

doi: 10.6046/gtzyyg.2020.04.10
引用格式: 卫虹宇,赵银娣,董霁红. 基于改进 RetinaNet 的冷却塔目标检测[J]. 国土资源遥感,2020,32(4):68-73. (Wei H Y,Zhao Y D,Dong J H. Cooling tower detection based on the improved RetinaNet [J]. Remote Sensing for Land and Resources ,2020, 32(4):68-73.)

基于改进 RetinaNet 的冷却塔目标检测

卫虹宇, 赵银娣, 董霁红

(中国矿业大学环境与测绘学院,徐州 221116)

摘要: 冷却塔排放容易造成大气污染,利用高分辨率遥感影像对冷却塔进行检测,可以为废气排放治理提供决策数据。针对传统算法在高分辨率遥感影像目标检测中检测精度不高、检测速度慢等问题,采用无采样机制改进 RetinaNet 目标检测框架从而提取冷却塔。首先,将数据集标注为工作中的冷却塔和非工作中的冷却塔;然后,根据数据集中目标类别数、训练中正样本的比例等特点对分类子网络最后一层的偏置项进行初始化并确定类别自适应阈值,此外,通过回归损失来设置分类损失的调整比例以避免损失函数被众多负样本所支配;最后,采用 ResNet50 提取图像特征,利用特征金字塔网络(feature pyramid networks,FPN)模块生成多尺度卷积特征金字塔,对每层特征进行检测框回归以及类别置信度计算。实验结果表明:对于高分辨率遥感影像冷却塔目标检测,该算法相比原始 RetinaNet 模型在保证检测速度的同时提高了检测精度,证明算法的有效性。

关键词: 卷积神经网络;目标检测;高分辨率遥感影像;无采样机制

中图法分类号: TP 751.1 **文献标志码:** A **文章编号:** 1001-070X(2020)04-0068-06

0 引言

冷却塔目标检测对大气污染治理有重要作用。冷却塔的排放物容易造成大气能见度恶化,需要对其位置和工作状态进行监测。随着遥感技术的发展以及遥感影像空间分辨率的提升,利用高分辨率遥感影像进行冷却塔目标检测,有利于监测特定区域内冷却塔的分布、预估冷却塔的排放量并及时防治区域频发的雾霾天气^[1-2]。

传统的目标检测算法,多是使用不同大小的滑动窗口在待检测图片上产生大量候选区域,对每一候选区域提取方向梯度直方图(histogram of oriented grids-ents,HOG)、尺度不变特征变换(scale-invariant feature transform,SIFT)等特征,最后利用训练完成的分类器进行分类。这种方法提取的特征泛化性不强,且滑动窗口计算开销大,在大量、多类目标检测中效率较低^[3]。随着计算机硬件的发展以及大数据时代的到来,卷积神经网络(convolutional neural network,CNN)迎来了发展的高峰。2012 年,Krizhevsky 等^[4]提出的 AlexNet 使用 CNN 大幅提高了图像分类的准

确率,随后,国内外学者针对扩增网络深度以及解决由此产生的过拟合、计算量大、网络性能退化等问题,提出了 VGG,Inception 和 ResNet 等深度卷积网络,在图像分类的应用中得到了很好的反响^[5]。

基于 CNN 在特征提取方面的优越性,有学者提出将其应用于目标检测中。然而,在识别图像中目标种类的同时,还需要用检测框准确标定目标在图像中的位置。为了解决这一问题,Girshick 等^[6]提出了区域卷积神经网络(regions with CNN,R-CNN),相比于传统的目标检测算法,检测精度和检测效率都得到了提高,但也存在占用的磁盘空间较大、训练时间成本大、训练步骤繁琐等问题^[7]。为此,学者们进行研究并相继提出 SPP-Net^[8],Fast R-CNN^[9]和 Faster R-CNN^[10]等目标检测方法。R-CNN 系列检测方法称为二阶目标检测法,需要先获取候选区域,再对每个候选区域进行分类,尽管在检测精度和检测速度上逐步完善,但是其计算量仍然很大。为了实现实时目标检测,学者们提出并发展了一阶目标检测法,如:YOLO^[11]和 SSD^[12]等,显著提高了检测速度,但在检测精度上难以超越二阶目

收稿日期: 2020-01-15; 修订日期: 2020-02-23
基金项目: 国家重点研发计划项目“大型煤电基地土地整治关键技术”(编号: 2016YFC0501105)、中央高校基本科研业务费专项资金项目“点云与序列影像融合的实景三维模型构建”(编号: 2015XKMS050)和自然资源部退化及未利用土地整治工程重点实验室开放基金课题“基于深度学习的城市高分遥感变化检测研究”(编号: SXDJ2019-4)共同资助。
第一作者: 卫虹宇(1996-),女,硕士研究生,主要研究方向为遥感目标检测。Email: 925074644@qq.com。
通信作者: 赵银娣(1980-),女,博士,副教授,主要研究方向为遥感数据处理、模式识别。Email: zhaoyd@cumt.edu.cn。

标检测法的检测水平。目标检测训练过程中正样本数量远小于负样本数量,二阶目标检测法会先区分前景和背景,并采用非极大值抑制算法抑制掉一部分负样本,减小了正负样本数量的差距,但检测速度也因此降低,而一阶目标检测法直接对类别极不平衡的样本进行细分类,容易导致分类器训练失败。文献[13]指出正负样本不平衡是一阶目标检测法在精度上低于二阶目标检测法的主要原因,提出了 RetinaNet 兼顾了检测精度和检测速度。

高分辨率遥感影像背景复杂且冷却塔的尺度、位置、形态多样,给目标检测带来挑战。本文以 RetinaNet 为冷却塔目标检测框架,由于原始 RetinaNet 采用 Focal Loss,通过计算所有样本的损失值并赋予不同的权重以解决正负样本不平衡的问题,但是需要较多的超参数调优。为解决该问题,引入文献[14]提出的无采样机制,通过使用初始化策略和调整分类损失权重来保持训练稳定性。

1 改进的 RetinaNet 冷却塔目标检测

1.1 RetinaNet 框架

RetinaNet 框架如图 1 所示。RetinaNet 由特征

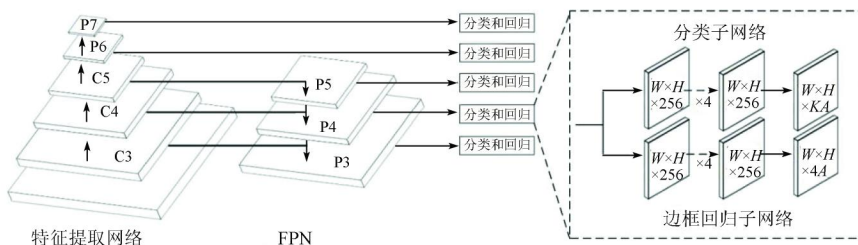


图 1 RetinaNet 框架

Fig. 1 RetinaNet structure

1.2 无采样机制

本文算法采用无采样机制,在解决训练中正负样本不平衡问题的同时避免较多的超参数调优,主要包括 3 个方案。

1.2.1 最优偏置初始化

在正负样本不平衡状态下,为了缓解梯度爆炸问题并加速模型的收敛,将分类子网络最后一层的偏置项 b 用先验概率 λ 初始化为:

$$b = -\ln \frac{1-\lambda}{\lambda} \quad (1)$$

在训练初期,样本的置信度分数接近初始化的先验概率 λ ,模型对所有样本的交叉熵损失函数 L_{CE} 为:

$$L_{CE} = -\ln \lambda - \left(\frac{N}{N_f} C - 1 \right) \ln(1 - \lambda) \quad (2)$$

式中: N 为总样本数; N_f 为正样本数; C 为类别数。

提取网络、特征金字塔网络 (feature pyramid networks, FPN)、分类子网络和边框回归子网络组成。使用 ResNet^[15] 作为特征提取网络,将卷积层 conv3, conv4 和 conv5 输出的特征图表示为 $\{C3, C4, C5\}$ 。在 ResNet 架构上利用 FPN 来生成丰富的多尺度卷积特征金字塔,金字塔包括 5 层特征映射 $\{P3, P4, \dots, P7\}$,其中 $P3-P5$ 是通过 FPN 中的自顶向下和横向连接对 ResNet 输出的 $C3-C5$ 进行计算得到的, $P6$ 是在 $C5$ 上进行卷积核为 3×3 ,步长为 2 的卷积运算得到的, $P7$ 是通过在 $P6$ 上应用 ReLU 以及卷积核为 3×3 ,步长为 2 的卷积运算得到的。将低层位置信息与高层语义信息融合,从而增加检测效果的鲁棒性。在金字塔的各层特征图上,对于每个锚点生成尺度为 $\{2^0, 2^{\frac{1}{3}}, 2^{\frac{2}{3}}\}$ 、长宽比为 $\{1:2, 1:1, 2:1\}$ 的 9 种候选框,即本文中 $A=9$,每个候选框包括 K 维独热编码的类别向量和 4 维坐标向量。将每层特征图的候选框输入到分类和边框回归子网络,分类子网络预测候选区域中 K 个目标类别的置信度,边框回归子网络计算候选框和真值框之间的偏移量,2 种网络结构基本相同,卷积层都使用 $W \times H$ 的卷积核,本文中为 3×3 ,子网络间不共享参数^[16]。

公式中第一项为前景,第二项为背景。

为了使 L_{CE} 最低,求 L_{CE} 的偏导数得到

$$\frac{\partial L_{CE}}{\partial \lambda} = -\frac{1}{\lambda} + \left(\frac{N}{N_f} C - 1 \right) \frac{1}{1 - \lambda} \quad (3)$$

当 $\lambda = \frac{N_f}{NC}$ 时, $\frac{\partial L_{CE}}{\partial \lambda} = 0$,此时 b 为最优偏置。

1.2.2 指导性损失函数

训练时负样本过多会导致模型倾向于将所有样本估计为负样本以获取极低的损失值。为了避免损失函数被众多负样本所支配,需要对分类损失和回归损失的权重进行调整。由于回归损失只计算正样本,不会受到正负样本不平衡的影响,所以可以通过回归损失来确定分类损失的调整比例 r ,即

$$r = \frac{\omega_{reg} L_{reg}}{L_{cls}} \quad (4)$$

式中： L_{reg} 和 L_{cls} 分别为回归损失和分类损失； ω_{reg} 为回归损失的权重系数。

则所有样本的总损失函数 L 为：

$$L = \omega_{\text{reg}} L_{\text{reg}} + \omega_{\text{cls}} L_{\text{cls}} r \quad , \quad (5)$$

式中 ω_{cls} 为分类损失的权重系数。

1.2.3 类别自适应阈值

为了进一步提升模型效果,缓解因正负样本不平衡而引起的置信转移问题,对每个类别采用不同的过滤阈值以保留更多的正样本,第 j 类的阈值 θ_j 为：

$$\theta_j = \frac{N_i}{N} \frac{N_j}{\sum_{j=1}^c N_j} \quad , \quad (6)$$

式中 N_j 为第 j 类样本数。

2 实验与分析

2.1 数据集

实验数据裁剪自谷歌影像,截取的图片均为双曲线自然通风冷却塔,使用开源图像标注软件 LabelImg 进行图像标注,标注方法与 Pascal VOC 数据集一致。数据集共 1 200 张图像,图像尺寸不一,如图 2 所示,检测目标包括工作中的冷却塔 (cooling tower working, CTWW) 和非工作中的冷却塔 (cooling tower resting, CTWR)。每张图像包含的目标数不一,数据集中共 1 450 个 CTWW 目标,1 344 个 CTWR 目标,目标角度、背景环境各不相同,给检测带来一定难度。



(a) CTWW 示例 (b) CTWR 示例

图 2 数据集示例

Fig. 2 Partial samples of dataset

2.2 模型训练

将数据集按 8:2 随机划分为训练验证集和测试集。实验基于 10 GB 内存的 I5 处理器,结合 NVIDIA GTX 1060 (6 GB 显存) 图形处理器 (graphics processing unit, GPU),在 Ubuntu16.04 系统上搭建 CUDA9.0 和 CUDNN7.5 的 GPU 运行环境,采用 Pytorch 深度学习框架训练模型。训练时先构建基于迁移学习的模型,使用在 ImageNet 数据集预训练的 ResNet50 模型对权重进行初始化。在正负样本匹配中,设置重叠率 (intersection over union, IoU) 阈值

生成正负训练样本,把 IoU 超过 0.7 的候选框记为正样本,小于 0.3 的记为负样本。在无采样机制的参数设置中, $\frac{N}{N_f} \approx 1\,000$, $C=2$,则最优 λ 约为 5×10^{-4} ,

ω_{reg} 和 ω_{cls} 都为 1,所有正样本中 CTWW 的比例为 52%,CTWR 的比例为 48%。采用随机梯度下降优化器进行参数优化,动量为 0.9,权值衰减系数为 0.000 1,初始学习率为 0.001,训练 50 个 epoch,每个 epoch 迭代 1 000 次,在 30 个 epoch 后学习率衰减为 0.000 1。

2.3 评价指标

本文采用平均准确率 (average precision, AP) 和平均准确率均值 (mean average precision, mAP) 作为冷却塔检测精度的评价指标。AP 是召回率在 0~1 之间的准确率均值,可以准确地衡量模型对单类目标的检测效果;mAP 是不同类间 AP 的均值,其数值越接近 1 说明模型在全类别中的检测性能越好。其中,准确率 P 和召回率 R 的定义分别为：

$$P = T_p / (T_p + F_p) \quad , \quad (7)$$

$$R = T_p / (T_p + F_N) \quad , \quad (8)$$

式中： T_p 为真正样本； F_p 为假正样本； F_N 为假负样本。

为了全面评价模型的性能,比较模型在测试集中检测单幅图像的平均检测时间来评估检测速度。

2.4 结果分析

利用本文算法在冷却塔测试集上进行检测,并与其他常用算法对比,包括传统目标检测算法的词袋模型 (bag of words, BoW) 和 HOG 算法以及基于 CNN 的 Faster R-CNN,SSD 和 RetinaNet 算法,实验结果如表 1 所示。

表 1 冷却塔目标检测结果对比

Tab. 1 Comparison of cooling tower object detection results

方法	AP/%		mAP/%	速度/ (s 张 ⁻¹)
	CTWW	CTWR		
BoW	26.51	21.12	23.82	22.402
HOG	58.51	50.46	54.49	17.180
Faster R-CNN	97.74	90.48	94.11	0.109
SSD	90.73	89.66	90.19	0.046
RetinaNet	97.85	92.77	95.31	0.054
本文算法	98.69	93.55	96.12	0.049

由表 1 可知,对于高分辨率遥感影像冷却塔目标检测,本文算法相比其他算法在保证实时检测的同时提高了检测精度,证明本文算法的有效性。相比 BoW 和 HOG,表 1 中基于 CNN 的算法的检测效果都具有明显优势。BoW 使用 SIFT^[17] 提取特征,由于冷却塔没有明显的角点,SIFT 无法提取有用的特征点,因而 BoW 检测精度较低;HOG^[18] 是在网格密集且大小统一的细胞单元内进行计算,并且采

用了局部对比度归一化,所以 HOG 的 mAP 值比 BoW 高了 1 倍多,但对于遥感影像中的复杂背景,2 种传统方法的特征提取能力都不及 CNN。与 Faster R-CNN 和 SSD 的检测效果进行比较,RetinaNet 兼顾了检测精度和检测速度,这是因为 RetinaNet 采用一阶目标检测法中直接在特征图上进行预测的策略,并使用 Focal Loss 解决训练中正负样本不平衡的问题,在保持一阶目标检测法的速度优势的同时大幅提高了检测精度。相比原始的 RetinaNet 算法,本文算法的 mAP 值提升了 0.81 百分点,检测速度提升了 0.005 s,说明无采样机制在提升 RetinaNet 对正负样本的调节性能的同时可以减少计算损耗。对比表 1 中各算法对 CTWW 和 CTWR 的检测精度,检测

CTWW 的 AP 值明显高于 CTWR,这是由数据集中各类样本数量不均衡导致的。

图 3 显示了本文算法在冷却塔测试集的部分正检(图 3(a)–(d))和误检(图 3(e)–(h))结果。对检测结果进行分析,遥感影像中冷却塔目标的环境复杂并且一幅影像中存在多个目标,对于在大小、工作状态和坐落方位上存在较大差异的冷却塔都能够有效检测出,说明本文算法能够在复杂环境的干扰下有效提取冷却塔特征。分析误检影像,本文算法容易将烟雾遮挡的地物误检为工作中的冷却塔,而具有圆环特征的较小目标易被误检为非工作中的冷却塔,说明背景的多样性和地物间的相似性对模型检测造成一定影响。

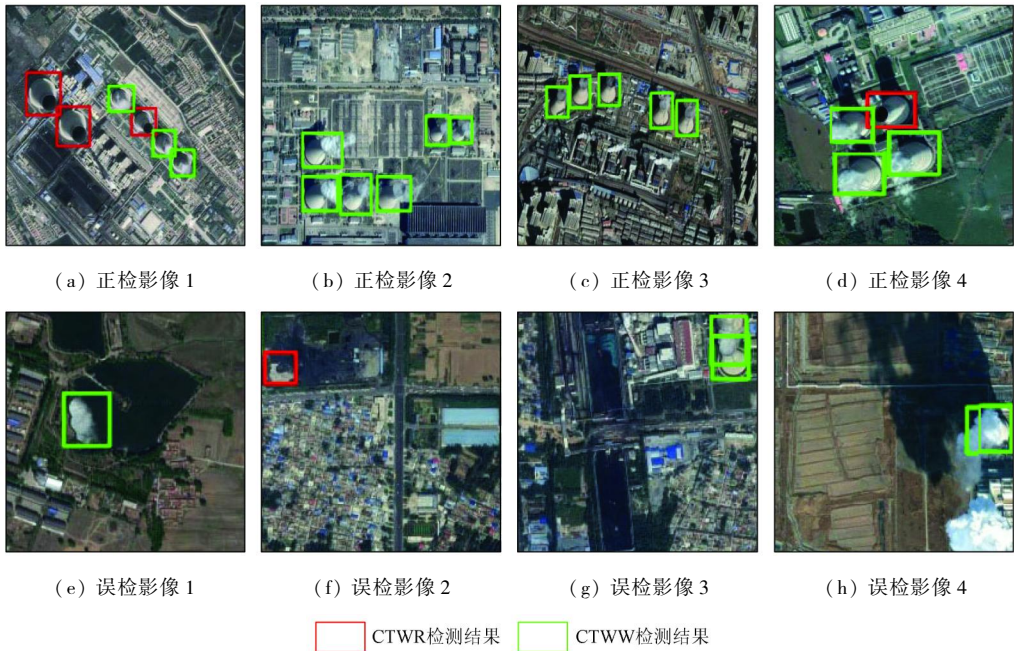


图 3 冷却塔测试集上的目标检测结果

Fig. 3 Object detection results on the cooling tower test set

实际应用中通常在较大范围的场景中进行冷却塔目标检测,为了验证本文算法的实用性,在较大图像分辨率的谷歌遥感影像上进行检测,结果如图 4 所示。可以看到,冷却塔目标在不同场景中所占的比例较小,受背景的干扰较大,在 3 个大型场景中

检测出 15 个目标,其中 12 个冷却塔目标被全部检出且分类正确,说明本文算法具有较强的定位能力和泛化能力,但无法避免相似地物的干扰,在场景 2(图 4(b))中出现 3 个误检目标。

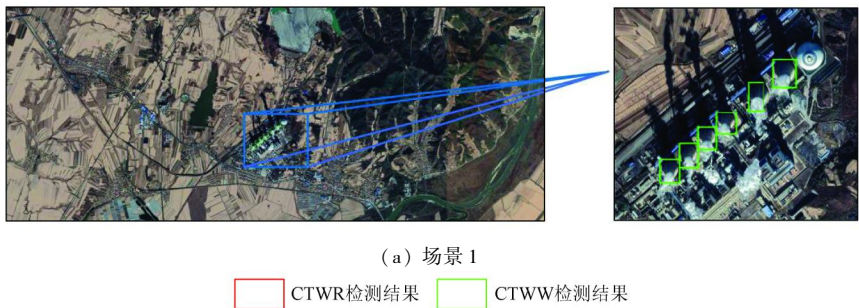


图 4-1 遥感影像冷却塔目标检测结果

Fig. 4-1 Results of cooling tower detection in remote sensing images

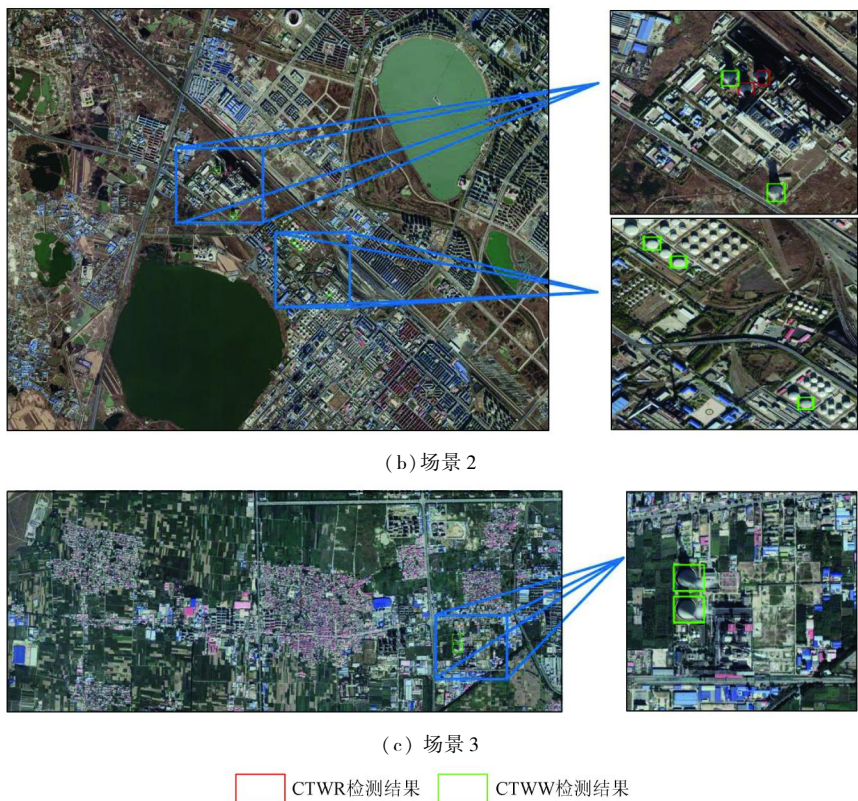


图 4-2 遥感影像冷却塔目标检测结果

Fig. 4-2 Results of cooling tower detection in remote sensing images

机,可基于本实验成果做进一步研究拓展。

3 结论

为解决传统算法在高分辨率遥感影像目标检测中检测精度不高、检测速度慢等问题,本文提出一种改进的 RetinaNet 目标检测方法。采用无采样机制替换原始 RetinaNet 中的 Focal Loss,避免了较多超参数的调节并且解决了训练中正负样本不平衡的问题。实验结果表明:

1)CNN 相比 BoW 和 HOG 等传统方法具有较强的特征提取能力,解决了传统方法提取的特征泛化性不强的问题。

2) RetinaNet 结合一阶目标检测法和二阶目标检测法的优势,在冷却塔数据集上取得了不错的检测效果。

3)对于高分辨率遥感影像冷却塔目标检测,本文算法相比其他算法在保证检测速度的同时提高了检测精度,并且在较大图像分辨率的遥感影像中检测效果良好,证明本文算法的有效性和实用性。

本研究还需要继续尝试新算法以优化模型,提高冷却塔检测精度;同时可以尝试样本优化,减小各类样本的数量差距,使模型在训练中充分学习目标特征,从而提升模型精度。此外,视频卫星^[19]技术的出现为区域内冷却塔实时监测提供了新的契

参考文献(References):

[1] 白永清,祁海霞,刘琳,等. 武汉大气能见度与 PM_{2.5} 浓度及相对湿度关系的非线性分析及能见度预报[J]. 气象学报,2016, 74(2):189-199.

Bai Y Q, Qi H X, Liu L, et al. Study on the nonlinear relationship among the visibility, PM_{2.5} concentration and relative humidity in Wuhan and the visibility prediction[J]. Acta Meteorologica Sinica, 2016, 74(2):189-199.

[2] 孙俊英,张璐,沈小静,等. 大气气溶胶散射吸湿增长特性研究进展[J]. 气象学报,2016, 74(5):672-682.

Sun J Y, Zhang L, Shen X J, et al. A review of the effects of relative humidity on aerosol scattering properties[J]. Acta Meteorologica Sinica, 2016, 74(5):672-682.

[3] 徐刚,岳继光,董延超,等. 深度卷积网络卫星图像水泥厂目标检测[J]. 中国图象图形学报,2019, 24(4):550-561.

Xu G, Yue J G, Dong Y C, et al. Cement plant detection on satellite images using deep convolution network[J]. Journal of Image and Graphics, 2019, 24(4):550-561.

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//26th Annual Conference on Neural Information Processing Systems, 2012:1097-1105.

[5] 谢奇芳,姚国清,张猛. 基于 Faster R-CNN 的高分辨率图像目标检测技术[J]. 国土资源遥感,2019, 31(2):38-43. doi: 10.6046/gtzyyg.2019.02.06.

Xie Q F, Yao G Q, Zhang M. Research on high resolution image object detection technology based on Faster R-CNN[J]. Remote

Sensing for Land and Resources,2019,31(2):38-43. doi:10.6046/gtzyyg.2019.02.06.

[6] Girshick R,Donahue J,Darrell T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.

[7] 回天,哈力旦·阿布都热依木,杜晗. 结合 Faster R-CNN 的多类型火焰检测[J]. 中国图象图形学报,2019,24(1):73-83.

Hui T, Halidan A, Du H. Multi-type flame detection combined with Faster R-CNN[J]. Journal of Image and Graphics,2019,24(1):73-83.

[8] He K,Zhang X,Ren S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(9):1904-1916.

[9] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV),2015:1440-1448.

[10] 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.

[11] Redmon J,Divvala S,Girshick R,et al. You only look once: Unified,real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:779-788.

[12] Liu W,Anguelov D,Erhan D,et al. SSD:Single shot multibox detector [C]//European Conference on Computer Vision. Berlin, Springer,2016:21-37.

[13] Lin T Y,Goyal P,Girshick R,et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision,2017:2980-2988.

[14] Chen J, Liu D, Xu T, et al. Is sampling heuristics necessary in training deep object detectors? [EB/OL] (2019-11-16). <http://arxiv.org/abs/1909.04868>.

[15] 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. IEEE,2016:770-778.

[16] 刘革,郑叶龙,赵美蓉. 基于 RetinaNet 改进的车辆信息检测[J]. 计算机应用,2020,40(3):854-858.

Liu G,Zheng Y L,Zhao M R. Vehicle information detection based on RetinaNet[J]. Journal of Computer Applications,2020,40(3):854-858.

[17] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision,2004,60(2):91-110.

[18] Dalal N,Triggs B. Histograms of oriented gradients for human detection [C]//IEEE Computer Society Conference on Computer Vision and Pattern. IEEE,2005.

[19] Zhang X,Xiang J,Zhang Y. Space object detection in video satellite images using motion information [J]. International Journal of Aerospace Engineering,2017,1024529:1-9.

Cooling tower detection based on the improved RetinaNet

WEI Hongyu, ZHAO Yindi, DONG Jihong

(School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Cooling tower emissions pollute the atmosphere. Using high-resolution remote sensing images to detect cooling towers can provide decision-making data for the treatment of exhaust emissions. Aiming at the problems such as low detection accuracy and slow detection speed of traditional algorithms in high-resolution remote sensing image object detection, the authors improved the RetinaNet by adopting a sampling-free mechanism to detect the cooling towers. First, Images in dataset were labeled as working cooling towers and resting cooling towers. Then, based on the number of object categories in the dataset and the proportion of positive samples in training, the bias term of the last layer in the classification subnetwork and class-adaptive threshold were determined. In addition, the regression loss was used to determine the adjustment ratio of the classification loss to avoid loss functions to be dominated by numerous background examples. Finally, ResNet50 was used to extract image features, and the FPN module was used to generate a multi-scale convolution feature pyramid. Detection boxes regression and category confidence calculations were performed for each layer of features. The results show that, for cooling tower detection on high-resolution remote sensing images, the proposed algorithm can improve the detection accuracy while ensuring the detection speed compared with RetinaNet, which proves the effectiveness of the proposed algorithm.

Keywords: convolutional neural network; object detection; high-resolution remote sensing image; sampling-free mechanism

(责任编辑: 张 仙)