

DOI:10.3876/j.issn.1000-1980.2022.04.016

基于 Mask R-CNN 的防波堤复杂护面块体检测和分割方法

高林春^{1,2},王收军³,陈松贵²,赵旭²,陈汉宝²

(1. 天津理工大学计算机科学与工程学院,天津 300384;
2. 交通运输部天津水运工程科学研究院港口水工建筑技术国家工程实验室,天津 300456;
3. 天津理工大学机电工程国家级实验教学示范中心,天津 300384)

摘要: 针对斜坡式防波堤护面层块体个数统计效率和精确率低的问题,提出了基于 Mask R-CNN 深度学习网络的斜坡式防波堤扭王字块体的识别和分割方法。该方法利用 Mask R-CNN 深度学习网络学习实验室采集图像的特征信息,通过调整交并比(IOU)阈值得到评价指标最好的模型,并将该模型应用于现场防波堤图像护面块体的识别和分割。测试结果表明,IOU 取 0.5 时,目标分割的平均精确率为 91.83%,平均召回率为 92.94%;将训练得到的网络用于识别无人机航拍现场的防波堤图像,扭王字块识别率可达 90.7%,且拍摄角度和高度对识别精度影响不大。Mask R-CNN 深度学习网络可实现密集、复杂护面块体的准确识别,具有良好的移植性和通用性。

关键词: Mask R-CNN;防波堤;块体识别;扭王字块

中图分类号:TU122 文献标志码:A 文章编号:1000-1980(2022)04-0121-06

Identification and segmentation technology of complex armour blocks of rubble mound breakwater based on Mask R-CNN

GAO Linchun^{1,2}, WANG Shoujun³, CHEN Songgui², ZHAO Xu², CHEN Hanbao²

(1. School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China;
2. Tianjin Research Institute for Transport Engineering, National Engineering Laboratory for Port Hydraulic Construction Technology, Tianjin 300456, China;
3. National Demonstration Center for Experimental Mechanical and electrical engineering Education, Tianjin University of Technology, Tianjin 300384, China)

Abstract: In order to address the low statistical efficiency and accuracy of the number of armour unit blocks of slope breakwater, a method for the recognition and segmentation of the slope breakwater accropodes based the Mask R-CNN deep learning network was proposed. Firstly, this method used the Mask R-CNN network learning laboratory to collect the feature information of the image. Secondly, the model with the best performance evaluation index was obtained by adjusting the IOU threshold. Finally, the trained Mask R-CNN network was applied in the recognition and segmentation of the armour blocks of the on-site breakwater image. The test results show that when the IOU is 0.5, the average accuracy rate of target segmentation is 91.83% and the average recall rate is 92.94%. Using the trained model to detect the breakwater images taken by drones in actual projects, the identification rate of accropodes is 90.7%, and the shooting angle and height have little effect on the identification accuracy. Therefore, the Mask R-CNN deep learning network can realize the accurate recognition of dense and complex armour layer blocks with good portability and versatility.

Key words: Mask R-CNN; breakwater; block identification; accropode

防波堤护面层是防波堤与波浪直接接触的部分,承受波浪的冲击作用,保护防波堤其他部位不受破坏。

基金项目: 国家自然科学基金(52001149);中国科协青年人才托举工程(2018QNRC001)
作者简介: 高林春(1992—),男,博士研究生,主要从事机器学习和计算机视觉研究。E-mail:337503806@qq.com
通信作者: 陈松贵(1987—),男,副研究员,博士,主要从事波浪-结构物相互作用研究。E-mail:chengs05@163.com
引用本文: 高林春,王收军,陈松贵,等. 基于 Mask R-CNN 的防波堤复杂护面块体检测和分割方法[J]. 河海大学学报(自然科学版), 2022,50(4):121-126.
GAO Linchun, WANG Shoujun, CHEN Songgui, et al. Identification and segmentation technology of complex armour blocks of rubble mound breakwater based on Mask R-CNN [J]. Journal of Hohai University(Natural Sciences), 2022,50(4):121-126.

斜坡式防波堤的破坏可以定义为其功能的部分或全部丧失,通常与护面层的水力失稳有关^[1-2]。Husdon^[3]使用护面块体移动的百分比来描述破坏的特征。JTS 154-1—2011《防波堤设计与施工规范》给出了每100 m²护面层需要安放的块体个数。通常判断护面层是否破坏的手段之一是比较一定面积护面层块体的个数和该面积内实际应安放的块体个数。

目前,通常采用人工手动统计方法得到护面块体的个数。该方法需要操作人员去现场进行实地统计或通过图像进行统计,效率低、精度差。随着人工智能的兴起,深度学习算法发展迅猛,计算机代替人工识别护面块体成为可能。由于防波堤护面层块体的摆放存在较多粘连,当使用深度学习算法检测和分割块体时分割困难,因此护面块体的精确分割是实现护面块体个数统计的重要前提。

传统的目标检测算法在背景和目标区分明显时效果较好,但是需要人为的设计特征,算法鲁棒性较差,时间复杂度较高,难以为复杂的护面块体识别提供有效方法^[4-5]。2012 年以来,深度卷积神经网络(convolutional neural network, CNN)发展迅速,在 ImageNet 图像识别大赛中将图像分类错误率从 25% 降至 15.3%^[6]。Ciresan 等^[7]用 CNN 来挑战语义分割,打破了 CNN 只用于目标分类的常规做法, CNN 开始用于图像分割领域。CNN 不需要人为的设计特征,可以通过训练网络提取图像特征,大大提高了算法的鲁棒性。Girshick 等^[8]提出了 RCNN(regions with CNN),将深度学习算法应用于目标检测,通过训练深度学习网络进行特征提取,然后将训练得到的模型用于目标定位和图像分割。He 等^[9]于 2017 年提出 Mask R-CNN 深度学习网络可以对密集的目标进行识别和分割,且算法具有较高的识别和分割准确率。

本文将 Mask R-CNN 深度学习网络用于复杂的防波堤扭王字块的检测和分割,并准确识别了图像中扭王字块的边界,可为斜坡式防波堤健康状态监测提供算法支持。

1 检测和分割方法

1.1 图像采集

采用从海康威视 DS 网络摄像机拍摄的视频中截取的防波堤护面层扭王字块图像作为训练样本,试验中扭王字块单个重量为 25 kg,块体最大高度 42 cm,拍摄时相机距扭王字块表面 3 m。该训练样本为大小不一的彩色图,如图 1(a)所示。在使用 Mask R-CNN 深度学习网络训练时,首先要制作训练使用的数据集,即对样本进行标注。由于训练图像中扭王字块过于密集,标注的图像过多会加大训练的时间开销。因此,本文共标注 15 张防波堤护面层扭王字块图像作为训练数据,其中,训练集 12 张图像,测试集 3 张图像。

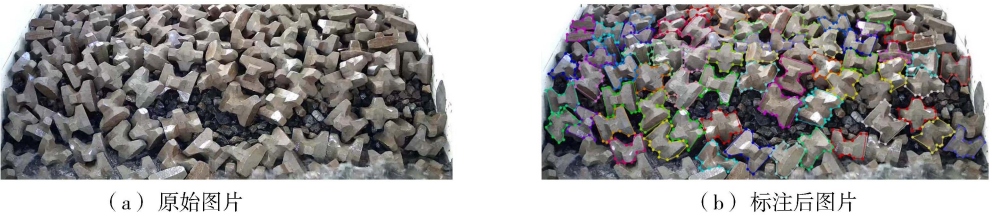


图 1 标注前后的扭王字块

Fig.1 Accropodes before and after labeling

1.2 图像标注

选择图像中所有扭王字块作为目标,使用标注软件 labelme 的多边形功能将所有扭王字块的边缘轮廓标注出来,不同的块体用不同的标签,如 unit1, unit2。Labelme 在标注完成后会自动生成一个与标注图像命名相同的 json 文件来保存所有的标注信息,标注后的图像如图 1(b)所示。标注点形成了一个闭合的多边形,多边形内部为扭王字块目标。图像中的不同扭王字块会用不同颜色的闭合多边形包围,以表示不是同一块体。对数据集中的所有图像都进行标注,生成与图像一一对应的 json 文件,此时每个 json 文件只包含与之对应的图像的信息,而输入 Mask R-CNN 深度学习网络中的 json 文件应包含数据集中的所有标注信息,因此需要将所有的 json 文件合并成一个 json 文件,输入到 Mask R-CNN 深度学习网络中进行训练。

1.3 基于 Mask R-CNN 的护面块体检测

1.3.1 Mask R-CNN 深度学习网络基本结构

Mask R-CNN 深度学习网络主要框架如图 2 所示,是基于 Faster RCNN^[10]改进而来的,使用 RoIAlign 代

替了 RoIPooling,并添加了一个用于预测目标掩膜 (Mask)的新分支。图 2 上方灰色部分为 Faster RCNN 框架,下方红色部分是在 Faster RCNN 基础上新增的实例分割网络。常用的 Mask R-CNN 深度学习网络的主干网络有 ResNet50 和 ResNet101,主干网络的选择会直接影响特征提取的最终结果。研究表明,在相同的条件下,使用 ResNet101 比 ResNet50 的网络效果更好^[11]。

采用 ResNet101 与 FPN(feature pyramid networks)相结合的网络结构对扭王字块图像进行特征提取。ResNet101 网络包含 101 个卷积层,通过卷积进行特征提取,形成 stage1、stage2、stage3、stage4、stage5 共 5 层特征图,不同阶段的特征图大小和维度均不同。stage5 包含很强的语义信息,但是它的特征图最小,空间信息损失很大。为了充分利用 ResNet101 网络各个阶段提取的特征,使用 FPN 网络将不同阶段的特征图进行融合,并将融合后的特征图输入区域候选网络。FPN 为特征金字塔网络^[12],它在自底向上特征提取过程后增加了自顶而下的特征融合过程,并将两个过程横向连接,即将底层的空间信息和高层的语义信息融合,从而提升模型的检测性能,也可以更好地检测图像中的小目标。

区域候选网络对特征图中的感兴趣区域(ROI)进行前景或背景的二值分类,并对包含扭王字块体的边界框回归,使得边界框尽量贴合块体大小。

通过区域候选网络过滤一部分 ROI,剩下包含扭王字块的 ROI,这些 ROI 被输入 ROIAlign 中调整为固定的尺寸,然后分别输入分类和边框回归两个网络。分类和边框回归网络利用训练的 ROI 分类器和边框回归器来进行扭王字块的识别,将 ROI 分为背景或扭王字块,并调整 ROI 的边界框,最后准确地检测到扭王字块。分割掩膜(Mask)网络是一个卷积网络,取 ROI 分类器分类结果是前景的 ROI 为输入生成其掩码,掩码的大小和形状与扭王字块一致,通过掩码对扭王字块进行分割。最后将识别与分割结果相结合,得到分割后的图像。

1.3.2 模型参数

使用 Windows10 操作系统,显卡为 NVIDIA Quadro RTX 5000,显存为 16GB。使用 Mask R-CNN-FPN-ResNet101 做主干网络。

预训练模型使用 COCO2014 数据集训练得到的权重模型,在此基础上训练适用于扭王字块识别和分割的 Mask R-CNN 模型。COCO2014 数据集是微软在 2014 年发布的大型图像数据集,包含 82 783 个训练样本,40 504 个验证样本以及 40 775 个测试样本,是目标检测竞赛的常用数据集之一。

对最终选择的参数组合通过多次试验进行比较:①取 max epoch 为 30,初始学习率 0.001,动量默认 0.9。试验表明,网络对块体的识别效果不佳,即网络欠拟合,模型复杂度较低。②取 max epoch 为 60,初始学习率 0.001,动量默认 0.9,结果表明,网络会存在将一个目标检测为两个的现象,即网络过拟合,泛化能力较差。

载入预训练模型的权重对手动标注的扭王字块训练集进行训练,根据模型训练的情况对参数进行微调,最终的参数设置如下:训练集样本数为 12,测试集样本数为 3,循环次数为 100,批大小为 1,迭代次数为 50 次,学习率为 0.001,权重衰减系数为 0.0005。训练网络需要运行时间约 11 h,训练完成后分割需要几秒钟。

检测模型性能的重要评价指标是精确率和召回率。精确率表示正确分类的目标占整个被分为该类目标个数的百分比;召回率表示正确分类的目标占该类所有目标个数的比例。在计算精确率和召回率时,使用交并比(intersection over union, IOU)来评估预测质量,IOU 值越大,表明预测结果与真实值越接近。不同的 IOU 阈值与精确率和召回率的关系如表 1 所示。精确率和召回率的计算公式如下:

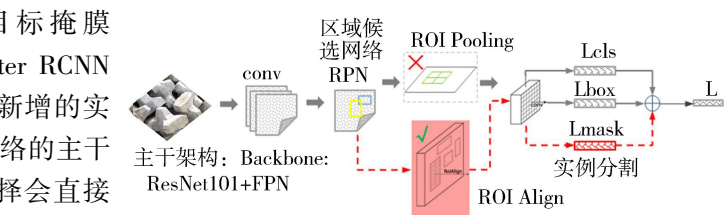


图 2 Mask R-CNN 深度学习网络框架
Fig. 2 Framework of Mask R-CNN deep learning network

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{1}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{2}$$

式中: P 为精确率; R 为召回率; N_{TP} 为存在的目标块体并被正确检测出来的数量; N_{FP} 为将其他目标预测为块

体的数量; N_{FN} 为将目标块体预测错误的数量。

从表1可以看出,随着 IOU 阈值的递增,精确率和召回率逐步下降,当 IOU 阈值为 0.9 时,精确率和召回率大幅减小。通过比较不同 IOU 阈值下的精确率和召回率,本文选择 0.5 作为 IOU 阈值,此时精确率和召回率取得最大值。

2 方法验证

2.1 原图和分割后的图像

目标块体是指图像中需要识别的扭王字块体。图3是包含多个目标块体的原图和分割后的图像。由图3(a)可以看到图像中有许多有一定重叠度的目标块体;图3(b)为使用训练好的 Mask R-CNN 深度学习网络识别后的结果,多个目标块体分别被不同颜色的掩码覆盖并各自处于一个独立的虚线框内,每个虚线框上方是该框内目标的预测标签和得分,即该虚线框中目标块体的类别和目标块体属于该类别的概率。本文试验只对扭王字块一个类别进行训练,因此最后预测的结果只有一个类别。



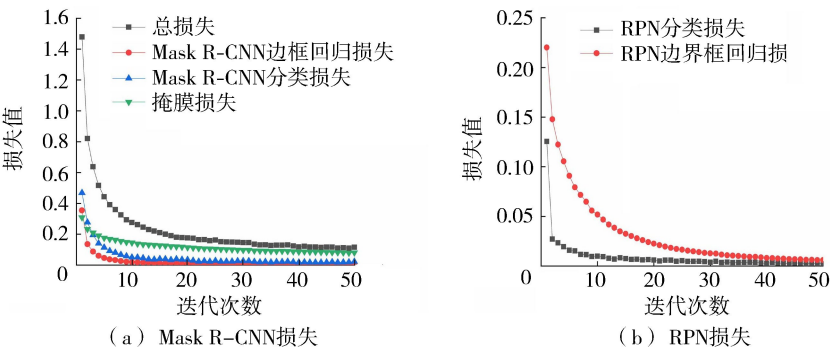
(a) 原图 (b) 分割后的图像

图3 原图与分割后的图像样本

Fig.3 Samples of original image and segmented image

为了验证方法的可靠性,首先,对 Mask R-CNN 模型训练过程的损失函数进行分析;其次,选择 3 张图像作为验证集,利用 Mask R-CNN 深度学习网络对其进行分割,再进行精确率和召回率的分析。

Mask R-CNN 深度学习网络在训练的过程中需要计算预测结果和标注值间的损失并进行反向传播来优化模型参数,损失变化曲线如图4所示。



(a) Mask R-CNN损失 (b) RPN损失

图4 Mask R-CNN 深度学习网络目标分类 Loss 及 Mask R-CNN 深度学习网络目标掩膜 Loss 曲线

Fig.4 Target classification Loss and target mask Loss curves of Mask R-CNN deep learning network

Mask R-CNN 深度学习网络的损失分两部分组成:①Mask R-CNN 深度学习网络的 heads 损失,包括边框回归损失、分类损失、掩膜损失;②RPN 网络的损失,包括 RPN 前景/背景分类损失和 RPN 边界框回归损失。如图4(a)所示为组成 Mask R-CNN 损失的五部分损失。如图4(b)所示为 RPN 对候选框前景或背景的判断损失和 RPN 提取的边界框损失。从图4可以看出,随着迭代次数的增加,各部分损失的衰减趋势相同,迭代 50 次后,总的损失降为 0.12,掩膜损失降为 0.08,剩下四部分损失都小于 0.02,对数据拟合效果好。

2.2 Mask R-CNN 深度学习网络测试结果

将训练好的 Mask R-CNN 深度学习网络在测试集上进行测试,最终得到的平均精确率为 91.83%,平均召回率为 92.94%。在检测过程中,由于目标块体之间有一定的重叠度,加上目标块体的密集性以及图像边

缘处的块体无法识别,会有漏检的情况发生,从而影响识别的精确率和召回率。

2.3 工程实例验证

采用训练得到的模型对无人机航拍的某港口防波堤护面层图像进行分割识别。该防波堤扭王字块重 25 t,块体最大高度 4.2 m,航拍高度为 30 ~ 50 m。选择 6 张图像进行预测,如图 5 所示,其中图 5(a)(b)为防波堤同一位置、不同高度拍摄的图像;图 5(c)(d)为同一高度、不同角度拍摄的图像;图 5(e)(f)的拍摄高度和角度相同,图 5(e)中的白色框内为规则摆放的块体,其余块体为不规则摆放,图 5(f)中块体均为不规则摆放。分别对图 5(a) ~ (f) 识别结果进行统计分析,每张图像中识别出的块体个数占图像中总块体个数的百分比分别为 90.7%、89.3%、87.7%、86.9%、70.6% 和 82.4%。

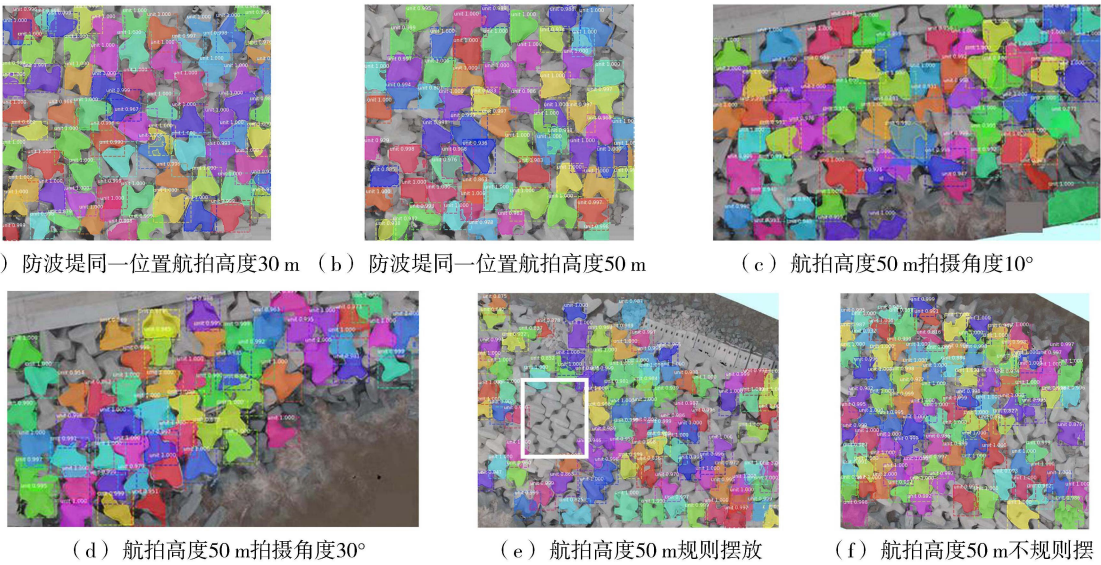


图 5 实际防波堤护面块体预测结果

Fig.5 Prediction results of actual breakwater armour blocks

从预测结果可以看出:①使用实验室块体图像作为训练样本得到的网络模型可以对防波堤护面层的原型图像进行识别且识别效果较好,最大识别精确率为 90.7%,说明该方法具有良好的移植性。②图 5(a) ~ (d) 中块体均为不规则摆放方式,图 5(a) 的拍摄高度为 30 m,图 5(b) 的拍摄高度为 50 m,识别块体个数占比之差为 1.4%;图 5(c)(d) 的拍摄角度不同,识别块体个数占比之差为 0.8%,由此可见,拍摄高度和角度对于不规则摆放块体的识别效果影响不大,说明该方法具有良好的通用性。③通过对比图 5 的 6 幅图可知,影响防波堤护面扭王字块识别精度的主要因素是块体的摆放规则。对于图 6(a) 中规则摆放方式,由于在训练集中缺少该摆放方式,因此识别精度较差(图 5(e)),而对于如图 6(b) 所示的不规则摆放方式,实验室训练的网络具有良好的识别精度(图 5(f))。因此,可以使用实验室物理模型数据集训练的网络去识别原型中采集的图像,节约了人力物力,大大减少了工作量,为防波堤护面层健康监测提供了技术支持。

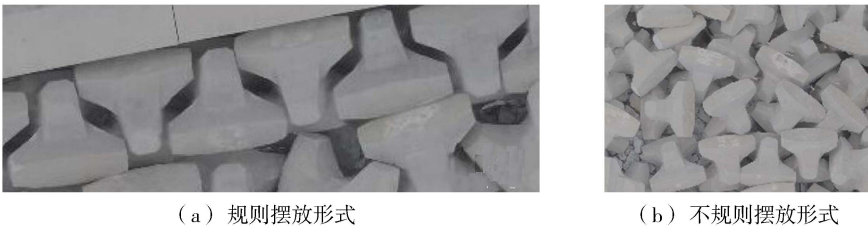


图 6 防波堤护面块体的摆放形式

Fig.6 Placement form of breakwater armour block

3 结 语

本文将 Mask R-CNN 深度学习网络用于防波堤复杂护面块体的检测,能够使用实验室块体图像训练好

的网络自动检测图像中的目标块体并对图像中目标块体的个数进行统计,解决了由于块体粘连识别和分割困难的问题。Mask R-CNN 深度学习网络对比较密集和有一定重叠度的扭王字块进行较为准确的识别和分割,未出现识别错误的情况;目标分割的平均精确率为 91.83%,平均召回率为 92.94%;使用训练的网络对训练集中没有的防波堤护面层原型图像进行识别,在块体为不规则摆放时,块体的识别率可达 90.7%,且拍摄角度和高度对识别精度影响不大,证明了该方法具有良好的移植性和通用性;同时,可以使用实验室物理模型数据集训练的网络去识别防波堤原型中采集的图像,提供了一种使用实验室物理模型训练结果来识别原型块体的新方法。

参考文献:

[1] Coastal Engineering Manual[R]. Washington D. C. : United States Army Corps of Engineers, 2002.

[2] The Rock Manual. The Use of Rock in Hydraulic Engineering[R]. London: CIRIA, 2007.

[3] HUDSON R Y. Laboratory investigation of rubble-mound breakwaters[J]. American Society of Civil Engineers, 1959, 85: 93-121.

[4] 徐渊,许晓亮,李才年,等. 结合 SVM 分类器与 HOG 特征提取的行人检测[J]. 计算机工程,2016,42(1):56-60. (XU Yuan, XU Xiaoliang, LI Cainian, et al. Pedestrian detection combined with SVM classifier and HOG feature extraction[J]. Computer Engineering, 2016, 42(1): 56-60. (in Chinese))

[5] 李忠伟. 支持向量机学习算法研究[D]. 哈尔滨:哈尔滨工程大学,2006.

[6] KRIZHEVSKY A,SUTSKEVER I,HINTON G. ImageNet classification with deep convolutional neuralnetworks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 84-87.

[7] CIRESAN D C, MEIER U, GAMBARDELLA L M, et al. Convolutional neural network committees for hand-written character classification[C]//Proceedings of the 2011 International Conference on Document Analysis and Recognition. Beijin: IEEE Computer Society, 2011: 1135-1139.

[8] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. IEEE Computer Society, 2013, 32(1): 1-8.

[9] HE Kaiming, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 1-12.

[10] REN Shaoqing, HE Kaiming, ROSS G, 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): 1-4.

[11] CHENG X,ZHANG Y H,CHEN Y Q,et al. Pest identifi-cation via deep residual learning in complex background[J]. Computers and Electronics in Agriculture, 2017, 141: 351-356.

[12] LIN Tsungyi, DOLLÁR P, GIRSHICK R, et al. FPN: Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Visionand Pattern Recognition(CVPR). Honolulu, USA: IEEE, 2017: 1-5.

(收稿日期:2021 -08 -10 编辑:胡新宇)