

李棋, 路胜男, 马千里. 基于深度学习的裂纹实例分割方法[J]. 智能计算机与应用, 2024, 14(12): 18-25. DOI: 10.20169/j.issn.2095-2163.241203

基于深度学习的裂纹实例分割方法

李棋, 路胜男, 马千里

(南京邮电大学通信与信息工程学院, 南京 210003)

摘要: 本文针对传统裂纹检测方法的检测精度较低的问题, 提出一种先验知识改进和深度学习相融合的裂纹实例分割方法。使用单阶段网络 YOLOv5 进行实验验证, 先验知识改进包括使用裂纹图像分类模型过滤 mosaic 增强产生的错误样本并对相连接裂纹的标签信息进行融合, 实现训练集的纠正与增强; 使用基于交并比的 K-means++ 算法生成先验框, 加速训练回归收敛。最终实验验证, 先验知识改进不增加网络参数, 实现更高的检测精度, 同时还保持单阶段网络较高的检测速度, 验证了对于先验知识的改进可以有效地提高裂纹实例分割的检测精度, 增强网络模型的泛化性。

关键词: 裂纹检测; 实例分割; mosaic 增强; 先验框; K-means++

中图分类号: TP183

文献标志码: A

文章编号: 2095-2163(2024)12-0018-08

Crack instance segmentation method based on deep learning

LI Qi, LU Shengnan, MA Qianli

(School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Aiming at the problem of low detection accuracy of traditional crack detection methods, this paper proposes a crack instance segmentation method which integrates prior knowledge improvement and deep learning. The single-stage network YOLOv5 is used for experimental verification. The prior knowledge improvement includes the use of crack image classification model to filter the error samples generated by mosaic enhancement and the fusion of the label information associated with the connected cracks to realize the correction and enhancement of the training set. The K-means++ algorithm based on intersection ratio is used to generate prior frames and accelerate the training regression convergence. The final experiment verifies that the improvement of prior knowledge does not increase network parameters, achieving higher detection accuracy, while maintaining a high detection speed of single-stage network, verifying that the improvement of prior knowledge can effectively improve the detection accuracy of crack instance segmentation and enhance the generalization of network model.

Key words: crack detection; instance segmentation; mosaic enhancement; prior bounding box; K-means++

0 引言

研究可知, 裂纹是一种常见的工业质量缺陷。在建筑工程施工和使用的过程中不可避免地会因损坏而导致裂纹的产生。如果细小裂纹无法得到及时处理, 随着建筑工程的长期使用则会导致裂纹的继续扩大, 随之带来的危害也会增加, 如此就会极大影响在用建筑的安全性和耐久性。

传统的裂纹检测包括人工检查、图像处理和机器学习。其中, 人工检查会耗费大量人力资源, 同时检测结果主观性较大, 很难做到快速且精准的检

测^[1]; 图像处理由于算法复杂度较大, 且泛化性较低, 针对复杂环境检测效果差, 无法满足工业需求; 机器学习手动提取有效特征的难度较大, 泛化性较差, 处理大规模数据容易出现过拟合和欠拟合的问题。

近年来, 深度学习被广泛应用于各种领域, 例如图像和语音识别、自然语言处理等方面。许多学者在研究中将深度学习中的卷积神经网络 (Convolutional Neural Network, CNN) 应用在裂纹检测的应用中, 并得到不错的效果。CNN 就是利用层次化的结构实现特征提取和抽象, 对输入的数据进行处理和分类, 通过对网络迭代训练不断地优化更

作者简介: 李棋 (1998—), 男, 硕士研究生, 主要研究方向: 图像处理; 路胜男 (1998—), 女, 硕士研究生, 主要研究方向: 图像处理。

通信作者: 马千里 (1978—), 男, 博士, 副教授, 主要研究方向: 生物医学电子学, 图像处理, 深度学习。Email: maql@njupt.edu.cn。

收稿日期: 2023-06-16

哈尔滨工业大学主办 ◆ 学术研究与应用

新卷积核和全连接层的权重和偏置值来提高模型准确性,最终训练得到的模型可用来预测其输入图像。CNN 在一定程度上可以实现特征提取的自动化,实现端到端的图像预测,使整个系统更加简洁高效。而与此同时,深度学习也开始陆续应用于图像识别任务,CNN 通过卷积操作和池化操作可以提取图像的高维特征,克服传统机器学习需要专业知识设定特征的困难,并通过网络结构的加深,进一步提高模型的泛化能力。Silva 等学者^[2]运用深度学习网络对裂纹图像进行分类,使用 VGG16 网络进行迁移学习,最终得到 92.27% 的图像分类精确率。Mandal 等学者^[3]提出基于单阶段神经网络 YOLOv2 实现道路裂纹自动检测的方法,并得到 88.51% 的精确率和 87.10% 的回归率的测试结果。Arman 等学者^[4]运用双阶段神经网络 R-CNN 和 Faster R-CNN 实现道路损伤检测,并取得最佳精确率 98.02% 的结果。Tan 等学者^[5]将 Mask R-CNN 应用在道路裂纹的检测中,实现了裂纹目标的像素级别识别,并取得了不错的识别结果。朱苏雅等学者^[6]通过改进 U-Net 网络,运用后处理检测结果的方式,最终的准确率达到 90% 以上。

GN 基于深度学习的实例分割技术作为当前比较热门的研究方向,针对裂纹对象,实现了分类、定位和形态信息的预测^[7],可以通过形态信息有效地评估裂纹的危害性,并通过定位信息开展后续的维修和保养工作,具有极大的实用价值。但由于裂纹特征复杂,形态多样,通常收集采样的数据集特征较

弱,极大限制了深度学习网络的检测精度。因此,本研究通过对数据集和先验框等先验知识增强,提高数据集质量,最终实现网络模型检测精度和泛化性的提升。

1 相关技术说明

1.1 数据增强

数据增强也叫做数据扩增,表示在现有的数据集基础上进行额外处理得到新的数据,最终达到数据集和特征扩增的目的。通过数据增强获得结构合理的数据可以解决数据集单一容易造成的过拟合问题^[8]。

比较常用的数据增强有 2 种类型:几何变换和颜色变换。其中,几何变换包括翻转、旋转、裁剪、变形等操作;颜色变换则包括噪音、模糊、颜色变换等操作。2 种数据增强类型都是根据单张图像进行变换,虽然实现了数据集数量的扩增,但是图像特征增强的效果并不好。

到目前为止,mosaic 增强^[9]利用数据集中多张图像进行拼接,有效地丰富了信息,提高输入图像复杂度,避免产生过拟合。mosaic 增强流程如图 1 所示。图 1 中,mosaic 增强会将输入图像分成 4 部分,并随机从训练集中选择 3 张图像和当前输入图像进行中心点随机选取裁剪成原始图像大小,针对新图像中空白部分则会使用固定颜色进行填充,最后对拼接后图像进行随机旋转翻转、平移缩放和调节色域等图像增强操作。其中,蓝色框为目标框信息,黑点为随机中心点,黑色框为裁剪区域。

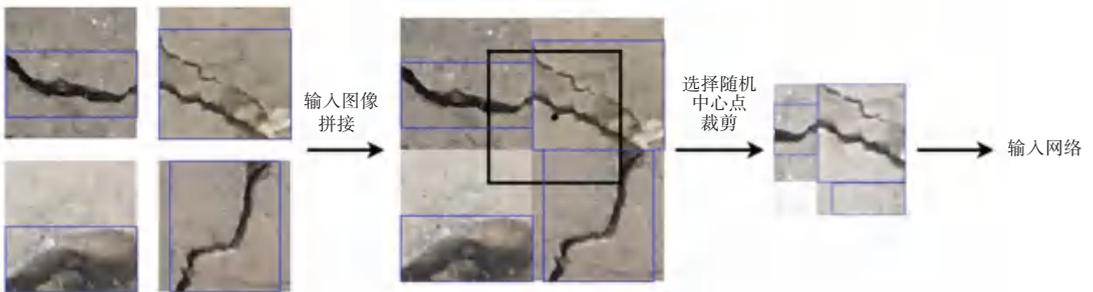


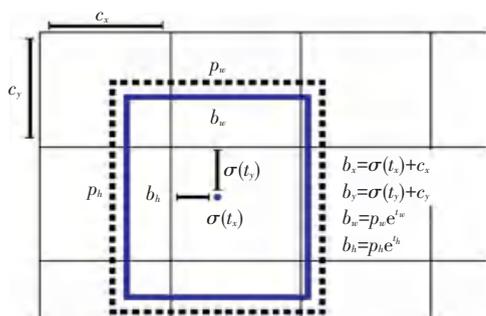
图 1 mosaic 增强流程图

Fig. 1 Flowchart of mosaic enhancement

1.2 先验框

先验框也叫做锚框,作为一种先验知识,在检测过程中,会遍历输入图像上所有可能存在目标的矩形框,并对其大小和位置进行偏移调整,最终选出真正存在物体的目标框,这种矩形框即可称为先验框。一般情况下,为了提高检测效果,通常会在同一个位置设置不同大小和长宽比的先验框^[10-12]。

先验框通常采用中心坐标和宽高进行描述,如图 2 所示,先验框在训练过程中,多会根据特征网格位置预设 4 个参数,分别是:中心点坐标 C_x 和 C_y 、宽高 p_w 和 p_h 。在训练过程中分别预测偏移参数 t_x 、 t_y 、 t_w 和 t_h ,使用图 2 中公式进行参数更新,最终通过 $Clou$ ^[13] 损失函数反向传播,将先验框收敛回归成目标框。

图2 先验框^[10]Fig. 2 prior bounding box^[10]

2 YOLOv5 检测算法及其改进

为体现本文研究的设计效果,采用 YOLOv5 7.0 版本中的 YOLOv5s 的实例分割网络进行实验,这里 s 表示该网络中的一个模型版本。具体是在 YOLOv5s 目标检测网络的基础上新增了 Proto 模块来获取图像原型掩膜,同时在检测头部分新增相应原型掩膜数量的权重参数预测,最终在对两者进行加权叠加后通过预测框裁剪和阈值处理得到掩膜预测信息。YOLOv5 实例分割网络结构如图 3 所示。

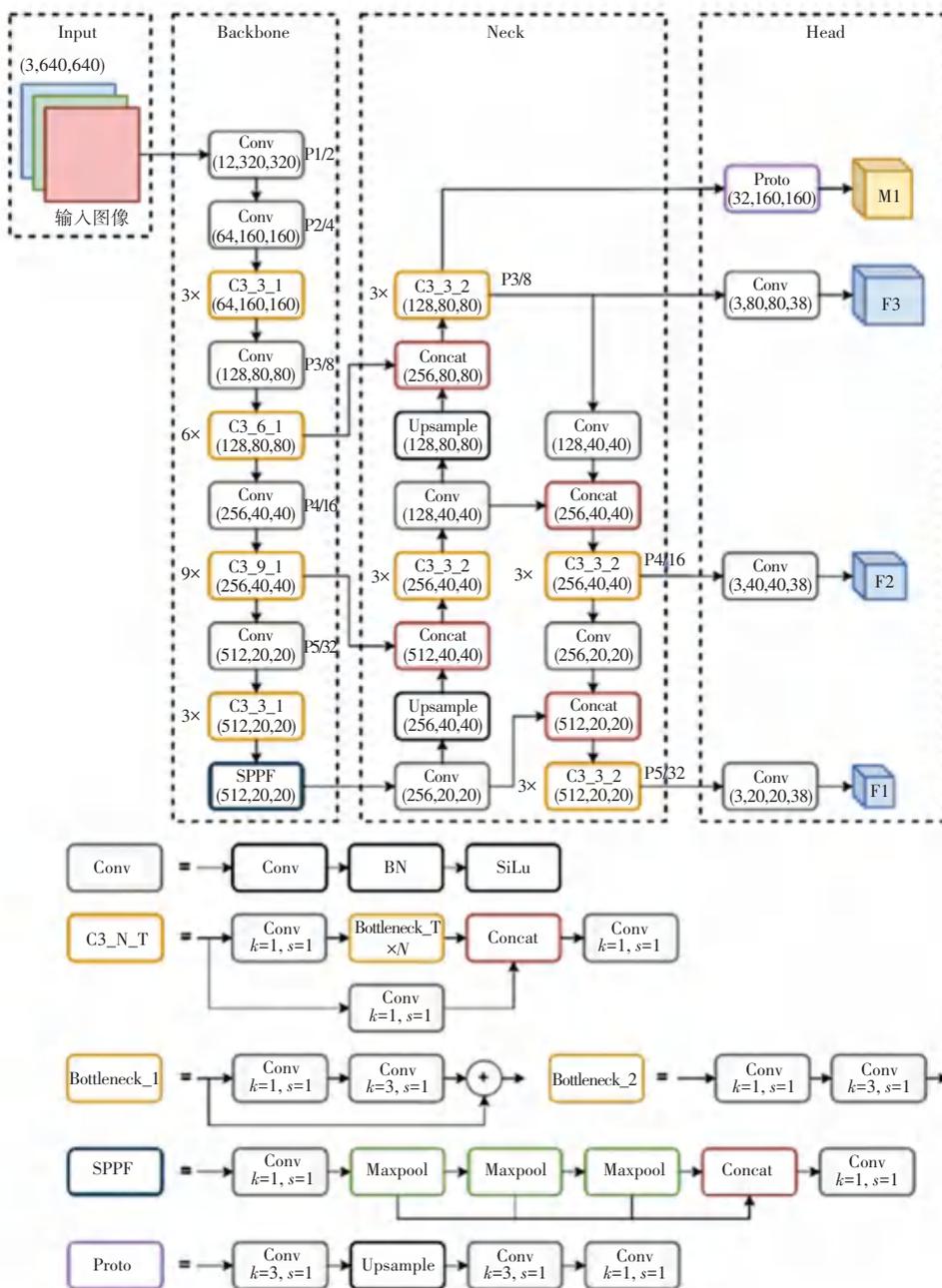


图3 YOLOv5 实例分割网络结构图

Fig. 3 Structure diagram of YOLOv5 instance segmentation network

2.1 改进 mosaic 增强

在 YOLOv5s 的输入端对训练集进行 mosaic 增强的先验知识处理, 其中 mosaic 增强对 4 张训练集样本图像裁剪后进行随机中心点拼接, 因此会对样本图像中目标对象造成不确定的裁剪, 同时还由于裂纹目标的形态复杂, 在真实框中往往占比较小, mosaic 增强后的样本图像会存在目标框中实际没有裂纹对象的问题, 如图 4 所示。图 4 中, (a) 为原始 mosaic 增强拼接后的图像, (b) 为原始 mosaic 增强带蓝色目标框信息图像。由于裂纹目标在目标框中一般是对角存在(参见图 1)。当在对角方向上进行裁剪时, 极大可能会造成裁剪后目标框中没有裂纹目标的情况, 这样一来就会给网络模型提供错误样本, 增加模型训练难度, 还会造成模型对样本图像误检, 影响检测精度。针对该问题, 本文提出改进 mosaic 增强, 使用裂纹图像分类模型对裁剪后目标框内的图像进行图像分类预测, 如果预测含有裂纹目标, 则保留目标框信息, 反之删除。运用此方法可

减少产生错误样本的概率, 实验进行改进 mosaic 增强后的训练集的目标框信息见图 4(c)。红色框为过滤后的目标框, 通过对裁剪图像进行分类就可一定程度上减少错误样本。其中, 裂纹图像分类模型使用 Efficientnet_b0 网络^[11]在同一数据集上训练得来。

由于裂纹具有连接性, 即多个普通目标相连接仍为多个目标, 但多个裂纹目标相连接, 从人类视觉经验而言则属于一个裂纹目标。基于此, 为增强裂纹特征, 本文提出使用膨胀和腐蚀操作对裂纹目标的连接性进行判断, 并以此为基础来合并标签信息, 如图 5 所示。由于标签信息为矩阵, 图 5 中绿色和蓝色分别表示 2 条裂纹, 通过膨胀分别产生橙色和紫色边缘像素, 红色部分则表示两裂纹相连接, 再通过腐蚀操作来平滑裂纹连接处像素, 得到的达到标签信息合并的作用, 得到的效果如图 6 所示。由图 6 可以看出, 合并标签信息后的图像具有更复杂的裂纹形态信息, 实现了训练集特征的增强。

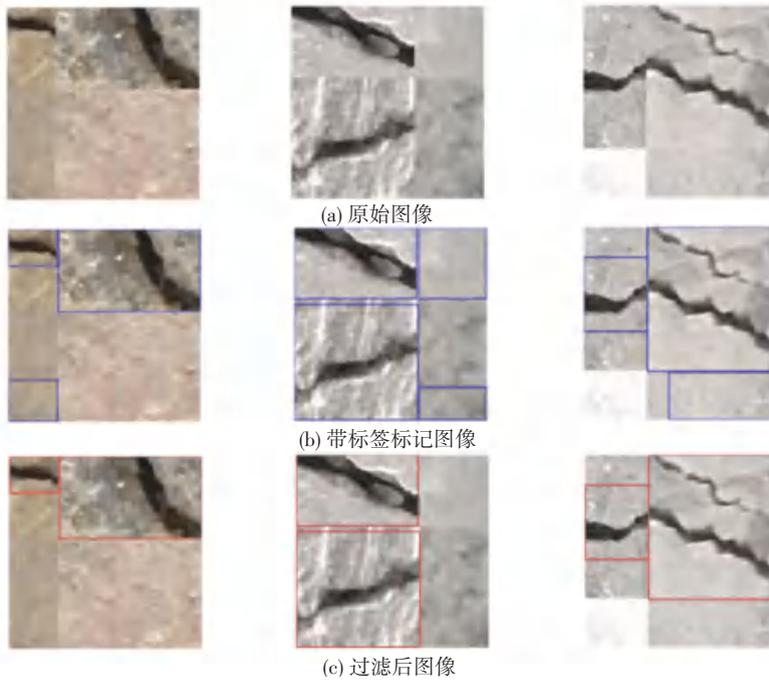


图 4 mosaic 增强效果

Fig. 4 Effect of mosaic enhancement

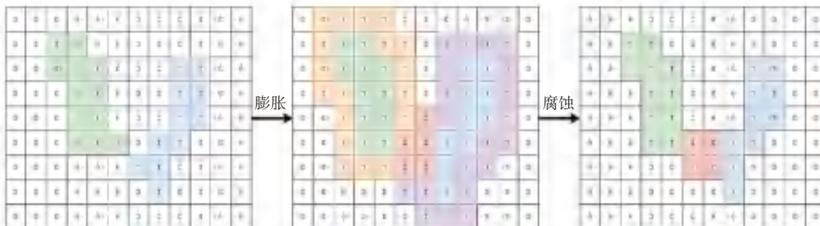


图 5 标签信息的膨胀与腐蚀操作

Fig. 5 Expansion and corrosion operations of label information

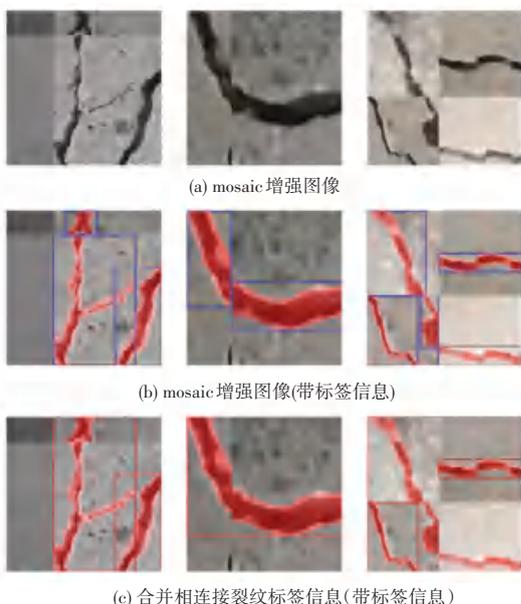


图6 合并相连接裂纹标签信息效果

Fig. 6 Label information effect combined with the connected crack

2.2 改进先验框

先验框作为最终预测框的一个先验知识,其质量优劣将极大影响最终训练模型的检测效果。实例分割任务中的先验框、即 YOLOv5s 实例分割网络中默认先验框由 COCO 数据集聚类而来,因此和其他数据集的匹配效果可能并非最佳,因此 YOLOv5s 提供 K-means 算法^[14]对数据集进行聚类,其流程如图 7 所示,通过在训练集所有目标框中随机选择 9 个宽高数据点作为初始聚类中心,并选取范式距离作为聚类的距离标准,每次迭代计算各个数据点与选定聚类中心的聚类距离,并将数据点分配给与数据点聚类距离最短的聚类中心簇类中,随后将每个聚类中心簇类的数据点采用取均值的方式来更新聚类中心,然后重复上述迭代计算和更新聚类中心操作,直到达到某一终止条件,通常终止条件为指定迭代次数或者聚类中心更新达到收敛。

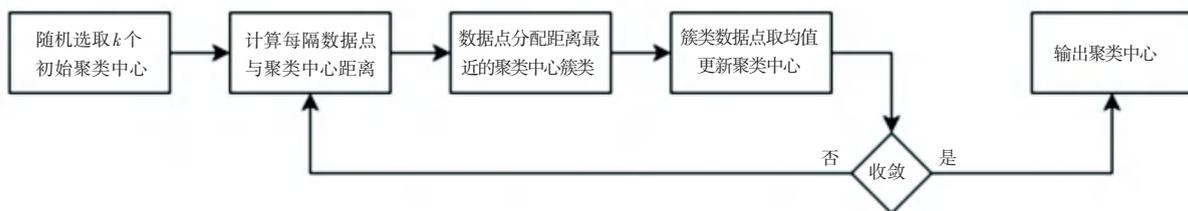


图7 K-means 算法流程图

Fig. 7 Flow chart of K-means algorithm

由于 K-means 算法初始聚类中心是随机选取,容易导致收敛至局部最优解,如图 8 所示。K-means 算法两次聚类同一组数据集的效果是不一样的,具有不稳定性,同时默认的范式距离的聚类标准也并不适合先验框收敛,因此本文提出基于交并比聚类标准的 K-means++ 算法^[15]。其中, K-means++

算法在初始聚类中心的选取时,是通过随机选取一个聚类中心,计算所有数据点与初始聚类中心的聚类距离,然后通过加权概率分布方法选取第 2 个聚类中心,随后迭代选取出所需聚类中心个数,使得算法初始聚类中心尽可能距离远,从而避免了收敛至局部最优解的问题。

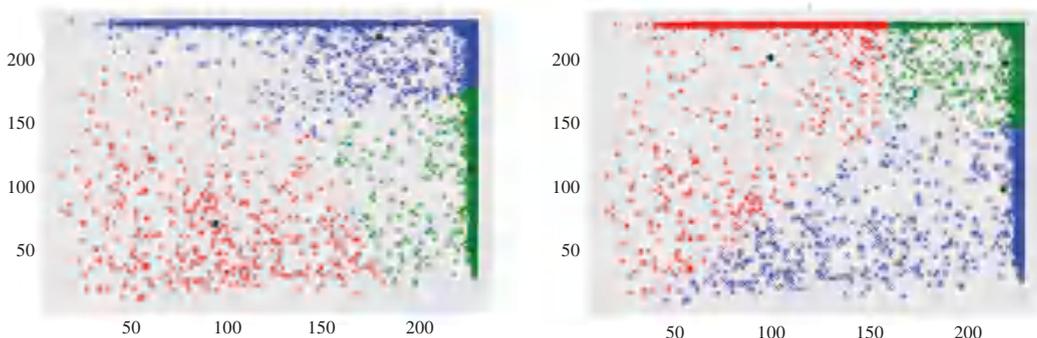


图8 K-means 算法聚类效果图

Fig. 8 Clustering effect of K-means algorithm

3 结果与分析

3.1 裂纹数据集

实验采用公开的 SurfaceCrackDetection 部分数据集^[16-17], 基础数据是从 METU 校园混凝土建筑的



图 9 裂纹数据集

Fig. 9 Crack data set

标签信息利用 labelme 模块包进行标记, 最终数据集使用 1 297 张裂纹图像和 133 张背景图像, 采用 7 : 3 : 3 的比例随机生成训练集、验证集和测试集, 对应的数量见表 1。

表 1 裂纹数据集图像统计

Table 1 Crack data set image statistics

参数	数量/张
总数据集数	1 430
训练集	770
验证集	330
测试集	330

3.2 实验环境及训练参数

本实验使用训练和检测模型的软件环境为 ubuntu 18.04.2, 使用 Python 3.9, PyTorch 1.13.1 深度学习框架, 采用显卡训练信息为 NVIDIA GeForce RTX 2080ti, 显存 11 GB。CUDA 版本为 11.4, CUDNN 版本为 9.1。训练时将输入图像缩放为 640×640 大小, 使用初始学习率为 1e-3, 训练的批尺寸为 64, 使用 Adam 优化器和交叉熵损失函数优化网络模型。出于数据集样本特征简单的考虑, 训练设置 dropout 参数为 0.5。

3.3 模型评估指标

实验使用平均精度均值 (mean Average Precision, mAP) 作为模型检测准确性能的评估指标, 由于实验为单分类任务, 因此 mAP 等价于平均精度 (Average Precision, AP)、综合衡量精确率 (Precision, P) 和召回率 (Recall, R), mAP 指标的计算公式如下:

墙壁和地板上采集的图像补丁获得。这些图像是在距离表面约 1 m 远的地方拍摄, 相机直接面对目标。用于训练、验证和测试的图像样本如图 9 所示。这里的图像依据是否含有裂纹目标分成裂纹和背景两类。

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

$$AP = \int_0^1 P(R) dR$$

$$P = \frac{TP}{FP + TP}$$

$$R = \frac{TP}{TP + FN} \quad (1)$$

其中, TP 、 FP 和 FN 表示预测结果。这里, T 和 F 分别表示预测正确和错误; P 和 N 分别表示预测为目标和背景。实验针对实例分割分别评估目标框和掩膜使用 mAP_{50}^{bbox} 、 mAP_{50-95}^{bbox} 和 mAP_{50}^{mask} 、 mAP_{50-95}^{mask} , 对应上标 bbox 和 mask 分别表示目标框评估和掩膜评估, 下标数值表示判断预测结果的交并比阈值; 50 表示阈值为 0.5, 50-95 表示从 0.50 到 0.95。每隔 0.05 计算一次 mAP 值, 最后求均值。

3.4 消融实验与分析

实验在测试集上进行验证, 训练的批尺寸为 16, 初始学习率为 1e-3, 学习衰减采用余弦衰减策略, 动量设置为 0.9, 权重衰减设置为 5e-4。输入网络中, 图像大小设置为 640×640, 并采用多尺度训练的策略, 每次迭代输入图像尺寸随机改为 320~992 之间、为 32 倍数的大小。

为公平对比 2 项改进点分别对 YOLOv5s 实例分割网络的贡献, 在裂纹测试集上进行消融实验, 通过各项改进点的组合设计了 4 组实验, 实验结果见表 2。表 2 中, 改进 1 为改进 mosaic 增强, 改进 2 为改进先验框。由此可以看出, 单独 2 项改进对于 4 项指标分别有所提升, 其中改进 1 提升最大, 分别提高

了1.9%、8.4%、0.8%和3.2%。2项改进组合之后,性能提升有明显叠加效果,最终分别提高了2.2%、10.4%、1.4%和5.7%,其中关于阈值50-95的提升更

明显,具体表现就是检测效果更加精准。由于先验知识改进并未修改网络结构,因此改进后的YOLOv5s实例分割网络仍保持了118 FPS的检测速度。

表2 消融实验数据

YOLOv5s	改进1	改进2	mAP_{50}^{bbox}	mAP_{50-95}^{bbox}	mAP_{50}^{mask}	mAP_{50-95}^{mask}	%
✓			95.1	74.5	95.0	64.4	
✓	✓		97.0	82.1	95.8	67.6	
✓		✓	96.3	78.1	95.2	65.3	
✓	✓	✓	97.3	84.9	96.4	70.1	

3.5 改进前后检测效果对比

为更直观地展示改进前后网络模型的实际检测效果,分别使用类数据集图像和数据集外图像进行检测,检测效果如图10、图11所示。通过改进前后

的检测效果对比,可以明显看出,改进后的网络模型具有更高的检测精度,漏检率有明显下降,并且对于数据集外的复杂形态裂纹图像也有着不错的泛化性。

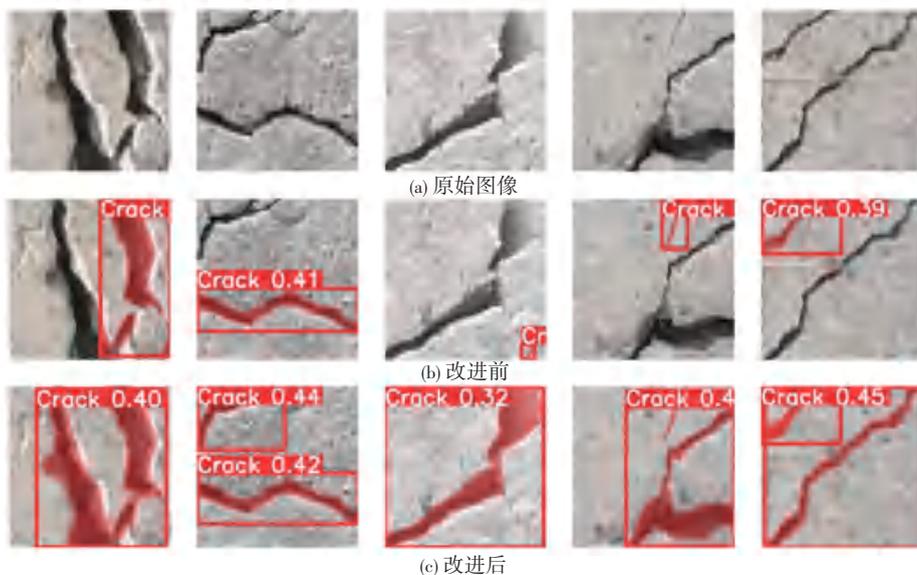


图10 类数据集图像检测效果

Fig. 10 Image detection effect of similar data set

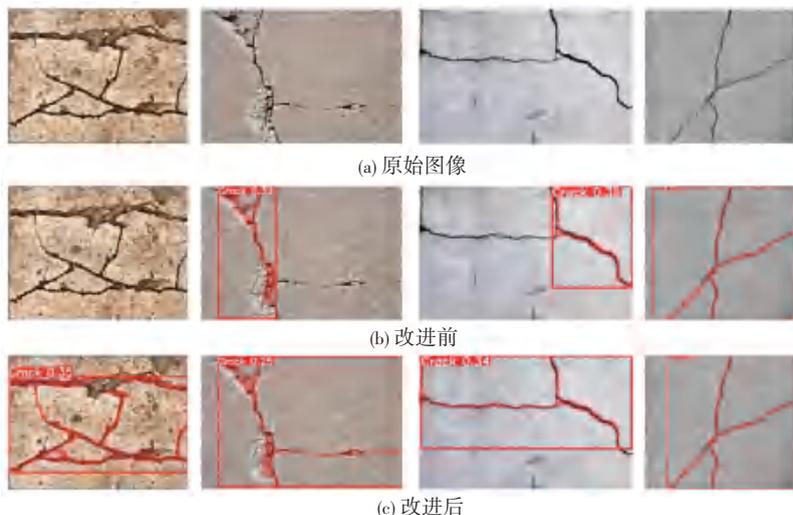


图11 数据集外图像检测效果

Fig. 11 Image detection effect outside the data set

4 结束语

实例分割是实现裂纹危害性评估的关键技术, 本文基于深度学习对裂纹实例分割进行研究。通过对数据增强和先验框两部分的先验知识改进, 研究中数据增强针对 mosaic 增强产生的错误样本使用裂纹图像分类模型进行过滤, 同时针对裂纹的连接性, 使用膨胀和腐蚀操作实现裂纹目标连接的标签信息合并; 先验框针对 K-means 算法容易收敛至局部最优解的问题, 提出使用 K-means++ 算法, 并使用更适合先验框收敛的交并比聚类标准, 最终在 YOLOv5s 实例分割网络上进行消融实验。实验验证 2 项改进可以有效提高先验知识质量, 达到了深度学习中实例分割网络模型的检测精度, 并具有不错的泛化性。

参考文献

- [1] ADHIKARI R S, MOSELHI O, BAGCHI A. Image - based retrieval of concrete crack properties for bridge inspection [J]. *Automation in Construction*, 2014, 39: 180-194.
- [2] SILVA W R L, LUCENA D S. Concrete cracks detection based on deep learning image classification [C]//*Proceedings of the 18th International Conference on Experimental Mechanics (ICEM18)*. Brussels, Belgium;dblp, 2018, 2(8):489.
- [3] MANDAL V, UONG L, ADU-GYAMFI Y. Automated road crack detection using deep convolutional neural networks [C]//*IEEE International Conference on Big Data (Big Data)*. Piscataway, NJ;IEEE, 2018: 5212-5215.
- [4] ARMAN M S, HASAN M M, SADIA F, et al. Detection and classification of road damage using R-CNN and faster R-CNN: A deep learning approach [C]//*Cyber Security and Computer Science: Second EAI International Conference*. Cham;Springer, 2020: 730-741.
- [5] TAN Chengjun, UDDIN N, MOHAMMED Y M. Deep learning-based crack detection using mask R-CNN technique [C]//*9th International Conference on Structural Health Monitoring of Intelligent Infrastructure*. Missouri, USA;dblp, 2019:1-7.
- [6] 朱苏雅, 杜建超, 李云松, 等. 采用 U-Net 卷积网络的桥梁裂纹检测方法 [J]. *西安电子科技大学学报*, 2019, 46(4):35-42.
- [7] 张继凯, 赵君, 张然, 等. 深度学习的图像实例分割方法综述 [J]. *小型微型计算机系统*, 2021, 42(1):161-171.
- [8] 朱晓慧, 钱丽萍, 傅伟. 图像数据增强技术研究综述 [J]. *软件导刊*, 2021, 20(5):230-236.
- [9] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection [J]. *arXiv preprint arXiv*, 2004. 10934, 2020.
- [10] LIU Wei, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]//*14th European Conference on Computer Vision (ECCV2016)*. Cham;Springer, 2016: 21-37.
- [11] ZHENG Zhaohui, WANG Ping, LIU Wei, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA;AAAI, 2020: 12993-13000.
- [12] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. *arXiv preprint arXiv*, 1804. 02767, 2018.
- [13] TAN Mingxing, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]//*International Conference on Machine Learning*. Florida, USA;PMLR, 2019: 6105-6114.
- [14] MACQUEEN J. Classification and analysis of multivariate observations [C]//*Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. Los Angeles, USA: University of California, 1967: 281-297.
- [15] ARTHUR D, VASSILVITSKII S. K-means++ the advantages of careful seeding [C]//*Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans, USA: ACM, 2007: 1027-1035.
- [16] ÖZGENEL Ç F, GÖNENÇ S A. Performance comparison of pretrained convolutional neural networks on crack detection in buildings [C]//*35th International Symposium on Automation and Robotics in Construction (ISARC 2018)*. Berlin, Germany;dblp, 2018: 693-700.
- [17] ZHANG Lei, YANG Fan, ZHANG D, et al. Road crack detection using deep convolutional neural network [C]//*IEEE International Conference on Image Processing*. Piscataway, NJ: IEEE, 2016:3708-3712.