

引文格式: 徐逸之,姚晓婧,李祥,等.基于全卷积网络的高分辨遥感影像目标检测[J].测绘通报,2018(1):77-82.DOI: 10.13474/j.cnki.11-2246.2018.0014.

基于全卷积网络的高分辨遥感影像目标检测

徐逸之^{1,2},姚晓婧¹,李祥^{1,2},周楠³,胡媛^{1,2}

(1. 中国科学院遥感与数字地球研究所,北京 100094;

2. 中国科学院大学,北京 100049; 3. 苏州中科天启遥感科技有限公司,江苏 苏州 215163)

摘要: 目标检测是遥感图像分析处理中的研究热点之一,具有十分重要的科研和应用价值。传统遥感影像目标检测方法多使用人工构造的浅层次特征,结合支持向量机、随机森林、Adaboost 等分类器进行目标识别,难以充分挖掘和利用影像中的深层特征。近年来,深度学习,特别是卷积神经网络在图像认知方面取得了巨大成功。在目标检测领域,以 Faster R-CNN 算法为代表的方法取得了突破性进展,检测精度大幅提高,检测速度达到了近实时的性能。但是, Faster R-CNN 算法由于使用了感兴趣区域(RoI)池化层,各个 RoI 计算不共享,因此检测速度依然有待提高。R-FCN 基于全卷积网络结构,同时采用位置敏感池化来引入平移变化,抵消全卷积网络造成的平移不变形问题,检测精度和效率都有了很大的提高。本文阐述了 R-FCN 算法原理,并运用于高分辨遥感影像目标检测,分析了不同参数和网络结构对 R-FCN 检测效果的影响,比较了利用 Fast R-CNN、Faster R-CNN 和 R-FCN 3 种算法进行飞机识别的性能。试验结果表明,利用 R-FCN 进行飞机识别定位可以达到 99.3% 的准确率和每张图 180 ms 的检测速度。

关键词: 高分辨率遥感;深度学习;全卷积网络;R-FCN;飞机检测

中图分类号: P237

文献标识码: A

文章编号: 0494-0911(2018) 01-0077-06

Object Detection in High Resolution Remote Sensing Images Based on Fully Convolution Networks

XU Yizhi^{1,2}, YAO Xiaojing¹, LI Xiang^{1,2}, ZHOU Nan³, HU Yuan^{1,2}

(1. Institute of Remote Sensing and Digital Earth, CAS, Beijing 100094, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Imagesky Remote Sensing Technology Co. Ltd., Suzhou 215163, China)

Abstract: Object detection is one of the research hotspots in remote sensing image analysis and processing, and has very important research and application value. Traditional remote sensing image target detection method uses the shallow artificial feature, and combines with classifiers, such as support vector machine, random forest and Adaboost, to realize object detection task. In recent years, deep learning, especially convolution neural network (CNN), has achieved great success in image recognition. In the field of target detection, the Faster R-CNN algorithm has made a breakthrough, the detection accuracy has been greatly improved, and the detection procedure has achieved nearly real-time performance. However, the Faster R-CNN algorithm uses RoI (region of interests) pooling layer, the RoI-wise calculation is not shared, so the detection speed is still unsatisfying. R-FCN, however, is based on the fully convolution network structure, and the position-sensitive pool is used to tackle the translation variance problem, which is concealed by the convolution neural network, thus lead to performance improvement. In this paper, the principle of R-FCN is described and applied to the detection of high-resolution remote sensing images. Different parameters and network structure are analyzed to find the best configuration. The performance of three popular algorithms, including Fast R-CNN, Faster R-CNN and R-FCN are compared. Experimental results show that R-FCN can achieve 99.3% precision and achieve at a test-time speed of 180 ms per image.

Key words: high resolution remote sensing; deep learning; fully convolution networks; R-FCN; aircraft detection

高分辨率遥感影像的目标分类与识别是遥感影像智能化信息提取中的关键技术,是精确制导、武器防御、海情监控等军事目标识别系统的关键,也是提升减灾应急、交通监管、渔业海事等民用系统智能化

水平的核心技术^[1]。随着传感器技术的飞速发展,遥感影像的空间分辨率不断提高。高空间分辨率遥感影像不仅降低了中低分辨率遥感影像中存在的混合像元问题,更提供了丰富的空间信息、几何结构和

收稿日期: 2017-04-05

基金项目: 江苏省测绘地理信息科研项目 (JSCHKY201720); 国家自然科学基金 (41701438); 科技基金性工作专项重点项目 (2014FY210800)

作者简介: 徐逸之 (1993—),男,硕士生,研究方向为地理信息系统研究与应用、数据挖掘。E-mail: sysu_xuyizhi@163.com

通信作者: 李祥。E-mail: lixiang01@radi.ac.cn

纹理信息,这为遥感影像上目标的检测提供了可能,因此研究基于高分辨率遥感影像的目标检测识别算法成为当前的迫切需求。传统遥感影像目标检测算法多使用人工构造的浅层次特征,如 HOG 特征^[2]、SIFT 特征^[3-4]等,识别效果严重依赖于专家知识及特征构造中的有效性,很难充分挖掘影像中深层特征。在目标检测的精度和效率方面,难以满足大范围自动化应用的需求,同时对于灾害应急等实时性要求较高的工程应用任务也十分局限。

近年来,深度学习算法在图像分类、自然语言处理、降维、目标检测、运动建模、人工智能等领域取得前所未有的成果^[5-11]。深度学习一般需大量的训练样本才容易提取其特征,这与遥感图像大数据的特性不谋而合。深度学习在遥感中的应用引起广泛重视,并在场景分类^[12]、高光谱图像分类^[13-14]等方面取得显著性成效。目标检测方面,Chen Xueyun^[15-16]使用深度置信网络和深度卷积网络进行了飞机和车辆的检测;Diao Wenhui^[17]提出了一种基于稀疏深度置信网络遥感影像目标检测方法,并在 QuickBird 影像上进行了飞机检测试验,达到了 88.9% 的准确率,每张图检测时间为 6.3 s; Han Junwei 等^[18]提出了一种基于深度玻尔兹曼机的遥感影像目标检测方法,并在 3 个光学遥感影像数据集上进行了试验,证明了方法的有效性;高常鑫^[19]使用栈式自编码器来提取高分辨率遥感影像中目标语义表征及上下文约束表征,以实现高精度目标检测。然而,这些方法都聚焦在使用深度学习算法来进行特征提取,然后使用提取的特征构造分类器。检测流程中并没有给出目标搜索的算法,而是需要使用扫描窗口来逐一判读目标位置,因而效率较低。Faster R-CNN 算法首次提出使用区域建议网络 (regional proposal network, RPN) 来提取候选区域,将候选区域提取与分类统一到同一框架,实现了端到端的目标检测过程,检测精度和效率都得到了大幅提高。王万国等^[20]使用 Faster R-CNN 算法开展了无人机巡检图像电力小部件识别研究,达到每张近 80 ms 的识别速度和 92.7% 的准确率。但是, Faster R-CNN 算法由于使用了 RoI 池化层,各个 RoI 计算不共享,因此检测速度依然较慢,无法达到实时性任务的需求。

全卷积网络 (fully convolutional networks) 为目标检测提供了一种新的高效检测框架,诸如残差网络 (ResNet)^[21] 和 GoogLeNets^[22] 等先进的图像分类网络均为全卷积网络。类似地,可在物体检测中用全卷积网络。然而,现有研究表明,直接使用全卷积网络进行目标检测效果远差于该网络的分类效果。

主要原因在于物体分类要求平移不变性越大越好,因而平移不变性较好的全卷积结构更受青睐,而物体检测需要一些平移变化的定位表示。为了解决这一问题, R-FCN (regional-fully convolutional networks)^[23] 算法采用专门的卷积层构建位置敏感分数图,每个空间敏感分数图编码感兴趣区域的相对空间位置信息。该方法在 ImageNet 挑战赛、COCO 目标检测任务中取得了较好的效果。

本文研究基于全卷积网络的 R-FCN 算法在高分辨率遥感影像目标检测中的应用,并以飞机识别为例,比较 Fast R-CNN^[24]、Faster R-CNN^[25] 和 R-FCN 3 种算法的检测精度和效率。

1 基于 R-FCN 的目标检测

1.1 整体框架

类似于区域卷积神经网络 (R-CNN),本文采用计算机视觉领域流行的目标检测策略,包括区域建议和区域分类两步。用 Faster R-CNN 中的区域建议网络 (RPN) 提取候选区域,该 RPN 为全卷积网络。效仿 Faster R-CNN,共享 RPN 和 R-FCN 的特征。R-FCN 用来对 RPN 提取的感兴趣区域分类。整体技术流程如图 1 所示。

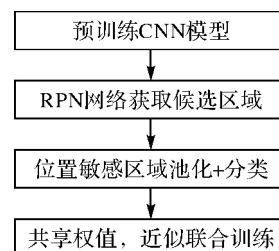


图1 基于 RFCN 的飞机检测流程

R-FCN 在与 RPN 共享的卷积层后多加 1 个卷积层,该卷积层在整幅图像上为每类生成 k^2 个位置敏感分数图,有 C 类物体外加 1 个背景,因此有 $k^2 (C+1)$ 个通道的输出层。 k^2 个分数图对应描述位置的空间网格。如 $k \times k = 3 \times 3$, 则 9 个分数图编码单个物体类。

R-FCN 最后用位置敏感 RoI 池化层,给每个 RoI 1 个分数。与 Faster R-CNN 的 RoI 池化方式不同,这里使用选择性池化。选择性池化图解:看上图的浅蓝色响应图像 (top-left),抠出浅蓝色方块 RoI,池化浅蓝色方块 RoI 得到浅蓝色小方块分数;其他颜色的响应图像同理。对所有颜色的小方块投票 (或池化) 得到各个类别的响应结果,也即 $C+1$ 个类别得分,最后使用 Softmax 分类器进行分类。如图 2 所示。

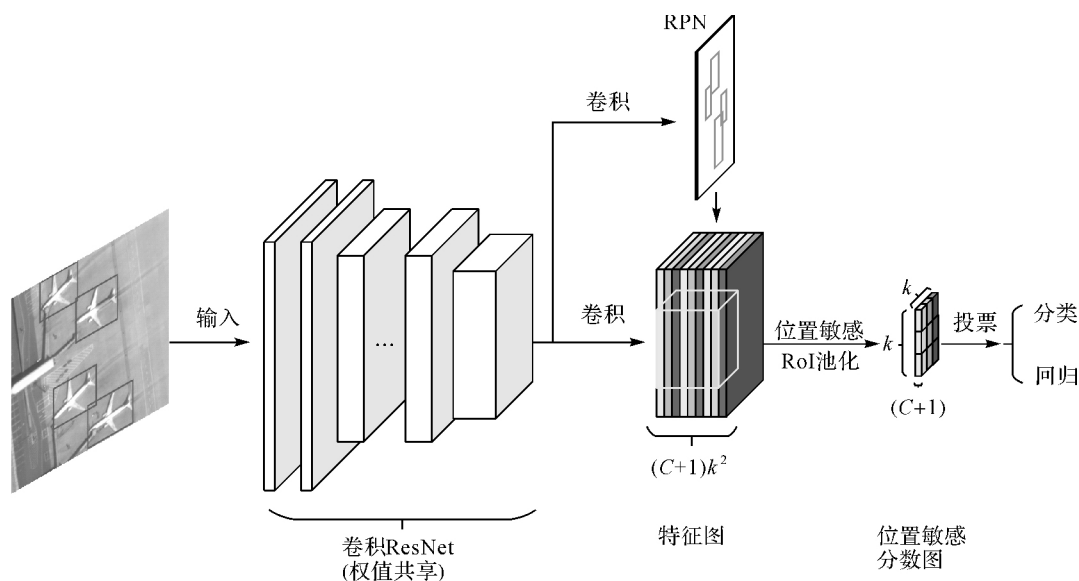


图2 R-FCN 网络结构

1.2 区域建议网络

RPN 网络的输入为一个卷积特征图,使用一个 $k \times k$ 滑动窗口(本文使用 3×3)扫描整个特征图,得到一个 $256-d$ 的特征图,然后使用大小为 1×1 的卷积核对其进行卷积运算,得到一个 $2k$ (类别)通道的分数图和一个 $4k$ (边界框)通道的分数图。试验中采用 3 种不同尺寸(128、256、512)和 3 种不同比例(1:1、1:2、2:1)组合成的 9 种基准小窗口对包含目标的窗口位置进行预测,使区域提议更准确。

1.3 位置敏感分数图

对 R-FCN 的卷积响应图像按 RPN 的结果分割出感兴趣区域,对部分通道(每类的每个相对空间位置通道中)的感兴趣区域分成 $k \times k$ 个网格,每个网格平均池化,然后所有通道再平均池化。对 1 个大小为 $w \times h$ 的 RoI,1 个桶(bin)的大小为 $\frac{w}{k} \times \frac{h}{k}$,最后 1 个卷积层为每类产生 k^2 个分数图。对第 (i, j) 个桶($0 \leq i, j \leq k-1$),定义位置敏感 RoI 池化操作如下

$$r_c(i, j | \theta) = \frac{1}{n(x, y) \in \text{bin}(i, j)} \sum z_{i, j, c}(x + x_0, y + y_0 | \theta)$$

式中 $r_c(i, j)$ 为第 c 类第 (i, j) 个桶的池化响应; $z_{i, j, c}$ 为 $k^2(C+1)$ 个分数图中的输出; (x_0, y_0) 为 RoI 的左上角坐标; n 为桶里的像素总数; θ 为网络的参数。

1.4 网络训练

每个 RoI 的损失函数为交叉熵损失与边界框回归损失的加权和

$$L(s, t, x, y, h) = L_{\text{cls}}(S_{c^*}) + \lambda [c^* > 0] L_{\text{reg}}(t, t^*)$$

式中 c^* 是 RoI 的真实类别标签, $c^* = 0$ 说明 RoI 为

背景。当 RPN 产生的区域建议当 RoI 与真实边框的 IoU(RoI 与真实边框交集的面积占它们并集面积的比例) 超过 0.5 时, 标定为正样本, 否则为负样本。 L_{cls} 使用 Softmax 损失函数, L_{reg} 使用 SmoothL1 损失函数, 计算方式参照文献[26], 同样设置 λ 为 1。

此外, 在网络训练过程中, 使用在线难例挖掘(OHEM)[28]。在线难例挖掘是一种 Hard Negative Mining 的方式, 用于启发式地选择出模型难以正确分类的样本, 从而使训练过程更为高效。假设每个图像前向产生 N 个区域建议, 计算所有建议的损失。按损失排序所有 RoIs, 选择损失最高的 B 个 RoIs, 反向传播过程只使用选择的难样本。

2 试验

2.1 试验数据

从谷歌地图上选取了 100 张包含飞机的图片, 原始图像大小近似为 600×600 像素, 并人工标注其中飞机样本外边框, 如图 3 所示。使用其中 50% 图片样本进行训练和验证, 剩余 50% 的图片样本用于测试。本次试验中, 用准确率和召回率来评判识别的准确性, 其中召回率为正确识别飞机数量除以测试图像中飞机总数量, 准确率为正确识别飞机数量除以模型识别出来飞机数量。

2.2 试验设置

本文 R-FCN 方法中 RPN 网络和检测网络使用 ResNet-50/ResNet-101 来构建。ResNet-50(101) 使用 50(101) 个卷积层, 后接 1 个平均池化层和 1 个全连接层。删除平均池化和全连接层, 只使用卷积层来计算特征图。使用 ImageNet 数据集对 ResNet-

51(101)网络进行了预训练。ResNet-51 中的最后一个卷积块是 2048-d,附加一个随机初始化的

1024-d 的 1×1 卷积层,放在 $k^2(C+1)$ 通道卷积层之前,用于减小尺寸。

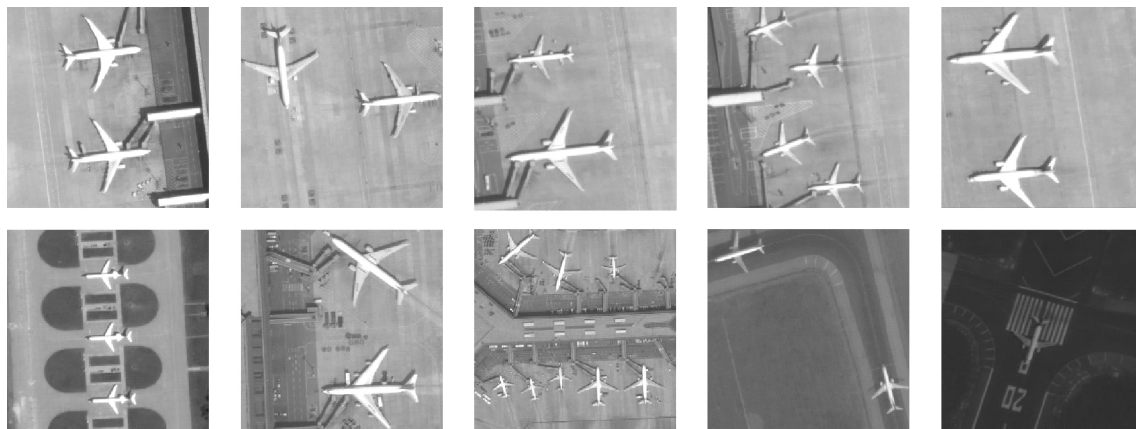


图3 飞机样本

在训练过程中,设置初始学习速率为 0.001,每迭代 2000 次降低为原来的 0.1 倍,学习动量为 0.9,正则化项系数为 0.001 6,批处理大小为 64。本文采用端到端(end-to-end)的方式来训练 RPN 网络和 R-FCN 网络。

本文所有试验均在同一台服务器上完成,采用 GPU 编程实现。试验中使用 GTX TITAN X 显卡(12 GB 显存)基于 Caffe 框架实现卷积神经网络模型。

3 结果与讨论

首先,本文试验了不同参数设置对平均准确率

均值(mAP)的影响。根据 R-FCN 论文结论可知,输入影像的尺寸、最大迭代次数及是否使用 OHEM 对检测精度有较大的影响,因此笔者进行了多组对比试验确定最优参数配置。表 1 展示了不同试验设置情况下的检测精度。从表 1 可知,不同参数配置下,R-FCN 算法精度变化不明显,均在 90%左右浮动;使用 OHEM 能提高 R-FCN 算法精度;增大输入影像尺寸和最大迭代次数,R-FCN 检测精度并没有提高,反而可能会降低。

表1 不同参数配置下 R-FCN 在测试集上 mAP 对比

(%)

参数配置	迭代次数				
	2000	4000	6000	8000	10 000
ResNet-50+scale = 600	90.3	90.0	90.6	90.4	90.0
ResNet-50+scale = 800	89.9	88.6	88.9	88.6	89.2
ResNet-50+scale = 600+OHEM	90.5	90.6	90.6	90.7	90.6
ResNet-50+scale = 800+OHEM	89.9	90.3	90.1	89.9	90.1
ResNet-101+scale = 600	90.5	90.4	89.8	89.9	89.9
ResNet-101+scale = 600+OHEM	90.5	89.0	89.1	88.6	88.6

然后,本文使用更深的 ResNet-101 构建 R-FCN 网络,参数调优过程同上,检测精度见表 1。从表 1 可知,使用更深的网络结构,检测精度几乎不变,导致这一现象的原因可能是网络太深造成了过拟合现象。使用以上确定的最优参数配置,随机从测试样本中挑选了 5 张图片进行测试,测试效果如图 4 所示。从图中可以看出,R-FCN 方法能够有效地识别各种尺寸、背景、光照条件下的飞机无错检和漏检的

情况。在检测 IoU 阈值设定为 0.5 的情况下,召回率达到 98.6%,对应的准确率为 99.3%。

最后,笔者比较了 R-FCN、Faster R-CNN、Fast R-CNN 这 3 种方法在测试集上的识别精度和速度,见表 2。需要说明的是,Fast R-CNN 算法使用 Selective Search 生成候选区域,而 Faster R-CNN 和 R-FCN 算法使用区域建议网络生成候选区域。为了使试验结果具有可比性,保持 3 种方法使用的候

选区域数量一致 ,都是 300 个。



图 4 基于 R-FCN 的飞机识别效果

表 2 在测试集上识别平均精度和速度

方法	准确率/(%)	召回率/(%)	mAP/(%)	检测时间/s
R-FCN+ResNet-50+OHEM	99.3	98.6	90.7	0.18
R-FCN+ResNet-101+OHEM	98.6	99.3	90.5	0.36
Faster R-CNN+ZF	82.9	96.5	90.1	0.08
Faster R-CNN+VGG	90.9	98.6	90.3	0.21
Faster R-CNN+ResNet-50	97.9	98.6	91.3	0.30
Faster R-CNN+ResNet-101	97.2	98.6	91.1	0.52
Fast R-CNN	88.8	88.7	88.4	2.42

从表 2 可以看出: ①Fast R-CNN 方法由于使用传统 Selective Search 方式确定候选区域 ,而过程采用 CPU 运算 ,因而速度远低于另外两种基于 GPU 运算的方法; ②Faster R-CNN+ZF/VGG16 由于网络结构简单(5 层 ,16 层) ,检测速度较快 ,但精度较低 ,出现了较多错检和漏检的情况; ③使用 ResNet-50(101) 网络结构的 Faster R-CNN 算法 ,检测精度得到了提高 ,但由于 Faster R-CNN 中 RoI 池化计算不共享 ,因而速度比同样网络结构的 R-FCN 方法慢; ④R-FCN 方法采用全卷积网络结构 ,RoI 池化层共享计算 ,检测精度高、速度快。综合考虑 ,R-FCN 方法目标检测效果优于 Faster R-CNN ,远优于 Fast R-CNN 方法。

4 结 语

本文探讨了基于全卷积网络的 R-FCN 算法在高分辨遥感影像目标检测中的有效性 ,并分析了不同参数和网络结构对 R-FCN 检测效果的影响 ,对比分析了 Fast R-CNN 算法、Faster R-CNN 算法和 R-FCN 算法在飞机识别中的效果。结果表明 ,R-FCN 方法比 Fast R-CNN 算法和 Faster R-CNN 识别精度更高 ,速度更快 ,识别准确率达 99.3% ,识别速度达到每张 180 ms ,基本满足高分辨遥感影像处理自动化实时处理需求。

参考文献:

[1] 刘扬 ,付征叶 ,郑逢斌.高分辨率遥感影像目标分类与

识别研究进展[J].地球信息科学学报 ,2015 ,17(9) : 1080-1091.
[2] DALAL N ,TRIGGS B.Histograms of Oriented Gradients for Human Detection [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05) . [S.l.]: IEEE ,2005.
[3] LOWE D G.Object Recognition from Local Scale-invariant Features [C] // Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra: IEEE ,1999.
[4] LOWE D G.Distinctive Image Features from Scale-invariant Keypoints [J] . International Journal of Computer Vision ,2004 ,60(2) : 91-110.
[5] HINTON G E ,SALAKHUTDINOV R R.Reducing the Dimensionality of Data with Neural Networks [J] . Science ,2006 ,313(5786) : 504-507.
[6] COLLOBERT R ,WESTON J.A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning [C] // Proceedings of the 25th International Conference on Machine Learning.Helsinki: ACM ,2008.
[7] BENGIO Y.Learning Deep Architectures for AI[J] . Foundations and Trends in Machine Learning ,2009 ,2 (1) : 1-127.
[8] MOHAMED A ,SAINATH T N ,DAHL G ,et al.Deep Belief Networks Using Discriminative Features for Phone Recognition [C] // IEEE International Conference on Acoustics ,Speech and Signal Processing (ICASSP) . Prague: IEEE ,2011.

- [9] CHAN T ,JIA K ,GAO S ,et al.PCANet: A Simple Deep Learning Baseline for Image Classification? [J].IEEE Transactions on Image Processing ,2015 ,24(12) : 5017-5032.
- [10] KAVUKCUOGLU K ,RANZATO M ,LECUN Y.Fast Inference in Sparse Coding Algorithms with Applications to Object Recognition [C] // Proceedings of OPT 2008. [S.l.]: [s.n.] 2008.
- [11] SILVER D ,HUANG A ,MADDISON C J ,et al.Mastering the Game of Go with Deep Neural Networks and Tree Search [J].Nature ,2016 ,529(7587) : 484-489.
- [12] ZHANG F ,DU B ,ZHANG L.Scene Classification via a Gradient Boosting Random Convolutional Network Framework [J].IEEE Transactions on Geoscience and Remote Sensing ,2016 ,54(3) : 1793-1802.
- [13] HU W ,HUANG Y ,WEI L ,et al.Deep Convolutional Neural Networks for Hyperspectral Image Classification [J].Journal of Sensors ,2015(2015) : 1-12.
- [14] MA X ,GENG J ,WANG H.Hyperspectral Image Classification via Contextual Deep Learning [J].EURASIP Journal on Image and Video Processing ,2015. <https://doi.org/10.1186/s13640-015-0071-8>.
- [15] CHEN X ,XIANG S ,LIU C ,et al.Aircraft Detection by Deep Belief Nets [C] // 2013 2nd IAPR Asian Conference on Pattern Recognition(ACPR 2013) .Naha: IEEE ,2013: 54-58.
- [16] CHEN X ,XIANG S ,LIU C ,et al.Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks [J].IEEE Geoscience and Remote Sensing Letters ,2014 ,11(10) : 1797-1801.
- [17] DIAO W ,SUN X ,DOU F ,et al.Object Recognition in Remote Sensing Images Using Sparse Deep Belief Networks [J].Remote Sensing Letters ,2015 ,6(10) : 745-754.
- [18] HAN J ,ZHANG D ,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.
- [19] 高常鑫 ,桑农.基于深度学习的高分辨率遥感影像目标检测 [J].测绘通报 ,2014(S1) : 108-111.
- [20] 王万国 ,田兵 ,刘越 ,等.基于 RCNN 的无人机巡检图像电力小部件识别研究 [J].地球信息科学学报 ,2017 ,19(2) : 256-263.
- [21] HE K ,ZHANG X ,REN S ,et al.Deep Residual Learning for Image Recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE ,2016.
- [22] SZEGEDY C ,LIU W ,JIA Y ,et al.Going Deeper with Convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE ,2015.
- [23] DAI J ,LI Y ,HE K ,et al.R-FCN: Object Detection via Region-based Fully Convolutional Networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Las Vegas: IEEE ,2016.
- [24] GIRSHICK R.Fast r-cnn [C] // Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE ,2005.
- [25] REN S ,HE K ,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.
- [26] SHRIVASTAVA A ,GUPTA A ,GIRSHICK R.Training Region-based Object Detectors with Online Hard Example Mining [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE ,2016.