YOLO ^[15]	SSD ^[13]	R-FCN ^[26]	FRCN-Deform ^[27]	FPN ^[18]	MSDN ^[28]	MSCNN
Airplane	0.884	0.830	0.874	0.956	0.961	0.983	0.964	0.998	0.994
Ship	0.773	0.776	0.847	0.937	0.983	0.892	0.931	0.972	0.953
Storage tank	0.853	0.525	0.427	0.617	0.725	0.817	0.914	0.838	0.918
Baseball diamond	0.881	0.963	0.931	0.995	0.994	0.984	0.947	0.991	0.963
Tennis court	0.408	0.629	0.658	0.860	0.907	0.859	0.944	0.973	0.954
Basketball court	0.585	0.688	0.870	0.944	0.978	0.927	0.959	0.999	0.967
Ground track field	0.867	0.984	0.975	0.987	0.981	0.988	0.990	0.986	0.993
Harbor	0.686	0.819	0.800	0.950	0.924	0.946	0.921	0.972	0.955
Bridge	0.615	0.793	0.903	0.966	0.934	0.947	0.838	0.927	0.972
Vehicle	0.711	0.639	0.704	0.745	0.884	0.816	0.900	0.901	0.933
mAP	0.726	0.764	0.799	0.894	0.928	0.917	0.931	0.956	0.960

通过分析不同算法的检测精度可知,基于VGG-16的FRCN网络的检测精度相对较低。事实上,FRCN目标检测网络仅基于固定感受野尺度的末端特征图生成预测,然而较大的步长降低了末端特征图的分辨率,导致在复杂背景下FRCN对多尺度数据特征的特征能力严重不足,严重制约了检测器的检测性能。

相较于YOLO网络,SSD的目标检测精度得到了大幅提高,这再一次证明了多尺度特征金字塔构型能够有效提升网络对多尺度目标的建模能力,从而达到提升多尺度目标检测精度的目的。对比分析FRCN-Deform网络与FRCN-VGG-16网络可知,带有可变形卷积结构和可变形池化结构的FRCN-Deform目标检测网络相较于基于传统的卷积和池化操作的检测网络具有更强的特征建模能力,以及对目标多尺度和形变的适应能力。因此,FRCN-Deform较之FRCN-VGG-16,检测精度也就更高。

与R-FCN相比较,FPN的检测精度提升了0.3%。事实上,FPN网络构造了多尺度特征金字塔,对多尺度特征进行融合,显著增强了网络对多尺度特征的学习能力。因此,采用FPN网络可显著提升对多尺度遥感目标检测的效果。本文提出的MSCNN对于大多数类别的目标均取得了最佳的检测效果,对于小尺度目标,如:飞机目标,较MSDN

检测精度提高了0.4%;油罐目标,较FPN提高了0.4%;车辆目标,较MSDN提高了3.2%。对于较大尺度目标,如:田径场目标,较MSDN检测精度提高了0.7%;桥梁目标,较SSD提高了0.6%。MSCNN在NWPU VHR-10公开数据集上的平均检测精度相较于MSDN提升了0.4%,相较于FPN提升了2.9%。由上述分析可知,本文提出的基于EFPN的MSCNN能有效提高多尺度目标的检测性能。

3.4 消融实验

1) EFPN

在目标检测框架MSCNN中,EFPN对于提升多尺度目标检测精度具有至关重要的作用。为了证明EFPN组件在检测框架中所发挥的性能,本文设计了一组对比实验,实验结果如表3第2~3行所示。EFPN取得了 $0.960@AP_{50}$ 以及 $0.824@AP_{75}$ 的平均检测精度,相较于RetinaNet取得了 $1.5%@AP_{50}$ 和 $1.5%@AP_{75}$ 的增益,表现出了更高的检测精度。此外,EFPN还取得了 $0.547@small$ 、 $0.578@medium$ 和 $0.701@large$ 的多尺度平均检测精度,相较于RetinaNet取得 $1.5%@small$ 、 $1.9%@medium$ 和 $1.9%@large$ 的增益,表现出了对多尺度目标检测的优越性。因此,在多尺度遥感目标检测中,EFPN组件相较于RetinaNet具有更为明显的优势。

表3 MSCNN消融实验参数

Table 3 Ablation experimental parameters of MSCNN

Backbone	Average precision					
	$I_{OU}=0.5;0.95$	$I_{OU}=0.5$	$I_{OU}=0.75$	Small	Medium	Large
RetinaNet	0.690	0.945	0.809	0.532	0.559	0.682
EFPN	0.706	0.960	0.824	0.547	0.578	0.701
EFPN-NoProj	0.700	0.950	0.819	0.544	0.562	0.698

2) EFPN-NoProj

相较于RetinaNet目标检测网络,EFPN保持了特征图的更高空间分辨率,并且生成了新的阶段。为验证EFPN的确生成了新的阶段 P_6 、 P_7 ,设计对比实验,其结果如表3第3~4行所示。采用EFPN作为基础网络进行训练和测试,接着修改EFPN的 P_6 和 P_7 ,即采用图2(c)中包含 1×1 Conv映射的Dilated bottleNeck结构代替图2(a)中的结构,并将新的修改置于 P_5 之后。为方便起见,新的网络记为EFPN-NoProj。EFPN-NoProj的网络结构如图3所示。

由表3第3~4行的检测结果可知,EFPN较

EFPN-NoProj的检测精度有所提升。同时,实验结果还证明了含有 1×1 Conv映射的Dilated bottleNeck结构的EFPN网络在保证特征图有较高分辨率的同时,确实生成了新的特征阶段。

3.5 多尺度目标检测

本文评估了所提目标检测框架RetinaNet、MSCNN,以及EFPN-NoProj网络在不同IOU阈值和不同边框尺度下的平均检测精度和召回率,结果如表4所示。

AP_{50} 是评价模型分类能力的有效指标, AP_{75} 能够体现出检测框架对边界框位置回归的能力。如表4所示,MSCNN取得了 $0.960@AP_{50}$ 以及 $0.824@$

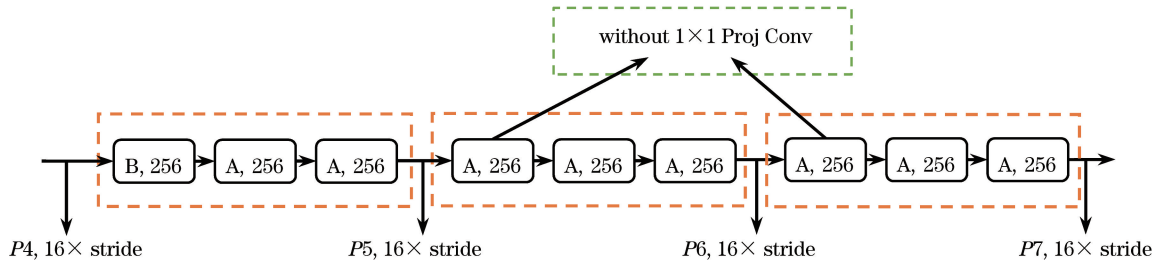


图 3 EFPN-NoProj 的网络结构

Fig. 3 Network structure of EFPN-NoProj

表 4 不同 IOU 阈值和不同边框尺度下的平均检测精度和召回率

Table 4 Average precision and average recall under different IOU thresholds and different bounding box areas

Backbone	Average precision			Average precision ($I_{OU}=0.5;0.95$)			Average recall ($I_{OU}=0.5;0.95$)		
	$I_{OU}=0.5;0.95$	$I_{OU}=0.5$	$I_{OU}=0.75$	Small	Medium	Large	Small	Medium	Large
RetinaNet	0.690	0.945	0.809	0.532	0.559	0.682	0.573	0.586	0.754
MSCNN	0.706	0.960	0.824	0.547	0.578	0.701	0.600	0.605	0.755
EFPN-NoProj	0.700	0.950	0.819	0.544	0.562	0.698	0.597	0.600	0.753

AP₇₅ 的平均检测精度, 相较于 RetinaNet 取得了 1.5% @ AP₅₀ 和 1.5% @ AP₇₅ 的增益。即 MSCNN 目标检测框架具有更高的分类能力和边框回归精度。此外, MSCNN 取得了 0.600 @ small、0.605 @ medium 和 0.755 @ large 的平均召回率, 相较于 RetinaNet 网络取得了 2.7% @ small、1.9% @ medium 和 0.1% @ large 的增益, 说明其对小尺度目标在召回率上具有一定优势。事实上, 与 RetinaNet 网络相比, MSCNN 在深层阶段的分辨率更高, 同时还生成了

更新的阶段。因此, MSCNN 可以在更深阶段检测到更小尺度的目标。

图 4 为部分实例检测结果, 从图中可以看出, 所提方法能够同时对多类多尺度目标进行准确检测。对于高分辨率光学遥感图像中的飞机、舰船、油管等排布相对密集的小目标, 以及车辆和网球场等包含复杂背景的目标, 所提方法均能给出准确的检测结果, 进一步证明了本文所提方法的有效性。

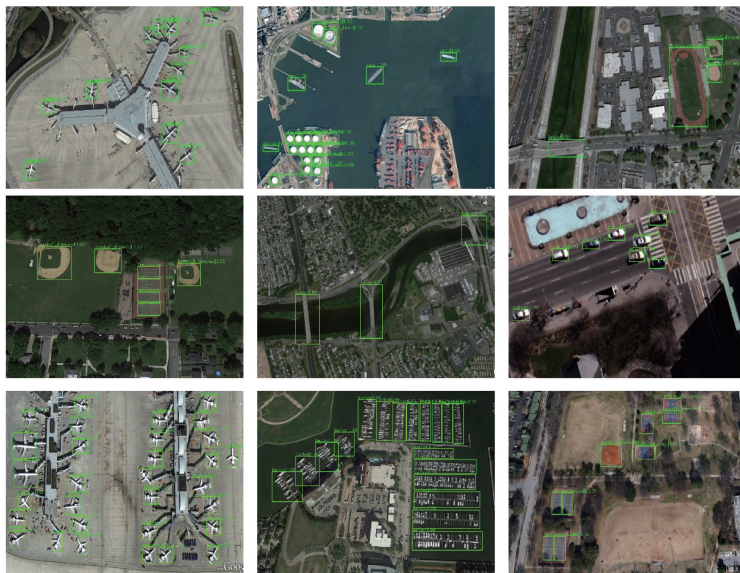


图 4 基于 MSCNN 的可视化检测结果

Fig. 4 Visual detection results of MSCNN

4 结 论

针对现有目标检测算法对复杂场景下多尺度目标检测精度低、泛化能力不足的问题,本文提出了一种基于多尺度卷积神经网络的遥感目标检测框架MSCNN。该方法首先设计了一种多尺度特征提取骨架,并在此基础上构造出一种深度特征金字塔EFPN,从而能够更加有效地对遥感图像中的多尺度目标特征的学习。此外,本文引入聚焦分类损失作为分类损失函数,进一步改善了目标检测网络对于遥感图像中难样本的挖掘能力。所提方法在NWPU VHRG10公开数据集上获得了0.960的平均检测精度,相较于其他遥感目标检测框架,实现了对多尺度遥感目标的高精度稳健检测。

参 考 文 献

- [1] Han J W, Zhang D W, Cheng G, *et al.* Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(6): 3325-3337.
- [2] Cheng G, Han J W. A survey on object detection in optical remote sensing images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 117: 11-28.
- [3] Cheng G, Han J W, Zhou P C, *et al.* Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 98: 119-132.
- [4] 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.
- [5] Zhong Y F, Han X B, Zhang L P. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 138: 281-294.
- [6] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [7] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE, 2005: 8588935.
- [8] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE, 2008: 10139902.
- [9] Fu C Y, Liu W, Ranga A, *et al.* Dssd: deconvolutional single shot detector[J/OL]. (2017-01-23) [2019-04-07]. <https://arxiv.org/abs/1701.06659>.
- [10] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2999-3007.
- [11] Zhang S F, Wen L Y, Bian X, *et al.* Single-shot refinement neural network 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, 2018: 4203-4212.
- [12] Chen Z, Zhang T, Ouyang C. End-to-end airplane detection using transfer learning in remote sensing images[J]. *Remote Sensing*, 2018, 10(1): 139.
- [13] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, *et al.* *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [14] 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, 2018: 3974-3983.
- [15] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [16] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems 28(NIPS 2015), December 7-12, 2015, Palais des Congrès de Montréal, Montréal Canada. Canada: NIPS, 2015: 91-99.
- [17] He K M, Gkioxari G, Dollár P, *et al.* Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2980-2988.
- [18] 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 (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 936-944.
- [19] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality 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, 2018: 6154-6162.
- [20] Han X B, Zhong Y F, Zhang L P. An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery[J]. Remote Sensing, 2017, 9(7): 666.
- [21] Ren Z J, Lin S Z, Li D W, *et al.* Mask R-CNN object detection method based on improved feature pyramid [J]. Laser & Optoelectronics Progress, 2019, 56(4): 041502.
任之俊, 蔺素珍, 李大威, 等. 基于改进特征金字塔的 Mask R-CNN 目标检测方法[J]. 激光与光电子学进展, 2019, 56(4): 041502.
- [22] Zhu M M, Xu Y L, Ma S P, *et al.* Airport detection method with improved region-based convolutional neural network[J]. Acta Optica Sinica, 2018, 38(7): 0728001.
朱明明, 许悦雷, 马时平, 等. 改进区域卷积神经网络的机场检测方法[J]. 光学学报, 2018, 38(7): 0728001.
- [23] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.
- [24] Chen L C, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFs[J/OL]. (2016-06-07) [2019-04-07]. <https://arxiv.org/abs/1412.7062>.
- [25] Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
- [26] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks[C] // Advances in Neural Information Processing Systems 29 (NIPS 2016), December 5-10, 2016, Centre Convencions Internacional Barcelona, Barcelona Spain. Canada: NIPS, 2016: 379-387.
- [27] Dai J F, Qi H Z, Xiong Y W, *et al.* Deformable convolutional networks[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 764-773.
- [28] Deng Z P, Sun H, Lei L, *et al.* Object detection in remote sensing imagery with multi-scale deformable convolutional networks[J]. Acta Geodaetica et Cartographica Sinica, 2018, 47(9): 1216-1227.
邓志鹏, 孙浩, 雷琳, 等. 基于多尺度形变特征卷积网络的高分辨率遥感影像目标检测[J]. 测绘学报, 2018, 47(9): 1216-1227.