

基于改进旋转区域生成网络的遥感图像目标检测

戴媛, 易本顺, 肖进胜*, 雷俊锋, 童乐, 程志钦

武汉大学电子信息学院, 湖北 武汉 430072

摘要 为了实现遥感图像中目标的快速准确检测, 解决遥感图像目标带有旋转角度的问题, 在卷积神经网络理论的基础上, 将旋转区域网络生成融入到 Faster R-CNN 网络中, 提出了一种基于 Faster R-CNN 改进的遥感图像目标检测方法。相对于主流目标检测方法, 本文算法针对遥感图像中的大多数目标都具有方向性不定且相对聚集的特点, 在区域候选网络中加入了旋转因子, 以便能够生成任意方向的候选区域; 同时, 在网络的全连接层之前增加一个卷积层, 以降低其特征图参数, 增强分类器的性能, 避免出现过拟合。将本文算法与几种主流目标检测方法进行对比分析后可知, 本文算法因融合了多尺度特征及旋转区域网络的卷积神经网络所提取的特征, 能得到更好的检测结果。

关键词 成像系统; 目标检测; 遥感图像; 深度学习; 旋转区域生成网络

中图分类号 TP183, TP751.1

文献标志码 A

doi: 10.3788/AOS202040.0111020

Object Detection of Remote Sensing Image Based on Improved Rotation Region Proposal Network

Dai Yuan, Yi Benshun, Xiao Jinsheng*, Lei Junfeng, Tong Le, Cheng Zhiqin

Electronic Information School, Wuhan University, Wuhan, Hubei 430072, China

Abstract In this study, the integration of the rotation region proposal network with Faster R-CNN network along with an improved remote sensing image object detection method based on the convolutional neural network is proposed. The aim is two-fold: 1) to realize rapid and precise detection of remote sensing image objects; 2) to address the problem caused by objects with rotated angle. Compared to the mainstream target detection methods, the proposed method introduces the rotation factor to the region proposal network and generates proposal regions with different directions, aiming at the characteristics of variable direction and relative aggregation of most targets in the remote sensing image. The addition of a convolution layer before the fully connected layer of the Faster R-CNN network has the advantages of reducing the feature parameters, enhancing the performance of classifiers, and avoiding over-fitting. Compared with the state-of-the-art object detection methods, the proposed algorithm is able to combine the features extracted by the convolutional neural network in the rotation region proposal network with the multi-scale features. Therefore, significant improvement in remote sensing image object detection can be achieved.

Key words imaging systems; object detection; remote sensing image; deep learning; rotation region proposal network

OCIS codes 110.2960; 100.3008; 150.4065

1 引 言

随着航天遥感技术的飞速发展, 遥感图像的分辨率不断提高, 基于遥感图像的目标检测在多个领域发挥着越来越重要的作用。目标检测^[1-2]主要包括预处理、特征提取、分类识别等阶段。根据提取图像特征方法的不同, 目标检测算法分为基于传统图

像处理与机器学习算法相结合的算法和基于深度卷积神经网络的算法。前者主要通过对方向梯度直方图等特征进行提取, 然后将提取的特征输入到支持向量机等分类器中学习分类规则, 最后利用训练完成的分类器实现目标的检测。然而, 这些特征都是人工设计的, 提取过程繁琐, 且属于浅层特征, 特征表达能力弱, 对于背景复杂多变的遥感图像目标的

收稿日期: 2019-07-29; 修回日期: 2019-09-02; 录用日期: 2019-09-19

基金项目: 国家重点研发计划(2016YFB0502602)、国家自然科学基金(61471272)

* E-mail: xiaojis@whu.edu.cn

检测效果欠佳。

近年来,深度学习^[3]有了新的进展,卷积神经网络^[4]被应用于目标检测。自从基于大型图像数据库 ImageNet^[5]的图像分类竞赛出现以来,各种深度学习检测算法^[6-8]相继被提出,这些算法主要分为 One-stage 和 Two-stage 两类。其中,2018 年提出的 YOLO v3^[7]是 One-stage 中的一种代表性算法,而 Two-stage 方法中的代表算法为 Faster R-CNN (Region-Convolutional Neural Network)^[8]等算法。R-CNN 系列算法已从 Faster R-CNN 发展到了 Mask R-CNN^[9],通过采用 RoIAlign,添加一个掩模分支,达到实例级别的检测分割。One-stage 算法在检测速度上优于 Two-stage 算法,但是在精度上则逊于后者。然而,这些主流的目标检测算法针对的是自然场景的图像,并未考虑目标的方向性等特征。

得益于深度学习在自然场景下对目标检测的巨大成功,大量基于深度学习的遥感图像目标检测^[10-12]研究不断涌现出来。遥感图像中待检测目标的一种重要特点是其方向及形状的不确定性,采用传统的水平框框选目标的方式无法准确地定位目标。而且,遥感图像中的目标大都密集分布,采用水平框的方式还可能会对检测后的非极大抑制(NMS)结果产生影响,导致检测效果变差。对于检测目标方向性多样的问题,已经有人提出了一些解决方法,但其中有很多是用于文本检测的,如:文献^[13]中提出了一种基于旋转区域的文本检测算法,该算法首先使用区域候选网络生成水平的候选框,再用多尺度池化后的特征来预测倾斜的文本框;文献^[14]中对 Faster R-CNN 网络中的区域生成网络

(Region Proposal Network, RPN)进行了改进,加入了旋转信息,使得 RPN 网络能够直接生成有角度的预测框来进行多方向性的文本检测。

综上所述,本文通过互联网收集了国内外公开的遥感图像数据集,对它们的特点进行对比分析后,从 Google Earth、资源卫星中心网站^[15]等数据源下载了遥感图像数据,然后采用人工标注建立了小型遥感图像的典型目标数据集。在卷积神经网络理论的基础上,本文提出了一个基于 Faster R-CNN 改进的遥感图像目标检测方法。针对遥感图像目标方向不定且分布聚集的特点,利用旋转区域生成网络取代区域生成网络,使网络能够生成任意角度的候选区域框,更加贴合目标的边界框,增大检测的准确率;同时,采用 RoIAlign 替代 RoIPooling,以减小池化过程中的误差;针对遥感图像尺度大的特点,修改了 Faster R-CNN 的分类网络,在全连接层之前添加一个卷积层,以减小生成特征图的维度,避免出现拟合现象,增强分类器的性能。将本文算法与 YOLO v2^[7]、YOLO v3^[7]等深度学习算法在 DOTA^[16]数据集上进行对比实验,实验结果表明本文设计的改进算法对遥感影像中目标的检测效果更好,可以实现多场景下多类遥感图像目标识别的检测任务。

2 Faster R-CNN 原理

考虑到遥感卫星图像检测的应用需求,本文采用检测精度较高的 Faster R-CNN^[8]深度检测框架作为算法原型。Faster R-CNN 算法的网络框架如图 1 所示。

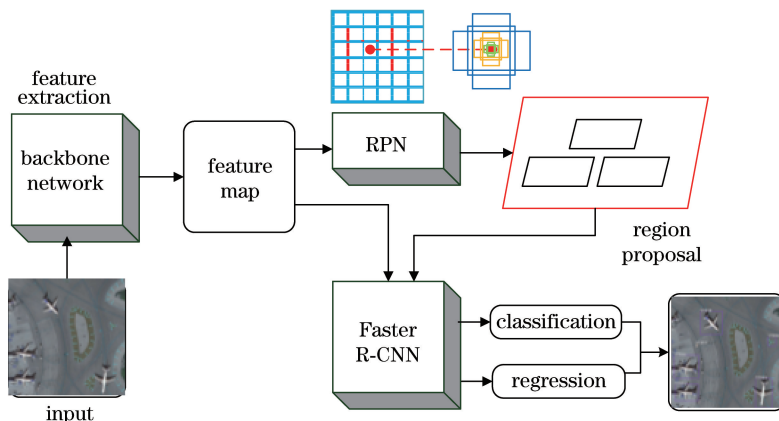


图 1 Faster R-CNN 网络模型结构

Fig. 1 Structure of Faster R-CNN network model

Faster R-CNN 使用基础网络进行特征提取,利用在图片分类任务上预训练好的基础网络来输

出中间层的特征图。其优势在于使用 RPN 产生候选区域框,RPN 使用卷积神经网络直接产生建

议区域,可以与分类网络共享特征提取网络,有效地减少了计算量,从而提高了整个目标检测过程的效率。

Faster R-CNN 网络包括基于区域的卷积神经网络 R-CNN。R-CNN 用全连接层为每个可能的目标类输出一个概率分数,R-CNN 将 RPN 输出的建议分为目标类和背景类,并对每个预测的目标类别的建议边框坐标进行调整,使输出的边框坐标更加准确。采用特征图对每个 RPN 生成的建议框的特征进行压缩,然后连接两个全连接层进行分类回归。

原始 Faster R-CNN 算法将 VGG16^[17] 网络的部分层作为浅层特征提取器,主要提取目标的点、边缘等基础特征,而且感兴趣区域作用在最后一层,从而影响了小目标的检测效果。遥感图像中的目标(比如船只、车辆等),具有方向性和密集分布的特点,此时,这类目标的检测与识别会受到方向的影响,因此采用普通的矩形框检测会导致检测框难以

精准地包围目标。

为解决以上问题,本文针对现有 Faster R-CNN 算法进行改进,主要包括三方面:1)重新设计特征提取网络;2)设计能产生旋转候选区域框的区域生成网络,使候选框更加贴合目标的边界框;3)对分类网络进行改进,以减小误差并减少网络的参数量。

3 基于改进旋转区域生成网络的目标检测方法

本文主要在以下三方面对现有 Faster R-CNN 算法进行改进:1)在基础网络阶段构建特征金字塔,并进行多尺度目标检测;2)将旋转区域生成网络(Rotation RPN, RRPN)取代 RPN,重新设计 Faster R-CNN 算法中的区域建议网络;3)用 RoIAlign 取代 RoIPooling,且在分类网络中添加一个卷积层。算法的整体框架如图 2 所示(cls 表示分类,loc 表示定位)。

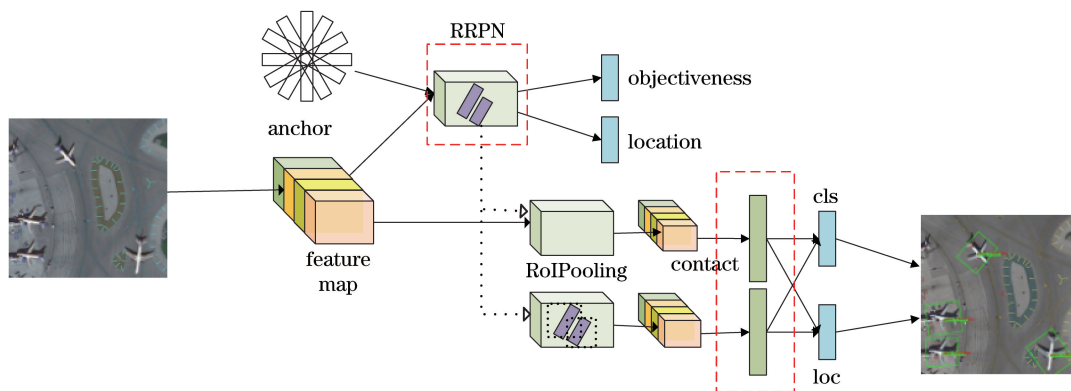


图 2 本文算法的网络结构

Fig. 2 Network structure of proposed algorithm

3.1 多尺度特征

遥感图像的分辨率过大,导致小目标很多,而 Faster R-CNN 算法使用 VGG Net 部分网络层作为浅层特征提取器,主要提取目标的点、边缘等基础特征,因此小目标的检测效果表现不佳。本文选择特征提取基础网络为 ResNet-101,同时加入了多尺度特征,以增强对小目标的检测能力。ResNet^[18] 的提出解决了传统网络加深时出现的退化问题,网络最深可达 152 层。与传统的网络相比,深度残差网络具有更好的泛化能力,同时还具有更低的复杂性。ResNet 相对于 VGG 网络更深,因此能够更好地学习到图像中的细节特征。ResNet 在卷积和池化层之间加入了批量归一层来加速训练,同时采用残差连接使得深度模型的训

练比较容易。残差网络引入了基于残差块的学习框架,输入可以通过跨层连接更快地向前传播。

Faster R-CNN^[8] 只根据基础网络最后一层的输出特征进行分类,计算量小,需要的内存少。然而,最后一层的特征属于高层特征,因此网络对于小尺度目标的特征表现力不足。一般说来,卷积神经网络的高层特征具有低分辨率、高层语义信息的特点;相反,低层特征具有高分辨率、低层语义信息的特点。把高层特征和低层特征结合起来,就能同时利用多尺度下的语义信息。根据文献[19]的思想,本文在特征提取部分加入了特征金字塔,以提升最终的检测效果,如图 3 所示。特征提取过程形成了自下而上的路径、自上而下的路径和横向连接。在形式上,对于宽度为 w 和高度为 h 的感兴

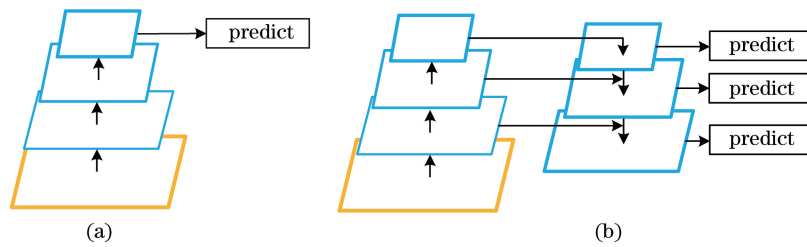


图 3 多尺度特征提取示意图。(a) Faster R-CNN 特征提取方法；(b)特征金字塔

Fig. 3 Diagram of multi-scale feature extraction. (a) Faster R-CNN feature extraction method; (b) feature pyramid

趣区域,假设 k_0 是基准值,代表特征图层数, k 为感兴趣区域对应的特征层级数,分配给特征金字塔的公式为

$$k = [k_0 + \text{lb}(\sqrt{wh}/224)]. \quad (1)$$

卷积神经网络的前馈计算是一条自下而上的路径,本文的基本网络使用每个残差块的特征激活输出,分别对应不同的卷积层输出,并具有不同的步长,这样能够很好地处理小目标。相比于只利用最

后一层卷积层提取的特征,这种结构可以利用更多的高层语义信息;对于小目标而言,在更大的特征图上面进行操作,增大了特征映射的分辨率,从而可以获得关于小目标的更有用的信息。

本文将 VGG16 网络和 ResNet101+FPN 网络作为基础网络,在 DOTA 数据集上进行实验,并通过验证集测试得到训练模型对典型目标的 AP (Average Precision)值,表 1 是测试结果。

表 1 不同基础网络对典型目标的提取结果

Table 1 Extraction results of different basic networks for typical object

Network	AP / %				
	Plane	Ship	Bridge	Harbor	Storage-tank
VGG16	79.2	42.1	18.5	43.1	44.5
ResNet101+FPN	82.1	44.3	21.6	45.3	47.4

由表 1 可以看出,使用 ResNet101+FPN 作为基础网络来提取特征,能够比较明显地提高检测准确率。此外,本文分别测试了两种方法的耗时。将 VGG16 作为基础网络时,检测 800 pixel×800 pixel 分辨率的图片的平均用时为 1.87 s,而将 ResNet101+FPN 作为基础网络时的平均用时为 2.14 s。ResNet101+FPN 有更多的层数和参数,因此增加了一定的计算量。

3.2 旋转区域生成网络

原始 Faster R-CNN^[8]中使用 9 个 anchor,目的是尽量生成大小及形状不同的包围框。对于自然场景下的目标,一般来说其形状及大小的变化不是很大,这些 anchor 足以包含大部分待检测目标。但是对于遥感图像,待检测目标的大小和形状变化很大,继续使用原始参数,不足以对不同形状的目标进行检测。因此,本文重新设计了旋转的 anchor,如图 4 所示。首先,如图 4(a)所示,针对部分尺寸较小的目标,重新设计了 4、8、16 和 32 的尺度;其次,如图 4(b)所示,由于遥感图像待检测的目标通常具有特殊比例,因此将纵横比更改为 1:1、1:2、1:3、1:4、

1:5、1:6、1:7 和 1:8,以覆盖各种遥感目标;然后,如图 4(c)所示,通过添加一个角度参数来控制 anchor 框的方向,分别对不同角度参数下的检测结果进行对比分析。角度过多会增加计算量,角度过少则会使角度误差偏大,从而影响平均精确度(mAP)。同时,考虑到角度计算表示的方便性,本文设置了 6 种不同的角度参数。角度间隔为 $\pi/12$,综合效果较好,在计算量和检测效果之间取得了较好的平衡。

由于算法在 RPN 阶段可以生成旋转建议框,因此,轴对齐的交并比(IoU)计算可能会导致偏差交互的 IoU 不准确,并使 RPN 的学习不准确,从而导致最后的检测结果不好。在常规的两个水平矩形框中,通过矩形框顶点的坐标能直接计算出相交矩形的顶点坐标。但如图 5 所示,两个带有角度的矩形框相交部分的形状是不定的。在本文中,计算倾斜 IoU 的方法如下:先计算出两个矩形的交点和一个矩形在另一个矩形内的顶点,然后计算两个矩形的并集,通过计算两个矩形的面积和,再减去每个矩形中非交集的部分,即图 5(a)、(b)中的黑色部分和阴影部分,就可计算 IoU 值。

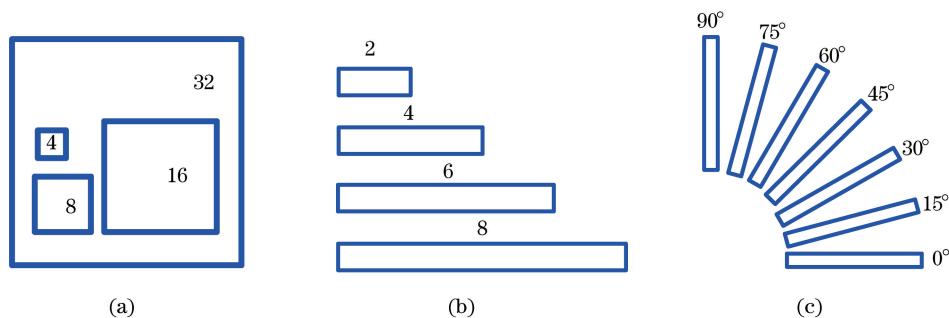


图 4 本文使用的 anchor 策略。(a)重新设计的尺度；(b)特殊比例；(c)角度参数

Fig. 4 Anchor strategy in our method. (a) Redesigned scale; (b) special ratio; (c) angular parameters

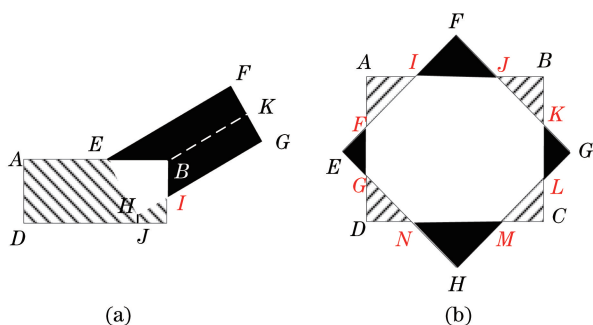


图 5 倾斜 IoU 的计算过程示意图。(a) 相交部分规则；(b) 相交部分不规则

Fig. 5 Calculation process of tilted IoU. (a) Regular intersection; (b) irregular intersection

NMS 被广泛用于目标检测,传统的 NMS 仅考虑 IoU(例如 IoU 阈值为 0.7),然而对带有角度的候选框直接进行 NMS 会影响检测。本文使用的 NMS

同时考虑了 IoU 和角度,由两个阶段组成:1)保持 IoU 大于 0.7 的框的最大 IoU;2)如果所有框的 IoU 均在 $[0.3, 0.7]$ 区间内,保持建议框与 ground truth 的最小角度差(角度差应小于 $\pi/12$)。对比轴对齐的边界框和倾斜的边界框时,可以在轴对齐的边界框上进行传统的 NMS,或者在倾斜的边界框上进行倾斜的 NMS。图 6 为执行不同 NMS 之后的检测结果。与图 6(b)相比,图 6(a)采用传统的 NMS 时漏检了 2 个距离比较近的目标中的 1 个。图 6(c)则显示了倾斜的 NMS 对倾斜目标检测更好的原因。从图 6 中可以看到:对于紧密相邻的倾斜目标,传统的 NMS 可能会遗漏一些待检测目标,因为轴对齐框之间的 IoU 可能很高;而倾斜的 NMS 能够检测到这个目标,倾斜的 IoU 值很低且角度差很小。

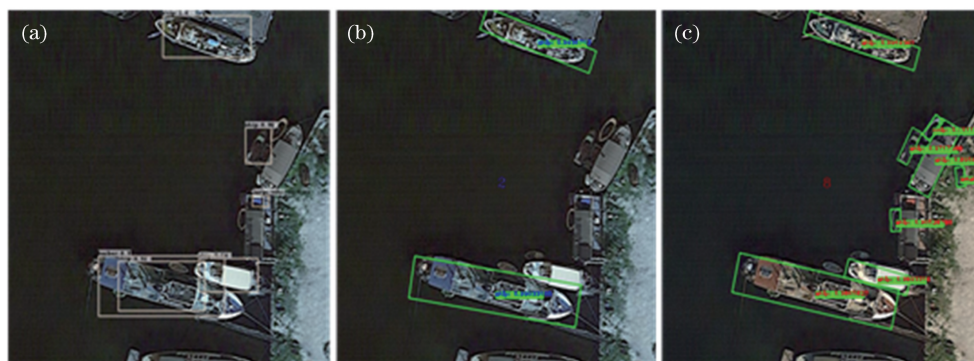


图 6 不同 NMS 检测结果的对比。(a)传统的 NMS 与普通框；(b)传统的 NMS 与旋转框；(c)倾斜的 NMS 与旋转框

Fig. 6 Comparison of results of different NMS detections. (a) Traditional NMS with common box; (b) traditional NMS with rotated box; (c) tilted NMS with rotated box

3.3 池化层及分类网络的改进

Fast R-CNN^[8]中提出的 RoIPooling 层能实现训练和测试的显著加速,并提高检测准确率。如图 7 所示,对不同大小的框,RoIPooling 也能得到固定大小的特征图^[20],图像通过卷积网络到达特征映射获得候选帧位置时以及 RoIPooling 对应每个

小网格的位置时存在浮点数舍入,这两次量化容易导致候选框位置偏离。首先要对 RoI 的连续坐标 (x_1, y_1) 和 (x_2, y_2) 进行取整量化,设坐标分量 x 的向下取整、向上取整分别表示为 $\text{floor}(x)$, $\text{ceil}(x)$, 然后在图 7(a)的实线框中,采用累加操作计算特征图上离散的特征值 $w_{i,j}$,进行最近邻采样操作;圆圈

代表像素点,框是 ROI 映射到特征图上的范围。由图 7(a)可以看出,量化操作将导致图像与特征图之

间不匹配。特征图上经量化后的感兴趣区域的池化结果 $r_{\text{pooling}}(x_1, y_1, x_2, y_2)$ 的计算公式为

$$r_{\text{pooling}}(x_1, y_1, x_2, y_2) = \frac{\sum_{i=\text{floor}(x_1)}^{\text{ceil}(x_2)} \sum_{j=\text{floor}(y_1)}^{\text{ceil}(y_2)} w_{i,j}}{(\text{ceil}(x_2) - \text{floor}(x_1) + 1) \times (\text{ceil}(y_2) - \text{floor}(y_1) + 1)} \quad (2)$$

Mask R-CNN 算法中提出 RoIAlign 池化方法,能有效减小 RoIPooling 量化操作中产生的误差。如图 7(b)所示, RoIAlign 不采取量化操作,采用双线性插值方法将图像上的像素值转换为浮点数,从而将整个特征聚合处理转换为连续操作。为了消除 RoIPooling 离散的量化错误,对每个连续点(图 7(b)中点 (a_i, b_i))都需要计算该点至 $N=4$ 的上、下、左、右 4 个离散点范围内的采用双线性插值操作得到的中心点 (a_i, b_i) ,即用(3)式对每个连续点 (a_i, b_i) 都操作 1 次。框是 ROI 区域在特征图上对应的区域,于是 ROI 区域的池化结果 $r_{\text{align}}(x_1, y_1, x_2, y_2)$ 可表示为

$$r_{\text{align}}(x_1, y_1, x_2, y_2) = \sum_{i=1}^N f(a_i, b_i) / N, \quad (3)$$

式中: $f(\cdot)$ 为表示特征图上的特征值; N 表示特征点的个数。

卷积神经网络全连接层的连接方式与卷积层、池化层不同,包含了大量的参数,而由 Lin 等^[21]的研究可知,全连接层容易导致过拟合,进而使网络的泛化能力降低。对分类网络进行改进的主要思想是降低全连接层的参数量,以减小计算量,同时防止过拟合。因此,本文对分类网络进行的修改如图 2 中分类网络处的虚线框所示,在全连接层之前添加一个卷积层来减少特征图的参数量,使分类器的性能

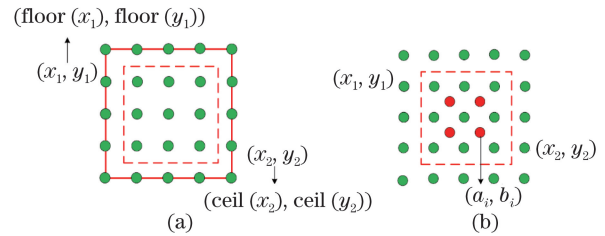


图 7 RoIPooling 和 RoIAlign。(a) RoIPooling; (b) RoIAlign

Fig. 7 RoIPooling and RoIAlign. (a) RoIPooling; (b) RoIAlign

更强大。本文使用 3×3 的卷积核,小卷积核不仅能实现大卷积核的作用,还可以减少参数量,加快计算速度。此外,该操作能防止融合特征的维度过大而导致出现的过拟合现象,同时可以减少 1/2 的特征尺寸,便于后续计算。

图 8 为改进方法与原始 RRPN 测试结果的对比,共有两组图。第一组图中包含两张检测结果图,主要是对车辆的检测;第二组图包含两张检测结果图,主要是针对船只的检测,分别为改进方法与原始方法的检测结果,可以看出改进后的测试结果更为准确。本文分别使用 RoIPooling 方法和 RoIAlign 方法在 DOTA 数据集上进行测试,并在验证集上测试了训练得到的模型对典型目标的 AP 值。测试结果如表 2 所示。

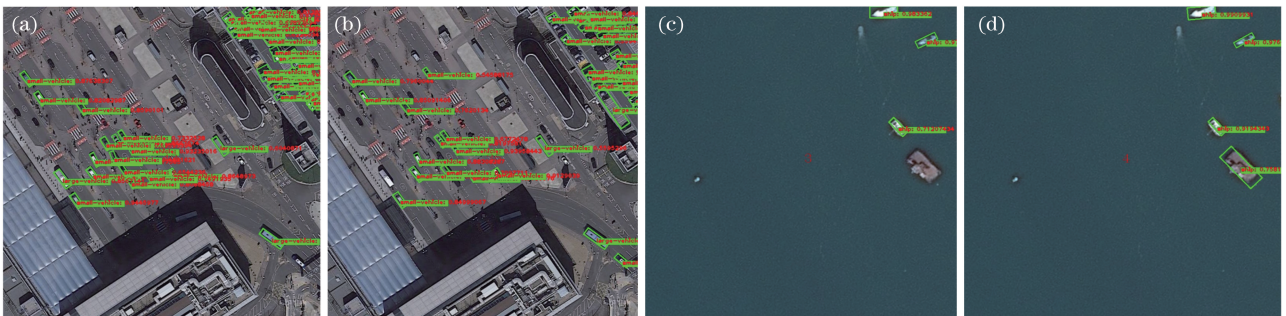


图 8 原始 RRPN 与改进方法的测试结果对比。(a)(c)原始 RRPN; (b)(d)改进方法

Fig. 8 Comparison of testing results between original RRPN and our method. (a)(c) Original RRPN; (b)(d) improved method

表 2 统计了两种池化方法对 5 类目标进行检测的结果,可以看出,与 RoIPooling 方法相比, RoIAlign

池化方法的检测准确率提高了 1~3 个百分点。因为遥感图像中的中小目标相对较多,且很密集,

表 2 不同 RoI 池化方法的检测效果对比
Table 2 Comparison of detection effects of different RoI pooling methods

Pooling method	AP / %				
	Bridge	Harbor	Storage-tank	Plane	Ship
RoIPooling	21.6	45.3	47.4	82.1	44.3
RoIAlign	23.9	47.0	48.5	83.8	47.4

RoIAlign 减小了池化过程中的像素偏差,所以与 RoIPooling 方法相比,检测率有一定提升。此外,本文还统计了两种方法的测试时间,使用多张 DOTA 数据集测试集中的图像进行测试,并计算了两种方法的平均检测时间。采用 RoIPooling 方法对分辨率为 $800 \text{ pixel} \times 800 \text{ pixel}$ 的图片进行检测,平均用时为 2.14 s,而使用 RoIAlign 方法进程检测的平均时间为 2.18 s,这说明算法改进之后不会对网络增加太多计算量。

使用原始分类网络和增加一个卷积层的分类网络在 DOTA 数据集上进行测试。本文在 DOTA 的训练集上进行实验,并通过验证集测试得到了 AP 值,如表 3 所示。

表 3 不同分类网络对 15 种目标的检测效果
Table 3 Detection results of different classification networks for 15 types of targets %

Category	AP	
	Original	+Conv
Bridge	23.9	27.5
Small-vehicle	31.6	32.4
Baseball diamond	67.6	67.3
Basketball court	47.5	46.3
Harbor	47.0	46.9
Ground-track field	40.2	44.6
Soccer ball field	41.2	42.4
Storage-tank	48.5	48.5
Large-vehicle	49.8	51.7
Plane	83.8	84.1
Roundabout	47.6	45.4
Tennis court	89.4	88.8
Helicopter	45.4	42.3
Ship	47.4	47.4
Swimming pool	39.8	38.1
mAP	50.05	50.25

从表 3 可以看出,改进的分类网络能够提高部分类别的检测准确率,比如桥梁、车辆、田径场、足球

场、飞机等目标。同时,本文分别测试了两种网络的耗时,使用多张测试集中的图像分别对不同的网络进行测试,并统计图像的检测时间。使用原始分类网络对 $800 \text{ pixel} \times 800 \text{ pixel}$ 分辨率的图片进行检测,平均用时为 2.14 s,而使用改进的分类网络进行检测的平均用时为 1.98 s。

本文针对现有 Faster R-CNN 算法作出如下改进:采用 ResNet 网络设计特征提取网络,引入 FPN 特征提取方法进行多尺度特征融合,替换原有的多尺度预测方法,用 RRPN 网络取代 RPN,用 RoIAlign 策略取代 RoIPooling 方法,并在分类网络中添加一个卷积层。

4 实验结果与分析

4.1 实验数据与平台

不同于自然图像,遥感图像中的目标大小不一,方向各异,背景环境复杂。本文收集、比较了网络上公开的多种遥感数据集,并分析了它们的优缺点。考虑到 DOTA^[16]数据集的数据质量较高,分辨率最高为 $4000 \text{ pixel} \times 4000 \text{ pixel}$,包含的类别多达 15 类,且样本较其他数据集更为均衡,尺度变化较大,因此将其作为实验的数据集。同时,从 Google Earth 和资源卫星中心网站上收集了部分数据,利用 roLabelImg 工具对这些数据进行了标注,为实验测试提供数据。样本集中的图片共计 2806 张,其中训练集中有 1411 张,验证集中有 458 张,测试集中有 937 张。本文选择的深度学习实验平台是 TensorFlow,利用 NVIDIA TITAN XP 进行训练测试。

本文采用 Faster R-CNN 的方式进行端到端的联合训练,为减少训练量,使用 ResNet 模型在 ImageNet 数据集中学习到的预训练参数。训练及测试过程中相关的网络参数设置如下:输入图像大小为 $800 \text{ pixel} \times 800 \text{ pixel}$,网络学习率为 0.0003。区域生成网络阶段的相关参数设置如下:IoU 正样本阈值设为 0.7,负样本阈值设为 0.3,以保持一定的正负样本比例。当 anchor 与 GT(ground truth)间的 IoU 大于 0.7 时,认为该 anchor 是正样本;当 IoU 小于 0.3 时,认为该 anchor 是负样本;对于 IoU 位于 2 个阈值之间的 anchor,即 IoU 在区间(0.3,0.7)内,认为背景和物体掺杂,对训练目标没有贡献,不参与训练;NMS 的 IoU 阈值设为 0.7;anchor 的设置参考 3.2 节。Fast R-CNN 检测分类的相关参数设置如下:Fast R-CNN 中 NMS 的 IoU 阈值设为 0.3,IoU 正样本阈值设为 0.5。

4.2 实验测试效果及不同算法的对比

本文利用改进算法的训练模型在 DOTA 数据集的测试集上进行相关测试,主要测试了桥梁、小型车辆、棒球场、篮球场、海港、田径场、足球场、储存罐、大型车辆、飞机、圆环、网球场、直升飞机、船只、游泳池等 15 类目标。表 4 是准确率(Precision)、召回率(Recall)及 AP(Average Precision)的测试指标。

表 4 为本文改进的算法在 DOTA 数据集上进行目标识别的结果,可见:飞机、网球场的识别效果较好,平均准确率超过了 80%;而桥梁的识别效果较差,平均准确率低于 30%。可能的原因是飞机、网球场具有显著的形状、颜色和纹理特征,且所处环境较为单一,识别起来相对容易;而桥梁一般处于地物比较密集的地区,长宽比极大,同时数据集中桥梁的数量也较少,因此识别难度较大。

为了说明本文改进算法的有效性,利用训练集训练了几种经典的深度学习目标检测算法,并对检测结果进行了对比。使用的模型分别为 YOLO v2^[7]、YOLO v3^[7]、Faster R-CNN^[7] 及原始的 RRPN 算法,结果对比如表 5 所示。算法 YOLO v2 和 YOLO v3 为 Darknet 框架下的开源代码,Faster R-CNN 为 TensorFlow 框架下的开源代码,参数基本参考开源代码中的默认参数,部分参数进行过适

当调整(由于参数量比较多,没有一一进行调整),选择检测效果较好的结果作为对比结果。

表 4 本文方法对 15 类目标的实验结果

Table 4 Experimental results of proposed method for 15 types of targets

Category	Precision	Recall	AP
Bridge	59.15	32.92	26.40
Small-vehicle	68.65	44.12	34.10
Baseball diamond	80.40	81.21	78.57
Basketball court	82.01	76.20	73.20
Harbor	77.95	61.41	56.00
Ground-track field	81.29	60.05	55.62
Soccer ball field	78.20	59.95	57.79
Storage-tank	81.40	52.82	51.50
Large-vehicle	62.68	76.20	56.91
Plane	94.10	87.36	86.50
Roundabout	75.52	60.67	56.05
Tennis court	97.04	91.29	91.16
Helicopter	82.05	65.31	61.88
Ship	74.01	55.35	50.10
Swimming pool	71.60	53.45	47.52
Average	77.74	63.89	58.89

表 5 不同算法对 15 类目标的实验结果

Table 5 Experimental results of different methods for 15 types of targets

Category	YOLO v2	YOLO v3	Faster R-CNN	RRPN	Proposed method
Bridge	14.18	10.03	41.82	23.88	26.38
Small-vehicle	13.08	14.79	3.85	34.65	34.15
Baseball diamond	52.79	9.09	72.83	67.61	78.57
Basketball court	42.43	2.27	55.81	47.48	73.21
Harbor	51.99	17.07	59.04	47.30	56.18
Ground-track field	32.57	4.81	84.68	40.19	55.64
Soccer ball field	31.67	0.146	63.60	41.15	57.78
Storage-tank	40.21	24.59	5.31	48.77	51.55
Large-vehicle	22.02	9.09	38.94	49.74	56.91
Plane	80.91	49.44	38.74	83.89	86.52
Roundabout	44.40	21.64	44.44	47.61	56.06
Tennis court	72.52	15.18	89.75	89.40	91.15
Helicopter	21.22	0.02	40.64	45.44	61.91
Ship	46.73	30.31	3.99	47.19	50.15
Swimming pool	34.31	7.54	22.71	39.78	47.55
mAP	39.87	14.40	44.41	50.08	58.91

表5表明,与YOLO v2、YOLO v3、Faster R-CNN等算法相比,改进后的算法在mAP上得到了较大提升。表5中的改进算法比YOLO v2(采取默认参数设置)检测得到的各类AP提高了3~40个百分点,尤其是对于车辆、球场等目标,AP显著提升,mAP提高了约19个百分点,表明了本文策略对于弥补YOLO v2缺陷的有效性。根据图9,YOLO v2得到的检测结果为普通矩形框,包含的背景部分较改进算法多。在表5中,除了桥梁、田径场和足球场,改进算法比Faster R-CNN(采取默认参数设置)的AP提高了1.4~46个百分点,mAP提高了约14.5个百分点。根据图10,原始的Faster R-CNN算法在检测密集目标,比如飞机时,检测效果不好,改进算法通过增加RPN中anchor的角度及尺度,对密集分布的目标具有更好的检测效果。与原始RRPN算法相比,改进RRPN对各类目标检测的

AP值有不同程度的提升,mAP约提高了8.8个百分点。根据图11所示的检测效果,针对网球场,改进算法在形状上更贴合目标。表5中YOLO v3算法的参数为默认设置,直接使用YOLO v3进行检测时效果不理想,尤其是车辆类别中的大小车辆,算法无法很好地进行检测,改进算法则在mAP得到了很大提高。同时根据图12,对比改进算法,YOLO v3对海洋背景下的船只进行检测时会出现漏检现象。

YOLO算法在目标检测上的AP分布极为不平衡,比如桥梁和小型车辆等目标,检测效果显著低于其他类别。同时,Faster R-CNN在对较小且分布密集的目标进行检测时,比如小型车辆、储存罐和船只等,检测效果远低于其他尺寸的中等目标,改进的算法则极大地优化了对这些类别目标的检测。

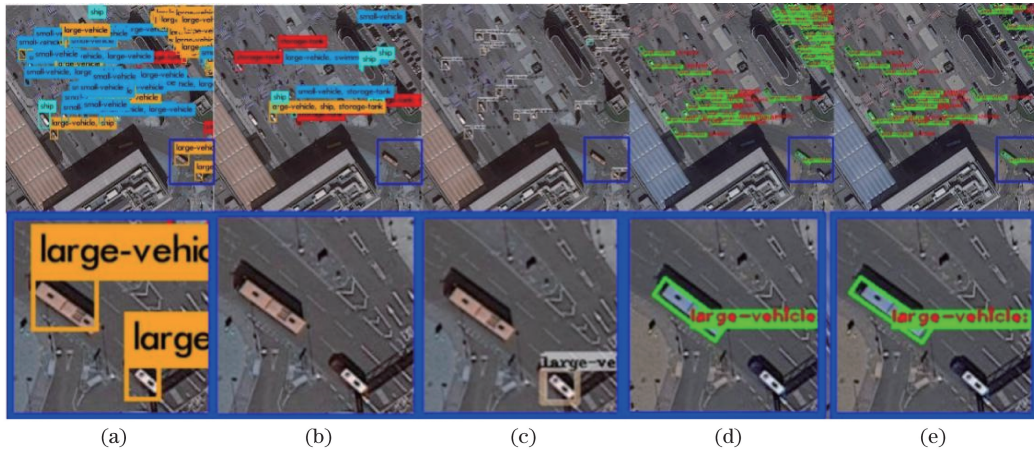


图9 不同算法对大型车辆的检测结果。(a) YOLO v2; (b) YOLO v3; (c) Faster R-CNN; (d) RRPN; (e)所提算法

Fig. 9 Detection results of different algorithms for large vehicle. (a) YOLO v2; (b) YOLO v3; (c) Faster R-CNN; (d) RRPN; (e) proposed algorithm

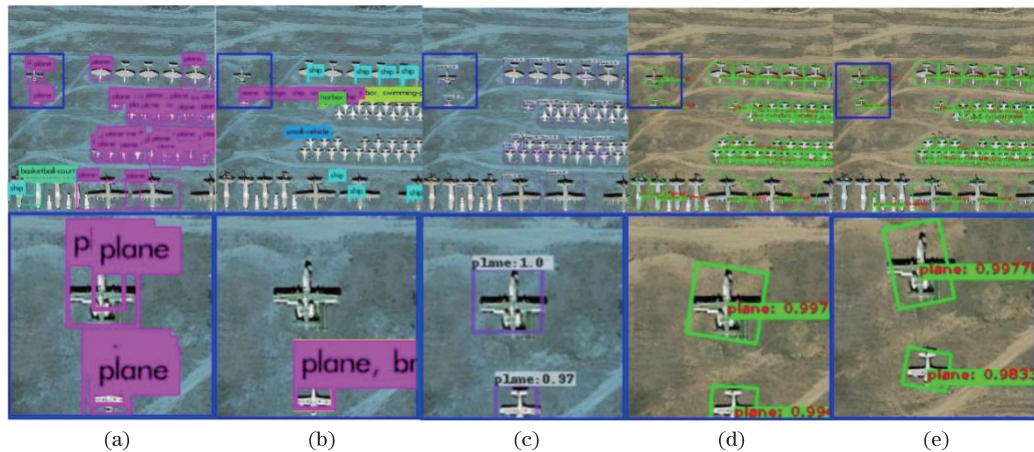


图10 不同算法对飞机的检测结果。(a) YOLO v2; (b) YOLO v3; (c) Faster R-CNN; (d) RRPN; (e)所提算法

Fig. 10 Detection results of different algorithms for airplane. (a) YOLO v2; (b) YOLO v3; (c) Faster R-CNN; (d) RRPN; (e) proposed algorithm

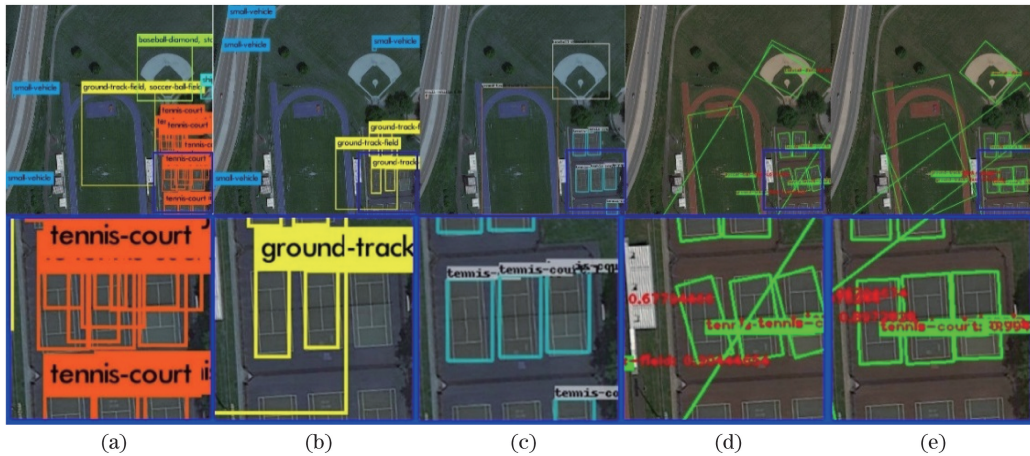


图 11 不同算法对网球场的检测结果。(a) YOLO v2; (b) YOLO v3; (c) Faster R-CNN; (d) RRPN; (e)所提算法
 Fig. 11 Detection results of different algorithms for tennis court. (a) YOLO v2; (b) YOLO v3;
 (c) Faster R-CNN; (d) RRPN; (e) proposed algorithm

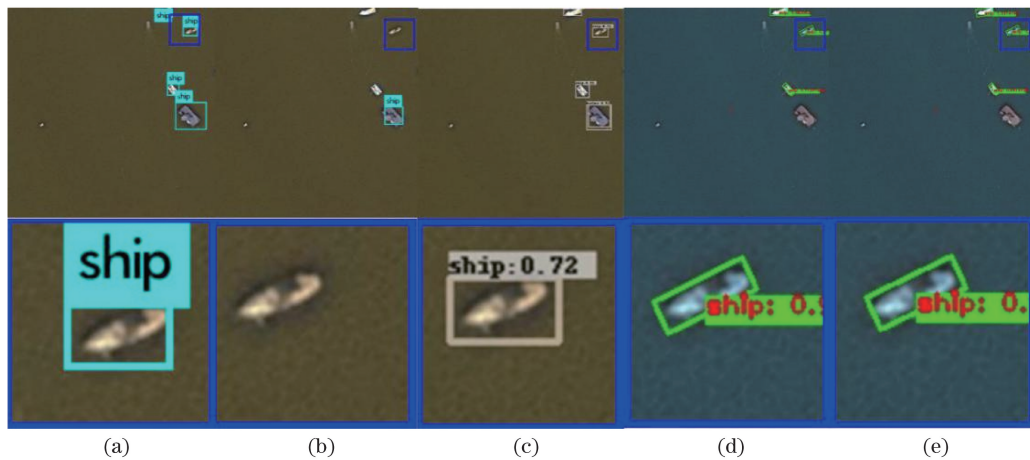


图 12 不同算法对船只的检测结果。(a) YOLO v2; (b) YOLO v3; (c) Faster R-CNN; (d) RRPN; (e)所提算法
 Fig. 12 Detection results of different algorithms for ship. (a) YOLO v2; (b) YOLO v3;
 (c) Faster R-CNN; (d) RRPN; (e) proposed algorithm

图 9~12 列出了 5 种不同算法对不同目标检测效果图的局部区域放大图,本文测试的目标有 15 类,可以看出本文方法对具有方向性的密集型目标(如车辆、飞机、网球场、船只等)的检测效果较其他算法更好。综上可知,本文改进的算法较 YOLO V2、YOLO V3、Faster R-CNN 算法在检测结果上有较大提升,尤其是在检测密集分布的目标和带有旋转角度的目标上,优势更为明显。实验结果表明,本文算法可以有效处理高分辨率遥感影像的目标检测问题。

5 结 论

随着遥感影像空间分辨率的不断提高,以及计算机软硬件技术的快速发展,卷积神经网络在遥感影像场景分类、目标识别、图像分割等领域得到了更

加广泛的应用。针对遥感图像中的典型海洋目标,着重研究了卷积神经网络用于目标检测的方法,并针对遥感图像的特点,改进了 Faster R-CNN 的网络结构,实现了对遥感图像典型目标检测精度的提升。针对遥感图像中大多数目标都是具有方向性不定且相互聚集的特点,在区域候选网络中加入了旋转因子,以生成任意方向的候选区域。在分类网络的全连接层前增加一个卷积层,以减少其特征参数,避免了过拟合现象的出现。实验对比结果表明,本文改进的算法能得到更好的检测结果。

参 考 文 献

- [1] Zou Z X, Shi Z W, Guo Y H, et al. Object detection in 20 years: a survey[J/OL]. (2019-05-16) [2019-07-28]. <https://arxiv.org/abs/1905.05055>.
- [2] Zhao Z Q, Zheng P, Xu S T, et al. Object detection

- with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.
- [3] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [4] Xiao J S, Tian H, Zou W T, et al. Stereo matching based on convolutional neural network[J]. Acta Optica Sinica, 2018, 38(8): 0815017.
肖进胜, 田红, 邹文涛, 等. 基于深度卷积神经网络的双目立体视觉匹配算法[J]. 光学学报, 2018, 38(8): 0815017.
- [5] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [6] 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.
- [7] Redmon J, Farhadi A. Yolov3: an incremental improvement[J/OL]. (2018-04-08) [2019-07-28]. <https://arxiv.org/abs/1804.02767>.
- [8] 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.
- [9] He K M, Gkioxari G, Dollar 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.
- [10] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved Faster R-CNN[J]. Acta Optica Sinica, 2018, 38(6): 0615004.
冯小雨, 梅卫, 胡大师. 基于改进 Faster R-CNN 的空中目标检测[J]. 光学学报, 2018, 38(6): 0615004.
- [11] 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.
- [12] Pei W, Xu Y M, Zhu Y Y, et al. The target detection method of aerial photography images with improved SSD[J]. Journal of Software, 2019, 30(3): 738-758.
裴伟, 许晏铭, 朱永英, 等. 改进的 SSD 航拍目标检测[J]. 软件学报, 2019, 30(3): 738-758.
- [13] Jiang Y Y, Zhu X Y, Wang X B, et al. R2CNN: rotational region CNN for orientation robust scene text detection [J/OL]. (2017-06-30) [2019-07-28]. <https://arxiv.org/abs/1706.09579>.
- [14] Ma J Q, Shao W Y, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [15] China national administration of GNSS and applications. China centre for resources satellite data and application provide services for monitoring earthquakes in Nyingchi prefecture in Tibet [EB/OL]. (2018-01-19) [2019-07-28]. <http://en.chinabeidou.gov.cn/c/785.html>.
- [16] 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.
- [17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. (2015-04-10) [2019-07-28]. <https://arxiv.org/abs/1409.1556>.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [19] Lin T Y, Dollar 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.
- [20] Jiang B R, Luo R X, Mao J Y, et al. Acquisition of localization confidence for accurate object detection [C]//Proceedings of the European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. New York: IEEE, 2018: 784-799.
- [21] Lin M, Chen Q, Yan S C. Network in network[J/OL]. (2014-03-04) [2019-07-28]. <https://arxiv.org/abs/1312.4400>.