

DOI:10.19651/j.cnki.emt.2107469

基于改进 Faster R-CNN 的小目标检测模型*

彭豪 李晓明

(太原科技大学 计算机科学与技术学院 太原 030024)

摘要: 针对工业大尺寸图像中小目标检测的平均精度均值低的问题,提出了一种改进的 Faster R-CNN-Tiny 模型。首先采用特征金字塔结构来对二阶检测器 Faster R-CNN 进行改进,来增强特征的表达能力,同时增加小目标特征映射分辨率,提高预测精度;其次将原本 ResNet 结构的最后一块改变为可变形卷积,自动计算各点的偏移,从最合适的地方取特征进行卷积,用以加强对小目标区域的特征提取;最后在提取感兴趣区域特征的时候,引入了内容的上下文信息,提高小目标检测的准确率。在工业中具有代表性的卫星遥感 UCAS-AOD 数据集以及天池瓷砖表面瑕疵质检数据集上进行对比试验。结果表明,改进后的 FRC-Tiny 模型相比原模型,其检测的平均精度均值分别提高了 5.57% 和 14.25%。

关键词: FRC-Tiny;特征金字塔;可变形卷积;工业小目标图像

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4

Small target detection model based on improved Faster R-CNN

Peng Hao Li Xiaoming

(School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: Aiming at the problem of low average precision of small target detection in industrial large-size images, an improved Faster R-CNN-Tiny model is proposed. Firstly, the feature pyramid structure is used to improve the second-order detector Faster R-CNN to enhance the feature expression capability and increase the resolution of small target feature mapping to improve the prediction accuracy. Secondly, the last piece of the original ResNet structure is changed to deformable convolution to automatically calculate the offset of each point and take features from the most suitable place for convolution, which is used to enhance the small target region. Finally, when extracting the features of the region of interest, the contextual information of the content is introduced to improve the accuracy of small target detection. The comparison tests are conducted on the representative satellite remote sensing UCAS-AOD dataset in industry and the quality inspection dataset of surface defects of tiles in Tianchi. The results show that the improved FRC-Tiny model improves the mean average precision of detection by 5.57% and 14.25%, respectively, compared with the original model.

Keywords: FRC-Tiny;feature pyramid networks;deformable convolutoin;industrial small target images

0 引言

近年来,大尺寸图像中的小目标检测成为了国内外学者的一大研究热点,在工业生产、卫星遥感、目标跟踪等领域发挥着重要作用^[1]。小目标检测^[2]一般是在大尺寸图像中精确定位小目标(一般尺寸小于 32 pixel×32 pixel 的目标)位置,例如寻找到遥感图像上的指定目标^[3](车、飞机等)^[4]、工业产品上瑕疵(瓷砖上的划痕、黑点等)的位置,然后标注出目标的类别。小目标检测的难点在于^[5]目标占原

图中的占比小,检测器无法提取充足且有效的特征,造成小目标检测结果不理想。其次,神经网络在学习中被大目标主导,小目标在整个学习过程被忽视,导致导致小目标的检测效果差,特别是网络层数多,小目标的特征信息将会丢失。

近年来,深度卷积神经网络在目标检测的多种方法中被广泛使用,根据对候选框的处理,将这类检测方法分为两类^[6]:1)基于回归的一段式目标检测方法,这类方法将整张图像作为输入,其目的是增加图像上目标的感受野,在图像

收稿日期:2021-08-03

* 基金项目:国家自然科学基金(61373099)项目资助

上不同位置回归出目标的位置及类别信息,最主要代表的方法为 YOLO^[7-9] 系列、SSD^[10] 等;2) 基于两段式的目标检测方法,先提取出图像上可能存在的目标候选框,再对每个区域候选框进行分类和位置回归,代表方法主要有 R-CNN^[11]、Fast-R-CNN^[12]、Faster-R-CNN^[13]、Mask R-CNN^[14] 等。第 1 类方法检测速度快,对大目标具有很好的适应性,但小目标容易漏检,第 2 种方法在小目标检测中精度相对高,但是特征提取,检测,分类各阶段单独工作,速度相对较慢,由于各阶段是分开的,可以单独进行改进优化,有很大提升空间。

最近,文献[10-17]提出了几种用于检测小目标的方法。文献[10]通过减小大对象的大小来扩充小对象数据,以克服数据不足的问题。除了数据扩充的方法之外,在不扩充数据集本身的情况下,也有一些增加所需特征信息的方法。文献[15]对所有特征图应用反卷积技术,以获得放大的特征图。然而,由于对所有特征图应用反卷积模块会增加模型复杂性和减慢检测速度的局限性。文献[16]通过结合不同尺度的特征信息和反卷积操作并获得了较高的精度

和速度。文献[17]使用文献[18]中的生成对抗网络以使用低分辨率特征作为 GAN 的输入来生成高分辨率特征。然而这些方法还有很大的改进空间,在本文中,建议使用上下文信息来帮助检测小目标。

本文提出的 FRC-Tiny 模型,在借鉴了 Faster-R-CNN 的基础上融合了其他比较好的方案,借鉴了 FPN^[19] 和 ResNet-101^[20]。引入了可变形卷积,使得在不同数据上检测更有优势,除此之外,引入了上下文信息,相比原模型,本文将整幅图像视为一个感兴趣区域(region of interest, ROI),通过 ROI 池化操作对图像全局特征进行提取,然后将候选框中的局部特征与全局特征进行相加,最后再进行分类和回归操作。在只增加 1 个 ROI 的特征提取和 1 个特征相加的操作的基础上可以有效提升预测准确率。

1 FRC-Tiny 小目标检测模型

1.1 模型结构

FRC-Tiny 模型采用了 Faster-R-CNN 作为基本框架,整体结构如图 1 所示,分为如下几个主要部分。

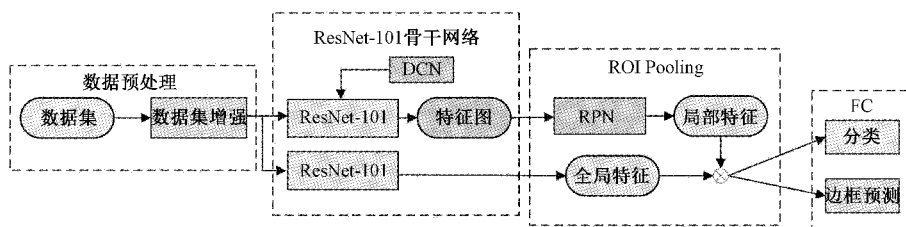


图 1 FRC-Tiny 模型结构体系

1) 数据预处理模块:对原始数据进行预处理与增强。

2) ResNet-101 骨干(Backbone)网络:主要由 ResNet-101 和可变形卷积构成,负责从输入的图像中提取特征,输出特征图集合。

3) ROI Pooling 部分:网络中引入了上下文(Context)信息,将全局特征与每一个候选框的局部特征相加,最后再进行分类和回归操作。

4) 全连接层:分类和边界框回归预测。

1.2 FRC-Tiny 模型实现

1) 数据增强

因原始图像分辨率大于 $1\ 000 \times 1\ 000$ 会增加计算开销,本文使用缩小两倍之后的图像作为网络的输入。选取 Resnet-101 作为主干网络,在提取特征的过程中,与输入图片相比,提取到的特征大小减小了 16 倍。原图到最后一层的卷积提取特征大小相比原图下降了 32 倍,这会严重地影响检测精度。在 FRC-Tiny 模型中,首先将图像切块,跟滑动切块不同,这里采取的是以检测目标为中心进行切块,切块大小为 $300\ \text{pixel} \times 300\ \text{pixel}$,图像不完整的用黑色填充,这样检测目标始终位于图像中心,具有很好的感受野,其次,对数据样本进行统计,对类别少的样本进行随机旋转,翻转,模糊操作,增加样本来提高网络对少量样

本类的识别率,提高鲁棒性。最后,采用 Mosaic^[21] 数据增强,将随机 4 张图片拼接成一张图片用于训练,相当于一下子训练四张图片,提高了训练速度,处理后的样本如图 2 所示。

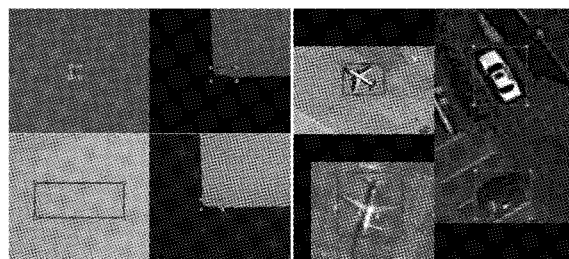


图 2 处理后的样本数据

2) ResNet-101 骨干(Backbone)网络

首先对图像低层的特征进行卷积核上采样操作,将高层特征中的语义信息和低层特征中的结构信息使用特征金字塔(feature pyramid networks, FPN)结构进行融合,用以增强特征的表达能力,除此之外,在特征金字塔中每一层中对候选框产生的位置进行充分的特征提取,有效提高小目标特征的映射分辨率,提高预测精度,此外,由于检测目标形状多样,传统的方形卷积核无法有效处理缺

陷,因此本文采用可变形卷积弥补这种缺陷,通过自动计算每个点的偏移量来完成对图像中最佳区域的特征提取。骨干网络如图 3 所示。

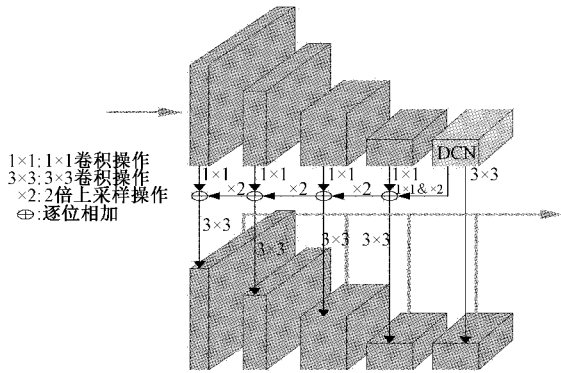


图 3 FPN 结构 ResNet-101 骨干网络

3) ROI Pooling 部分

Faster R-CNN 对候选框的操作有两步,第 1 步先生成候选框,第 2 步再对这些生成的候选框位置进行调整。当预测框的大小不是最佳时,可以将其向外或者向里进行调整。候选框的方向调整需要图像的全局特征信息来辅助,因此图像的全局特征信息对候选框的位置调整有重要的作用。本文将整幅图像视为一个 ROI,通过 ROI Pooling 操作对图像全局特征进行提取,然后将候选框中的局部特征与全局特征进行相加,最后再进行分类和回归操作。在只增加 1 个 ROI 的特征提取和 1 个特征相加的操作的基础上可以有效提升预测准确率。具体过程如图 4 所示。

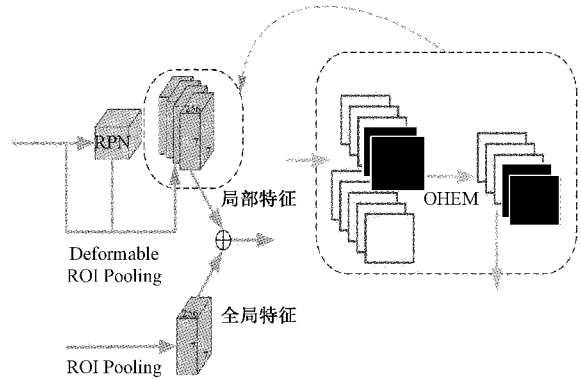


图 4 ROI Pooling 部分示意图

图 4 中,有黑色填充的方形框表示图片中提取出的带检测目标的样本,无黑色填充的方形框表示图片中没有检测目标的背景样本。数据集里会包含无目标样本和无缺陷样本,在训练过程中,采用一种策略,每次会随机选择 1 张无缺陷样本和 1 张缺陷样本,然后在缺陷图片中选择训练的正样本,在两张图片中选择负样本,两张图片的 ROI 合并进行 OHEM^[22] (online hard example mining) 选择正样本,之后再训练操作。这样可以有效利用无缺陷样本用以提高模型对背景信息的判别能力。

4) 全连接层

经过两个全连接层后,输出预测类别以及预测框。FRC-Tiny 模型结构如图 5 所示。

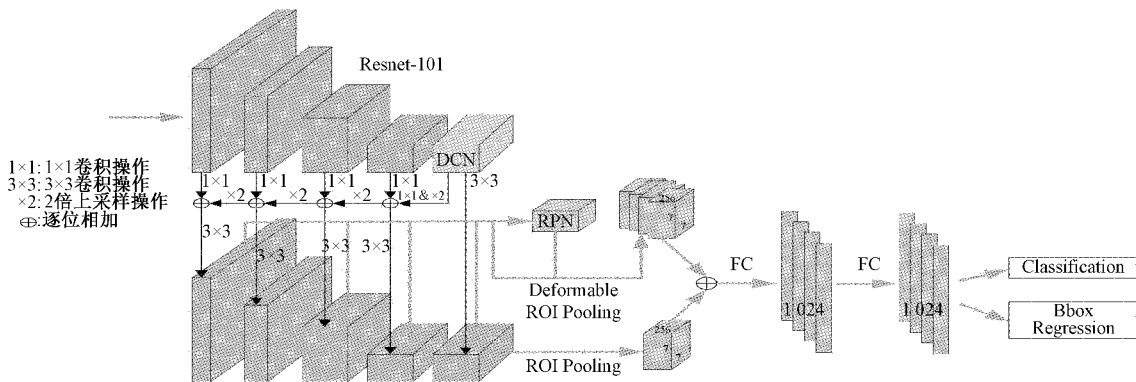


图 5 FRC-Tiny 模型结构

2 实验结果与分析

2.1 数据集及实验环境

本文采用的遥感数据集来自自由国科大于 2014 年发布的卫星遥感 UCAS-AOD^[23],其包含汽车 (Car) 和飞机 (Aircraft) 两类目标以及背景负样本,飞机数据集包括 600 张图像和 3 210 架飞机,而车辆数据集包括 310 张图像和 2 819 辆车,图像尺寸大约为 1 280 pixel×680 pixel。T.

业图片数据集来自于天池 2021 年广东工业智造创新大赛“瓷砖表面瑕疵质检”初赛数据集,包括 5 388 张带标记的训练集,1 762 张测试集,图像尺寸约为 8 192 pixel×6 000 pixel,有 0~5 共 6 类不同的样本标签,每张图的 bbox 的数量大部分集中在 1~5 个。本文的实验环境为 ubuntu18.04+ cudnn10.1+ cudnn9.0+ opencv3, GPU 为 NVIDIA GeForce GTX 2070 8 GB, CPU 为 Intel (R) Core(TM) i5-2500。

2.2 网络训练

本文训练图像数据集大小设置为 600×600 , 为了增强模型的泛化能力, 除了使用 Mosaic^[24-25]、旋转、翻转操作外, 还改变图像饱和度和色调。为了减少冗余, 采取非极大值抑制^[26] (NMS), 将 NMS 的 IoU 阈值固定为 0.65, 使用 2 600 个 RPN 提议对模型进行训练。统一将数据集转化为 VOC 格式, 方便训练。

2.3 评价指标

平均精度均值 (mean average precision, mAP) 和平均精确率 (average precision, AP) 常作为目标检测中常用的模型效果的评价指标, 平均精度是对目标检测的精确率和召回率的综合评价, 具有很好的衡量性。假设只有两类样本, TP 为正类样本被预测为正类, FP 负类样本被预测为正类, TN 负类样本被预测为负类, FN 正类样本被预测为负类。

1) 精确率为正类被预测为正类的图片占总的正类预测个数的比例。精确率计算公式如下:

$$precision = \frac{TP}{TP + FP} \quad (1)$$

2) 召回率为正类被预测为正类的图片占有标注图片的个数。召回率计算公式如下:

$$recall = \frac{TP}{TP + FN} \quad (2)$$

3) 平均精确率为精确率 (precision) 随着召回率 (recall) 变化的曲线在 recall 从 0~1 上的积分。AP 是计算单类别的模型平均准确度。通常来说, 平均精度越高, 代表该算法对该类目标的检测性能越好。

4) 均值平均精度将各类目标检测平均精度求均值, mAP 越大, 检测器检测效果越好。

2.4 实验结果及分析

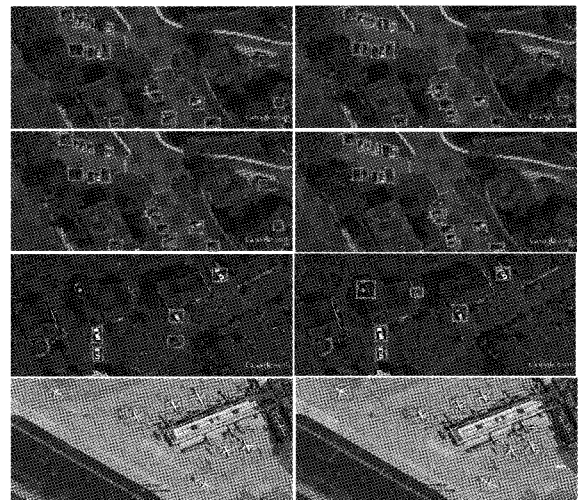
为了直观地体现模型各部分的作用, 从原始 Faster-R-CNN 开始, 逐个模块添加, 对比检测结果。结果如表 1 所示。FRC-Tiny 模型在 UCAS-AOD 与瓷砖瑕疵质检数据集上均获得了最高的 mAP, 分别为 76.38% 和 69.90%, 比起原始的 Faster-R-CNN 分别提高了 5.57% 和 14.25%。在 UCAS-AOD 预测过程中, 由于测试图像像素不算太大, 图像中目标密集程度较高, 使用的原图进行预测, 数据处理提升 mAP 为 2.82%。在瓷砖瑕疵质检数据集中, 测试图片大, 检测的瑕疵目标小, 且分布稀疏, 采用滑窗切割将测试图片切割成像素的小图进行测试, 其中每个滑窗重叠部分设置为 20 pixel, 测试完后将每个小图还原成原始图像, 采用 NMS 将有重复的标注框去

表 1 不同模块对检测精度的影响

组成部分	Faster-R-CNN	数据处理	FPN	Normal Sample	Deformable	Context
数据处理		✓	✓	✓	✓	✓
+FPN			✓	✓	✓	✓
+Normal Sample				✓	✓	✓
+Deformable					✓	✓
+Context						✓
UCAS-AOD/mAP	70.61	73.43	74.92	75.39	75.96	76.38
瓷砖瑕疵质检/mAP	55.65	60.72	63.09	65.27	66.31	69.90

除。因此, 数据处理对瓷砖瑕疵质检数据的 mAP 提升较大, 提高了 5.07%。

增加无目标的样本图, 增强了特征图的语义特征和模型对背景的判别能力, 能尽可能避免误检, 使 mAP 分别提高了 0.47% 和 2.18%。可变形卷积使得在卷积计算的过程中能够自动地计算每个点的偏移, 从而从最合适的地方取特征进行卷积, 让卷积的区域尽可能地集中在缺陷上, 让网络尽量避免小目标的漏检, 使 mAP 分别提高了 0.57% 和 1.04%。Context 上下文信息, 把整张图片的信息送给这个候选框当特征, 提高候选框的精度, 使 mAP 分别提高了 0.42% 和 3.59%。图 6 与 7 分别展示 Faster-R-CNN 与 FRC-Tiny 在两个数据集上的检测效果。



(a)Faster-R-CNN检测效果

(b)FRC-Tiny检测效果

图 6 在 UCAS-AOD 数据集上的检测效果比较

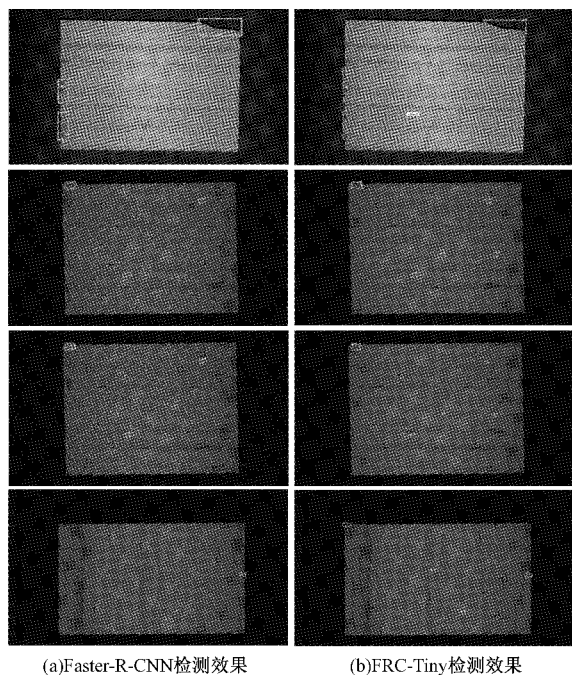


图 7 在天池瓷砖表面瑕疵质检数据集上的检测效果比较

3 结 论

本文针对工业大尺寸图像中小目标检测效果较差的情况,根据目标数据进行预处理,采用特征金字塔(FPN)结构来对二阶检测器 Faster R-CNN 进行改进,增强特征的表达能力,增加小目标特征映射分辨率,提高预测精度,并且将 ResNet 结构的最后一个 block 改变为可变形卷积来让卷积特征集中在小目标上,同时引入背景上下文信息,增加感受野,极大地改善了小目标漏检误检的情况。实验表明,本文提出的 FRC-Tiny 模型在 UCAS-AOD 数据集上比基准模型 mAP 提高 5.57%,在瓷砖瑕疵质检数据集上提高 14.25%。虽然 FRC-Tiny 模型在工业大尺寸图片的小目标检测中相比 Faster R-CNN 在平均精确率上有很大提升,但由于工业中不同数据环境存在差异,同时能够兼顾不同数据上的效率和精度是后续的研究方向。

参考文献

- [1] 姬晓飞,石宇辰,王昱,等. D-S 理论多分类器融合的光学遥感图像多目标识别[J]. 电子测量与仪器学报, 2020, 34(5):127-132.
- [2] BOSQUET B, MUCIENTES M, BREA V M. STDnet: A convnet for small target detection[C]. The British Machine Vision Conference (BMVC), Newcastle, United Kingdom, 2018: 253.
- [3] 王明阳,王江涛,刘琛. 基于关键点的遥感图像旋转目标检测[J]. 电子测量与仪器学报, 2021, 35(6):102-108.
- [4] 蔡彪,沈宽,付金磊,等. 基于 Mask R-CNN 的铸件 X 射线 DR 图像缺陷检测研究[J]. 仪器仪表学报, 2020, 41(3):61-69.
- [5] 解尧婷,张丕状. 基于改进的 YOLOv4 输电线路小目标检测[J]. 国外电子测量技术, 2021, 40(2):47-51.
- [6] KIM C, LEE Y, PARK J I, et al. Diminishing unwanted objects based on object detection using deep learning and image inpainting[C]. 2018 International Workshop on Advanced Image Technology(IWAIT), 2018: 1-3.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [9] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21-37.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [12] GIRSHICK R. Fast r-cnn[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [13] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [14] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [15] FU C Y, LIU W, RANGA A, et al. Dssd: Deconvolutional single shot detector [J]. ArXiv Preprint, 2017, ArXiv:1701.06659.
- [16] JEONG J, PARK H, KWAK N. Enhancement of SSD by concatenating feature maps for object detection[J]. ArXiv Preprint, 2017, ArXiv:1705.09587.
- [17] LI J, LIANG X, WEI Y, et al. Perceptual generative adversarial networks for small object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1222-1230.

- [18] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [J]. Advances in Neural Information Processing Systems, 2014, 27.
- [19] ADELSON E H, ANDERSON C H, BERGEN J R, et al. Pyramid methods in image processing[J]. RCA Engineer, 1984, 29(6): 33-41.
- [20] 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, 2016: 770-778.
- [21] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint, 2020, ArXiv:2004.10934.
- [22] 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, 2016: 761-769.
- [23] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network [C]. 2015 IEEE International Conference on Image Processing (ICIP), 2015: 3735-3739.
- [24] 化嫣然, 张卓, 龙赛, 等. 基于改进 YOLO 算法的遥感图像目标检测[J]. 电子测量技术, 2020, 43(24): 87-92.
- [25] 谢梦, 刘伟, 杨梦圆, 等. 深度卷积神经网络支持下的遥感影像飞机检测[J]. 测绘通报, 2019(6): 19-23.
- [26] NEUBECK A, VAN G L. Efficient non-maximum suppression [C]. 18th International Conference on Pattern Recognition(ICPR'06), 2006: 850-855.

作者简介

彭豪, 硕士研究生, 主要研究方向为图像处理、目标检测识别。

E-mail: 965534838@qq.com

李晓明, 博士, 教授, 主要研究方向为图像处理与分析、计算机视觉。

E-mail: lixmming@126.com