

改进 RetinaNet 的水泥路面露骨病害检测^①



石 丽, 裴莉莉, 陈 昊, 李 伟, 袁 博, 冯笑然

(长安大学 信息工程学院, 西安 710064)

通信作者: 李 伟, E-mail: grandy@chd.edu.cn

摘 要: 露骨病害是水泥路面常见的病害之一, 会严重影响路面使用年限和行车安全性能. 因此, 及时对露骨病害进行检测与修复十分重要. 针对传统的人工检测方法检测精度不高且检测效率低的问题, 本文提出一种基于改进 RetinaNet 模型的水泥路面露骨病害检测方法. 首先对人工和检测车采集露骨病害图像进行滤波、去噪等预处理操作, 构建模型训练数据集; 然后将 SE Net 结构嵌入特征提取网络, 并改进特征金字塔网络; 最后实现改进的 RetinaNet 对水泥路面露骨病害的检测. 结果表明, 改进的 RetinaNet 模型在露骨病害检测精度上比原模型提高了 4.9%, 达到 98.9%. 且在相同测试数据上, 相比 Faster R-CNN、SSD 和 YOLOv3 方法, 检测效果均提升显著.

关键词: 道路检测; 水泥露骨病害; 深度学习; 二值化; 改进 RetinaNet

引用格式: 石丽, 裴莉莉, 陈昊, 李伟, 袁博, 冯笑然. 改进 RetinaNet 的水泥路面露骨病害检测. 计算机系统应用, 2022, 31(4): 352-359. <http://www.c-s-a.org.cn/1003-3254/8436.html>

Detection of Exposed Distress of Cement Pavement Based on Improved RetinaNet

SHI Li, PEI Li-Li, CHEN Hao, LI Wei, YUAN Bo, FENG Xiao-Ran

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: Exposed distress is one of the common diseases of cement pavements, which seriously affects pavement service life and driving safety. Therefore, it is very important to detect and repair exposed distress in time. Traditional manual detection methods are low in both detection accuracy and detection efficiency. In view of this, our study proposes a method for detecting exposed defects of cement pavements based on an improved RetinaNet model. Firstly, preprocessing operations such as filtering and denoising are carried out on the exposed distress images collected by manual and inspection vehicles, and the model training data set is constructed. Then the SE Net structure is embedded into the feature extraction network, and the feature pyramid network is improved. Finally, the detection of exposed distress of cement pavements is realized by the improved RetinaNet. The results show that the improved RetinaNet model enhances the detection accuracy of exposed distress by 4.9% compared with the original model, which reaches 98.9%. In comparison with Faster R-CNN, SSD and YOLOv3 methods, the model significantly improved the detection effect for the same test data.

Key words: road detection; cement exposed distress; deep learning; binarization; improved RetinaNet

根据我国 2013-2030 年的交通发展规划, 未来公路的规模为 550 万~600 万公里. 其中, 在 460 万~520 万公里乡村公路中, 水泥混凝土路面占比达到 84%~87%. 露骨为水泥混凝土路面的常见病害, 路面出现露骨病害不仅会影响行车的安全性, 继续扩展还会形成

坑槽, 严重影响路面的使用年限. 尽早对露骨病害检测并加以养护, 可以延长路面的使用寿命, 降低路面维护成本. 目前露骨病害检测主要依靠人工识别, 这种方法不仅耗费大量人力物力, 而且检测结果受主观判断影响, 容易发生误检漏检的情况. 因此, 如何快速自动化

① 基金项目: 国家自然科学基金面上项目 (51978071); 中央高校卓越青年项目 (300102249301); 中央高校领军人才项目 (300102249306)

收稿时间: 2021-06-05; 修改时间: 2021-07-07, 2021-07-20; 采用时间: 2021-08-13; csa 在线出版时间: 2022-03-22

检测路面露骨病害已经成为亟待解决的问题。

随着深度学习的广泛应用,目标检测方法已经成为卷积神经网络领域的热点话题^[1,2]。基于卷积神经网络的方法主要包括两大类:一类是 one-stage 一阶段目标检测方法,此方法将整个目标检测过程融合成一个端到端的任务,只对图像进行一次处理即可得到目标的类别和位置信息,其典型的算法有 SSD^[3]、YOLOv3^[4]、FSSD^[5]等;另一类是基于候选区域的 two-stage 二阶段目标检测方法,此方法将生成候选框和候选框分类回归分成 2 个阶段,首先在图像上生成若干可能包含目标的候选框,其次使用卷积神经网络对候选框进行特征提取,最后使用卷积神经网络对目标区域进行分类和回归,其典型的算法有 Fast R-CNN^[6]、Faster R-CNN^[7]、SPP-Net^[8]、Mask R-CNN^[9]等。与二阶段目标检测算法相比,一阶段目标检测算法效率更高,但是在特征提取过程中,容易产生大量负样本,导致在训练时正负样本失衡,准确率和定位精度较低。Lin 等^[10]提出的 RetinaNet 模型利用 Focal Loss 函数解决样本失衡的问题,在 COCO 测试集上其检测精度高于当时最先进的二阶段检测算法。谭章禄等^[11]利用 RetinaNet 模型在煤矿监控中实现自动检测及提取人员信息,整体性能满足煤矿监控的需求。Zhang 等^[12]引入 Octave 卷积和加权特征金字塔结构以提高 RetinaNet 对车辆的检测性能,结果证明该方法对不同场景、不同规模的车辆目标具有良好的检测结果。姚青等^[13]选择组归一化作为 RetinaNet 的归一化方法,改进特征金字塔网络结构,改进后的模型对稻纵卷叶螟和二化螟为害状区域的平均检测精度为 93.76%。吴华运等^[14]对 RetinaNet 特征金字塔网络的特征融合过程进行优化,增加语义特征提取模块,在特征提取网络 ResNet 网络中添加了

膨胀卷积瓶颈模块,对医药空瓶表面气泡缺陷检测的准确率为 99.72%。张涛等^[15]在 RetinaNet 模型基础上,建立用来融合深层特征的 HFF 模块和融合浅层特征 LFF 模块,在 PASCAL VOC 和 MSCOCO 数据集上较原模型精度有提升。邓小桐等^[16]针对迷彩伪装人员目标特性,将空间注意力和通道注意力机制嵌入到 RetinaNet 网络结构中,并设计基于定位置信信息的检测框过滤算法,比原模型检测精度提升了 8.7 个百分点,达到 93.1%。

针对目前缺少关于水泥路面露骨病害检测方法以及人工检测费时、费力 and 检测精度不高的问题,本文以 RetinaNet 为基础,以 SE-ResNet101 作为特征提取网络,对特征金字塔结构进行改进,提出一种改进的 RetinaNet 模型,实现对水泥路面露骨病害的准确且快速地识别。

1 改进的 RetinaNet 模型

1.1 RetinaNet

RetinaNet 是 Lin 等在 2018 年提出的一阶段目标检测器,其结构如图 1 所示。RetinaNet 主要由 3 部分构成:如图 1(a) 为特征提取网络,实现对图像的特征提取操作,目前应用较广泛的特征提取网络有残差网络 ResNet^[17]、VGG-Net^[18]、GoogleNet^[19]等;如图 1(b) 为特征金字塔网络 (feature pyramid networks, FPN)^[20], 其将特征提取网络提取的特征图进行重新组合,形成具有各种尺度的相同维度特征,实现特征的精细化提取;如图 1(c) 和图 1(d) 表示用来分类和定位的分类和回归子网络 (fully convolutional networks, FCN)^[21], 分类子网络用于卷积对象分类,回归子网络用于卷积边界框回归。

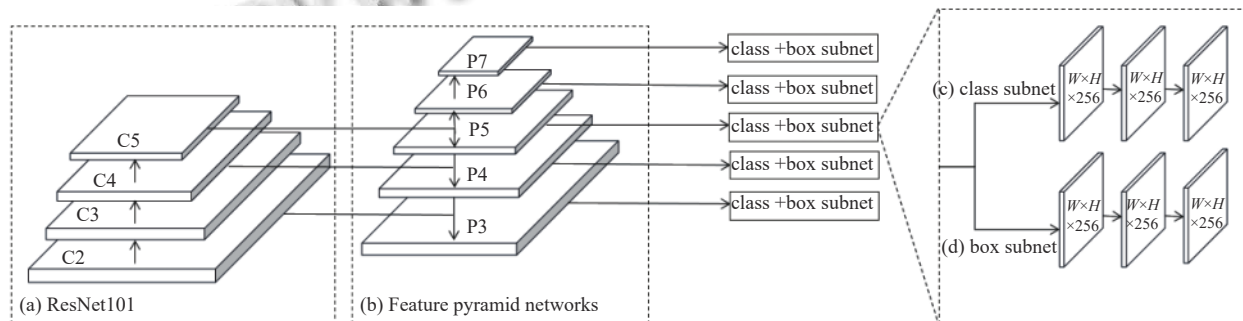


图 1 改进的 RetinaNet 结构图

1.1.1 特征提取网络

随着神经网络深度的增加,特征的提取抽象程度

就越高,理论上也就能产生更好的拟合效果。但是,随之会出现梯度弥散、梯度爆炸、学习率下降等问题。

为解决此问题, He 等^[17] 提出 ResNet 残差网络, 如图 2 为 ResNet 的残差单元. $H(x)=F(x)+x$ 表示最优解映射, 其中 x 为输入, $F(x)$ 为残差映射. 其采用快捷连接跨越的思想, 使用等值映射方式, 跳过中间层, 直接将一部分原始数据直接输出传到下一层. 此结构不会增加网络参数, 对网络计算的复杂度也不会有影响, 有效的解决网络的梯度爆炸、梯度弥散的问题.

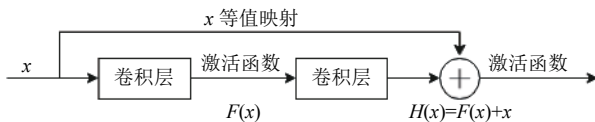


图 2 残差网络结构图

1.1.2 特征金字塔网络 (FPN)

如图 1(b) 为特征金字塔网络, 在 RetinaNet 模型中, 特征金字塔网络使得高层语义特征和低层位置特征相结合, 增强卷积神经网络的特征提取能力. 对 ResNet101 特征提取网络输出的特征图 C2~C5 进行上采样和横向连接操作, 产生特征融合后的特征金字塔网络结构 P3~P5. 具体操作方法为首先对 C5 使用步长为 1 的 1×1 卷积核卷积得 P5; 其次对 C4 使用步长为 1 的 1×1 卷积核卷积, 并相相对 P5 进行两倍上采样结果; 最后对上一步操作得到的特征图使用步长为 1 的 3×3 卷积核卷积得到 P4. 以此类推, 得到 P3.

1.1.3 分类回归子网络与损失函数

RetinaNet 网络借鉴 RPN (region proposal network) 中的 Anchor 的思想: 当卷积核滑动到特征图某位置时, 滑动窗口的中心点映射到原图上形成一个映射点, 此映射点对应一个尺度和纵横比, 成为一个 Anchor. 本文在 Anchor 生成时使用了 3 种不同尺寸 $\{2^0, 2^{\frac{1}{3}}, 2^{\frac{2}{3}}\}$ 和 3 种不同纵横比 $\{1:2, 1:1, 2:1\}$, 即可产生 9 种 Anchor. 如图 1(c) 和图 1(d) 为分类和回归子网络, 对生成的所有 Anchor 进行分类和回归.

分类子网络为每个 Anchor 预测露骨病害的概率. 将特征金字塔中某一层特征图与激励函数为 ReLU 的全卷积网络相连接, 使用 Sigmoid 函数预测水泥路面露骨病害的概率. 在分类子网络中, 检测器一般通过交叉熵损失函数 (CE loss) 判定预测值与真实值的差异程度. 目前普遍应用的 CE loss 如式 (1) 所示:

$$CE(p, y) = CE(p_t) = -\log p^t \quad (1)$$

其中, y 为聚焦参数; p_t 的定义如式 (2):

$$p_t = \begin{cases} p, & y = 1, p \in [0, 1] \\ 1 - p, & y \neq 1, p \in [0, 1] \end{cases} \quad (2)$$

在进行露骨病害检测时, 由于受路面车道线和阴影的影响, 出现了类不平衡和难易实例的检测问题. 本文在 CE loss 的基础上, 为解决中正负样本区域严重失衡的问题, 将正负样本增加权重因子 α , 降低负样本的权重, 增加正样本权重, 加权 CE loss 的定义如式 (3); 为解决难易样本区域失衡的问题, 增加调制系数 $(1 - p_t)$, 减少易检测的样本权重, 使得模型在训练时更加专注难检测的样本, 尺度动态可调损失函数 Focal Loss 定义式 (4).

$$CE(p_t) = -\alpha_t \log p^t \quad (3)$$

$$FL(p_t) = -\alpha_t (1 - p_t)^\beta \log p^t \quad (4)$$

其中, β 为聚集参数, 且 $\beta \geq 0, \alpha_t$ 定义如式 (5):

$$\alpha_t = \begin{cases} \alpha, & y = 1, \alpha \in [0, 1] \\ 1 - \alpha, & y \neq 1, \alpha \in [0, 1] \end{cases} \quad (5)$$

回归子网络与分类子结构相类似. 在回归子网络中, RetinaNet 使用了 Smooth L1 损失函数. 假定 p 代表预测区域与 Anchor 之间的变换关系, t 代表真实区域与 Anchor 之间的变换关系; \overline{p}_i^m 则表示第 i 个预测区域与 Anchor 之间的偏差, \overline{t}_i^m 第 i 个真实区域与 Anchor 之间的偏差; $m \in \{x, y, w, h\}$, (x, y) 表示区域中心点的坐标; (w, h) 表示区域的宽和高. d_j^m 为第 j 类缺陷的真实区域, t_i^m 为第 i 个预测区域和 Anchor.

$$\begin{cases} \overline{p}_i^x = (p_i^x - d_j^x) / d_j^w, & \overline{p}_i^y = (p_i^y - d_j^y) / d_j^h \\ \overline{p}_i^w = \log(p_i^w / d_j^w), & \overline{p}_i^h = \log(p_i^h / d_j^h) \\ \overline{t}_i^x = (t_i^x - d_j^x) / d_j^w, & \overline{t}_i^y = (t_i^y - d_j^y) / d_j^h \\ \overline{t}_i^w = \log(t_i^w / d_j^w), & \overline{t}_i^h = \log(t_i^h / d_j^h) \end{cases} \quad (6)$$

定义 Smooth L1 如式 (7):

$$SL1 = \begin{cases} 0.5(\overline{p}_i^m - \overline{t}_i^m)^2, & \text{if } |\overline{p}_i^m - \overline{t}_i^m| < 1 \\ |\overline{p}_i^m - \overline{t}_i^m| - 0.5, & \text{otherwise} \end{cases} \quad (7)$$

1.2 改进的 RetinaNet

为了更准确的获得水泥路面露骨病害的原始图像特征, 本文使用 ResNet101 作为特征提取网络. 由于路面病害的破损程度不同, 检测图像中的目标尺寸有差异, 为扩大感受野, 在 FPN 特征金字塔上做了改进. 并借鉴了 SE Net 的设计思想, 对残差网络进行改进.

1.2.1 ResNet101

本文采用深度残差网络 ResNet101 作为特征提取网络,其结构如图 3 所示.整个网络包含 101 个卷积层,其中每个括号内的结构代表一个结构块,每个结构块中包含 3 层卷积.在 Conv2 中 3×3, maxpool, stride2 代表卷积核为 3×3, 步长为 2 的最大池化操作, 1×1, 64 代表卷积核大小为 1×1, 输出通道为 64 的卷积操作, 其余与之同理.

1.2.2 改进的 FPN

在 RetinaNet 模型中, FPN 通过自底向上、自顶向下、横向连接等多种方式将 ResNet 不同层的特征图

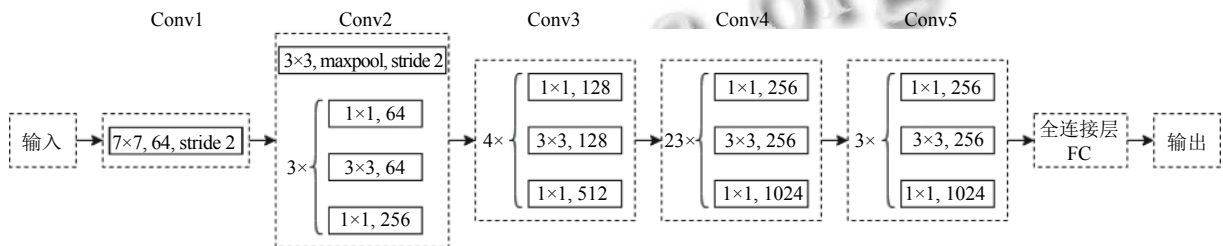


图 3 ResNet101 网络结构

1.2.3 基于 SE-ResNet101 的特征提取网络

图 4 为 SE-ResNet101 模块内部结构, SE-ResNet101 由 ResNet101 模块和 SE Net 模块两部分构成. SE Net (squeeze-and-excitation networks)^[22] 是基于加权特征图思想提出的一种网络结构, 在残差网络 ResNet101 中嵌入 SE, 促使神经网络更有针对性地提取目标特征, 以提高对目标的特征表达能力, 从而提升对露骨病害的聚焦.

SE Net 结构由压缩 Squeeze 和激励 Excitation 两个部分组成. 具体流程为: 首先 Squeeze 操作将 $C \times W \times H$ 的特征图通过全局平均池化压缩, 可得到 $1 \times 1 \times C$ 的全局信息描述符. 其次通过两个全连接层进行 Excitation 操作从而预测每个 channel 的重要性. 第一个全连接层输入为 $1 \times 1 \times C$, 输出 $1 \times 1 \times C/r$, r 为缩放参数; 第二个全连接输入为 $1 \times 1 \times C/r$, 输出为 $1 \times 1 \times C$. 然后使用 Sigmoid 激活函数输出一个原输入特征图各层权值的向量. 最后 Scale 操作将输出向量与原输入特征图相乘, 得到加权特征图, 来增强有效特征, 削弱低效或无效特征, 提取具有更强指向性的特征, 从而提高检测结果.

不同通道的特征进行一定的相乘相加运算, 使得加入注意力模块的网络学习到更多有用的图像信息, 也避免了网络学习不感兴趣区域的资源浪费, 增强网络的学习重点能力, 改善学习效率.

进行融合, 对小目标具有较好的检测效果. 本试验的数据集中, 露骨病害目标所占比例较大. 为了使模型有利于大目标图像的检测, 对本试验目标有更好的检测效果, 选择进一步增大模型的感受野, 具体操作为: 在图 1(b) P5 的基础上, 使用步长为 2 的 3×3 的卷积核卷积运算得到 P6、P7, 提升感受野的同时, 进一步对训练目标信息进行抽象, 获取更高维度的特征信息. 将更大感受野且更高维度信息整合后送入网络, 同时计算 class subnet 和 box subnet 的损失值, 进行反向梯度更新, 从而更加准确的进行下一阶段的特征提取操作, 提升整体检测性能.

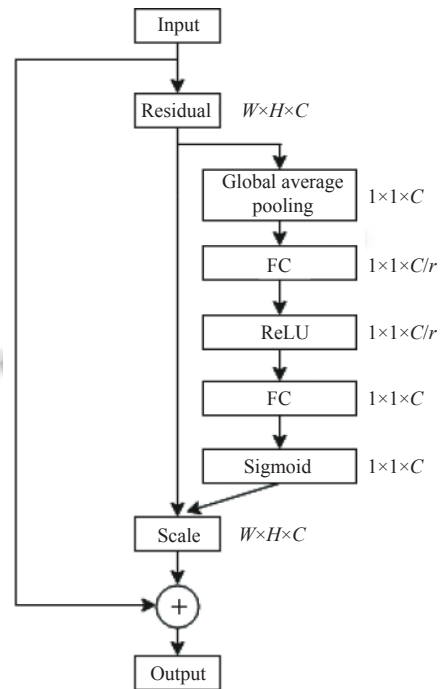


图 4 SE-ResNet101 模块内部结构

2 基于改进的 RetinaNet 模型的路面露骨病害目标检测

2.1 数据集构建

本文所使用的水泥路面露骨病害图像数据集由人

工手机拍摄及车载运动相机拍摄,采集区域主要是在中国陕西省西安市长安大学及其附近.最终拍摄数据集图像共506幅.部分露骨病害采集样本如图5所示.

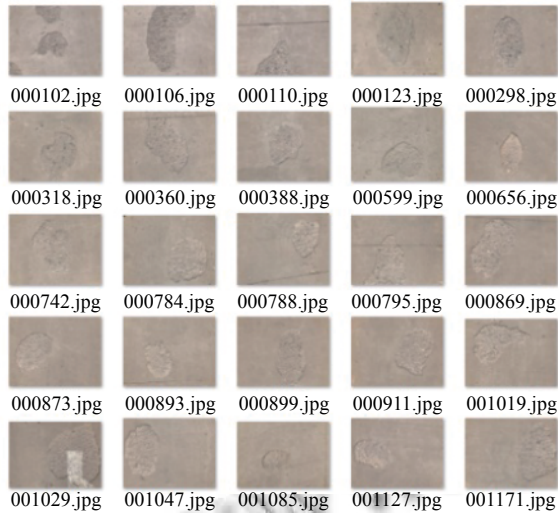


图5 部分水泥路面露骨病害图像

由图5可以看出,由于露骨病害图像源自不同拍摄设备、不同路段,因此图像背景十分复杂且样本数据集较少.基于上述情况,本文对数据集进行处理以提升其质量.首先为解决样本数据集数量不足的问题,本文采取镜像、旋转和调节像素等多种方式对数据集进行数据增广,最终样本的数量为1523.其次,为了尽可能减小背景及路况对检测的影响,需要对样本数据集进行预处理工作,具体流程图如图6所示.最后,使用Labellmg软件对数据集图像中的露骨病害目标进行标记,每个露骨病害目标对应于数据样本中位置注释的XML文件.

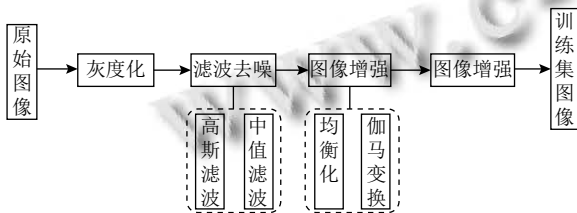


图6 露骨病害图像预处理流程图

2.2 数据集构建

本实验的操作系统为Ubuntu 14.04操作系统.具体的实验环境如表1所示.

从本文自制的1523个数据集中,选出90%即1370幅露骨病害样本进行训练,其余10%即153幅露骨病害样本进行测试.在分类训练集和测试集时,将图

像按照复杂度(阴影、车道线及其他异物)均匀地分给训练集和测试集以保证实验的有效性.模型训练过程中的参数设置如表2所示.

表1 实验环境

Name	Type
处理器	Intel(R) Core(TM) i7-7700K CPU @ 4.20 GHz
RAM	8.00 GB
GPU	GeForce GTX 1050 Ti
CUDA	8.0
Cudnn	7.5
OpenCV	3.7
操作系统	Ubuntu 14.04 64-bit
深度学习框架	Keras
编程语言	Python

表2 RetinaNet 模型训练参数

参数	值	参数	值
base_lr	0.001	momentum	0.9
steps	1000	weight_decay	0.0005
epochs	50	batchsize	2

2.3 实验评价指标

本文采用准确率 *Precision*、召回率 *Recall* 和检测精度 *AP* 对水泥路面露骨病害检测的效果进行评价,其计算公式如下:

$$\begin{cases} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \\ AP = \int_0^1 Precision(Recall) dRecall \end{cases} \quad (8)$$

式中, *Precision* 准确率表示正确检测为露骨病害目标在全部检测为露骨病害目标结果中的占比,其值越大说明误检的目标越少; *Recall* 表示检测正确的露骨病害目标在所有露骨病害目标的占比,其值越大说明漏检目标越少; *TP* 表示正确检测露骨病害目标数量, *FP* 表示误检为露骨病害目标的数量, *TN* 表示正确检测非露骨病害目标数量, *FN* 表示漏检的露骨病害目标数量. PR 曲线是以 *Precision* 作为纵坐标, *Recall* 作为横坐标,绘制的二维曲线, PR 曲线与坐标轴所围成的面积为平均准确率 (average precision, *AP*) 值. *AP* 值用来分析模型检测效果,衡量目标的检测精度. *AP* 值越大,说明其检测精度越高.

2.4 实验结果对比与分析

2.4.1 模型改进前后结果对比

为了验证本文方法改进后的 RetinaNet 对水泥路面露骨病害检测的有效性,分别对 RetinaNet 和改进后

的 RetinaNet 在测试集上的检测结果进行比较. 混淆矩阵如表 3. 由表 4 可知, 在检测精度方面, 改进后的 RetinaNet 比原 RetinaNet 的 AP 提升 4.9%; 在检测速度方面, 改进后的模型单张速度仅增加了 1.8 ms. 利用 Python 语言中的 Matplotlib 库绘制 PR 曲线如图 7, 可见改进后的模型整体性能得到了提升, 漏检率进一步的降低.

表 3 二分类混淆矩阵

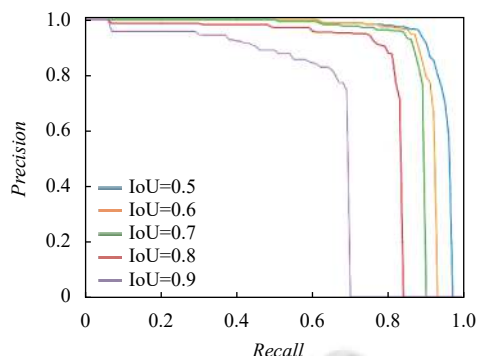
真实值	预测值	
	露骨病害	非露骨病害
露骨病害	真正类 (TP)	假正类 (FN)
非露骨病害	假反类 (FP)	真反类 (TN)

表 4 模型改进前后检测结果

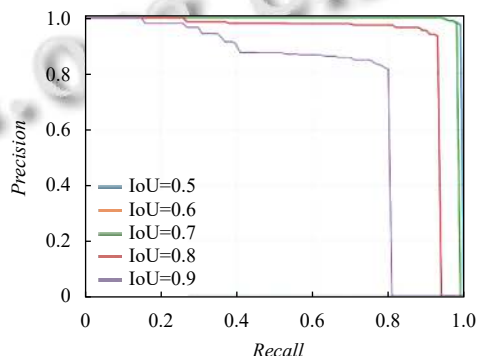
模型	AP (%)	单副检测速度 (ms)
RetinaNet	94.0	89.6
改进RetinaNet	98.9	91.4
对比	+4.9	-1.8

图 8 为 RetinaNet 改进前后目标检测算法对水泥里面露骨病害的识别效果的对比, 可以看出改进后 RetinaNet 算法相较于原算法有更好的检测效果.

根据改进后的 RetinaNet 输出结果与露骨图像实际标签, 得到混淆矩阵, 如表 5 所示.



(a) 原模型曲线图



(b) 改进后模型曲线图

图 7 RetinaNet 改进前后的 PR 曲线图



(a) RetinaNet 模型的识别结果



(b) 改进后 RetinaNet 模型的识别结果

图 8 RetinaNet 改进前后目标检测算法识别效果

表 5 露骨病害混淆矩阵

真实值	预测值	
	露骨病害	非露骨病害
露骨病害	152	1
非露骨病害	3	44

图 9 比较了使用 CE loss 和 Focal loss 的训练损失曲线, 从图 9 中可以看出, 在 50 轮的训练过程中, 本文使用的 Focal loss 函数相对于 CE loss 使得改进后的 RetinaNet 收敛速率更快. 在模型训练后期, 采用 Focal loss 在训练集上的损失值均低于 CE loss.

2.4.2 不同目标检测模型检测效果对比

为了进一步验证本文改进的 RetinaNet 模型对于水泥路面露骨病害检测的有效性,将本文模型与目前主流的目标检测模型 Faster R-CNN、SSD 以及 YOLOv3 对水泥路面露骨病害测试集进行检测对比,实验结果如图 10 所示.由图 10 可以看出,本文提出的改进 RetinaNet 模型相比于二阶目标检测模型 Faster R-CNN,单副检测速度减少 105.8 ms, AP 增加 4.1%;相比于二阶目标检测模型 SSD 和 YOLOv3,虽然检测速度略有增加,但检测精度增加 6.4% 和 5.2%.综合上述分析,本文提出的模型是一个对水泥路面露骨病害有效的检测模型.

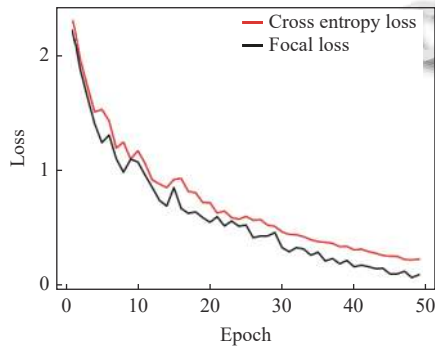


图9 使用 CE loss 和 Focal loss 的 loss 曲线对比图

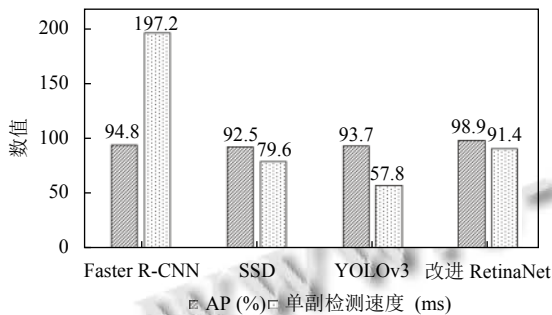


图10 不同模型检测效果对比

3 结论

本文以水泥路面露骨病害为研究对象,基于改进的 RetinaNet 网络对露骨病害自动化检测问题展开研究,主要结论如下:

(1) 采集路面露骨病害的图像,并进行灰度化、滤波去噪、图像增强等预处理操作,构建水泥路面露骨病害数据集;

(2) 在 RetinaNet 模型的基础上,利用 SE Net 的思想对特征提取网络 ResNet101 进行改进,增强特征图的特征表达能力;增加特征金字塔的层数,充分扩大目标的感受野完成对路面露骨病害图像的检测;

(3) 通过与 Faster R-CNN、SSD 和 YOLOv3 对比,发现改进后的 RetinaNet 模型对露骨病害的检测精度有显著提升,达到 98.9%.实现对露骨病害的准确检测,能够替代繁琐、效率低的人工检测方法.

本文提出的 RetinaNet 改进模型对水泥路面露骨病害自动化检测领域具有重要意义.但改进后的 RetinaNet 模型相比于原模型单幅露骨病害检测时间增加 1.8 ms,因此在保证识别精度的同时提升网络检测速度是之后研究工作的重点.

参考文献

- 孙朝云,裴莉莉,李伟,等.基于改进 Faster R-CNN 的路面灌封裂缝检测方法.华南理工大学学报(自然科学版),2020,48(2): 84-93.
- Feng XR, Xiao LY, Li W, *et al.* Pavement crack detection and segmentation method based on improved deep learning fusion model. *Mathematical Problems in Engineering*, 2020, 2020: 8515213.
- Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 21-37.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. *arXiv: 1804.02767*, 2018.
- Li ZX, Zhou FQ. FSSD: Feature fusion single shot multibox detector. *arXiv: 1712.00960*, 2017.
- Girshick R. Fast R-CNN. *Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 1440-1448.
- Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149. [doi: 10.1109/TPAMI.2016.2577031]
- He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916. [doi: 10.1109/TPAMI.2015.2389824]
- He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *Proceedings of 2017 IEEE International Conference on*

- Computer Vision. Venice: IEEE, 2017. 2980–2988.
- 10 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318–327.
 - 11 谭章禄, 陈孝慈. RetinaNet 图像识别技术在煤矿目标监测领域的应用研究. *矿业安全与环保*, 2020, 47(5): 65–70, 76. [doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826)]
 - 12 Zhang LY, Wang HT, Wang XY, *et al.* Vehicle object detection based on improved RetinaNet. *Journal of Physics: Conference Series*, 2021, 1757: 012070.
 - 13 姚青, 谷嘉乐, 吕军, 等. 改进 RetinaNet 的水稻冠层害虫为害状自动检测模型. *农业工程学报*, 2020, 36(15): 182–188.
 - 14 吴华运, 任德均, 吕义钊, 等. 基于改进的 RetinaNet 医药空瓶表面气泡检测. *四川大学学报 (自然科学版)*, 2020, 57(6): 1090–1095.
 - 15 张涛, 张乐. 一种基于多尺度特征融合的目标检测算法. *激光与光电子学进展*, 2021, 58(2): 0215003.16.
 - 16 邓小桐, 曹铁勇, 方正, 等. 改进 RetinaNet 的伪装人员检测方法研究. *计算机工程与应用*, 2021, 57(5): 190–196.
 - 17 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 770–778.
 - 18 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv: 1409.1556*, 2014.
 - 19 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 1–9.
 - 20 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 936–944.
 - 21 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. [doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683)]
 - 22 Hu J, Shen L, Albanie S, *et al.* Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011–2023. [doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372)]