基于 Xception 和 SA 的 YOLOv5 建筑裂缝检测方法

卞长庚,郝万君,马文琪
(苏州科技大学 电子与信息工程学院,江苏 苏州 215009)

摘 要:裂缝检测对于建筑的维修和加固、延长其使用寿命具有重要意义。针对建筑裂缝种类多和尺寸小造成裂缝检测精度低、速度慢的问题,提出了一种改进的 YOLOv5 裂缝检测算法,在提高检测裂缝精度的同时也提升了检测裂缝的速度。首先,引入轻量级网络 Xception 对主干网络轻量化,减少主干网络参数量以提升检测裂缝的速度;其次,使用空洞空间金字塔池化 ASPP(Atrous Spatial Pyramid Pooling)模块替换 SPP(Spatial Pyramid Pooling)模块,扩大感受野范围,加强主干网络提取裂缝特征的能力,避免因对主干网络轻量化而造成检测裂缝的精度降低;最后,添加 SA(Shuffle Attention)注意力机制,进一步加强网络提取裂缝特征的能力,提高裂缝检测的精度。通过在自制数据集上进行的实验表明,改进的算法 mAP 比原算法提高了 1.6%,速度为 50.8 f/s,比原算法提高了 2.7 f/s,满足建筑裂缝检测的精度和实时性要求,同时将改进算法与 Faster R-CNN、Mobile-SSD、YOLOv4-tiny 等算法进行对比,证明了该算法的优越性,更适合部署到硬件平台上。

关键词:裂缝检测;Xception;空洞空间金字塔池化;Shuffle 注意力
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2023)08-0159-06
doi:10.3969/j.issn.1673-629X.2023.08.023

YOLOv5 Building Crack Detection Method Using Xception and SA

BIAN Chang-geng, HAO Wan-jun, MA Wen-qi
(School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China)

Abstract: Crack detection is of great significance for the maintenance and reinforcement of buildings and for extending their service life. Aiming at the problems of low accuracy and slow speed of crack detection caused by the variety and small size of building cracks, an improved YOLOv5 crack detection algorithm is proposed, which improves the accuracy of crack detection and the speed of crack detection. Firstly, the lightweight network Xception is introduced to lighten the backbone network and reduce the number of backbone network parameters, thus improving the speed of crack detection. Secondly, the SPP (Spatial Pyramid Pooling) module is replaced with the ASPP (Atrous Spatial Pyramid Pooling) module to expand the receptive field, strengthen the ability of backbone network to extract fracture features, and avoid reducing the accuracy of crack detection due to the lightweight of backbone network. Finally, SA (Shuffle Attention) attention mechanism is added to further enhance the ability of network to extract crack features and improve the accuracy of crack detection. Through experiments on self-made datasets, the mAP of improved algorithm is 1.6% higher than that of the original algorithm, with a speed of 50.8 f/s, which is 2.7 f/s higher than that of the original algorithm, and meets the accuracy and real-time requirements of building crack detection. At the same time, the improved algorithm is compared with Faster R-CNN, Mobile-SSD, YOLOv4-tiny and other algorithms, which proves its superiority and is more suitable for deployment on hardware platforms.

Key words: crack detection; Xception; atrous spatial pyramid pooling; Shuffle attention

0 引 言

混凝土建筑由于自然环境的长期腐蚀、长时间负载过重以及施工不当等原因会产生裂缝,这会严重影响其以后的使用,并且会对人员的生命和财产造成安全隐患[1]。因此,尽早检测出裂缝并对建筑进行加固和维修,具有重要意义。

传统的裂缝检测方法依赖于人工操作,主观性大,不仅效率低下,且无法保证工作人员的安全[2]。而基于图像的裂缝识别算法,虽然替代了人工,但只是提取裂缝的一些浅层特征,检测的效果较差,且容易受到光照等外界因素的影响[3]。

近些年来,深度学习的发展为建筑裂缝检测提供

收稿日期:2022-10-14 **修回日期**:2023-02-16
基金项目:国家自然科学基金资助项目(51477109)
作者简介:卞长庚(1998-),男,硕士研究生,研究方向为深度学习、目标检测;通信作者:郝万君(1965-),男,博士,教授,研究方向为机器视觉。

了新的方法。基于深度学习的目标检测算法不依赖于人工,而是通过自主学习便可直接对输入的图像进行检测。目前,目标检测算法可分为两类,一类是一阶段的算法,如 SSD[4]、YOLO[5]等,另一类是二阶段的算法,如 Faster R-CNN[6]等。一阶段的算法虽检测速度较快,但检测精度不高,相反,二阶段的算法检测精度较高,但是推理速度较慢。随着深度学习的发展,众多研究者开始将目标检测算法应用于建筑裂缝的检测。

孙朝云等[7]采用 VGG 网络替换 Faster R-CNN 的主干特征提取网络,虽然提高了裂缝检测的精度,但模型复杂,速度较慢。李鹏程等[8]采用 MobileNet 对 SSD 主干做轻量化处理,提升了感受野范围,具有较快的检测速度,但检测裂缝的精度较低,以上算法无法兼顾检测精度和速度。而随着 YOLO 系列算法的发展,如今 YOLO 算法检测精度和速度都得到了很大提升。杨富强等[9]采用广义交并比对 YOLOv3 损失函数进行改进,并采用迁移学习进行训练,模型具有较好的检测精度。郝巨鸣等[10]在 YOLOv4 中引入 Ghost 模块,并加入高效注意力机制 ECA,降低了模型的复杂度,也提高了检测精度和速度。然而嵌入式设备的发展对网络模型大小、检测的精度和速度都提出了更高的要求。

YOLOv5 各方面的性能都优于之前的算法。为此,基于上述研究,该文在 YOLOv5 的基础上进一步改进,采用轻量级网络 Xception[11]对主干网络轻量化,同时加入 ASPP[12]模块扩大特征感受野范围,加强主干网络特征提取能力,最后在网络的颈部加入 Shuffle Attention[13]注意力机制,进一步提高裂缝检测的精度,通过在自行构建的数据集上进行训练和实验,验证了改进算法的有效性。

1 YOLOv5 网络模型

对于建筑裂缝的检测,考虑其精度和速度的要求,选取 YOLOv5s 作为检测的网络模型,其网络结构如图 1 所示。

YOLOv5 的网络结构主要由 Backbone、Neck、Head 共 3 部分组成,其输入端主要对图片进行缩放、数据增强以及自适应锚框计算。

主干 Backbone 主要是由 Focus、CBS、C3、SPP 等模块组成。其中,Focus 模块主要对图片进行切片操作,将原先图片的 RGB3 个通道扩展成 12 个通道;CBS 是基本的卷积层,由 BN 以及 SiLU 构成;C3 模块是由 CBS 模块与残差模块连接构成,在不增加网络深度的情况下实现了对残差特征的学习;SPP 模块对不同大小的卷积核进行池化、融合,提取图片重要特征。

颈部 Neck 是由 FPN 和 PAN 组成,FPN 采用上采样,将语义特征从高层传递到低层,而 PAN 采用下采样,将定位特征从低层传递到高层,提高了特征融合的能力。

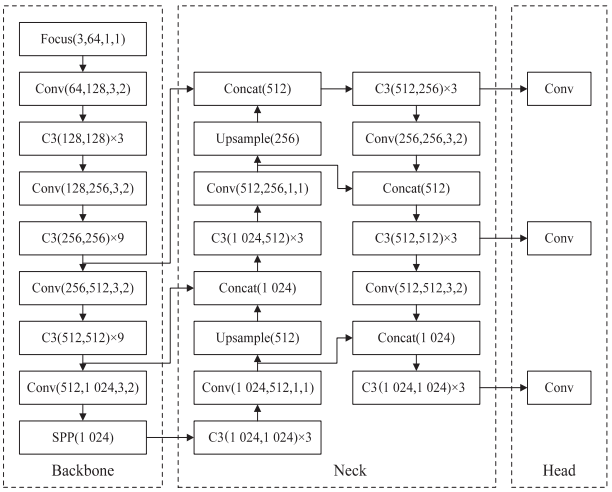

图 1 YOLOv5 网络结构

输出端 Head 负责预测,输出对象的概率、置信度和检测框的位置,YOLOv5 有 3 层负责预测,分别对大小不同的物体进行检测。

2 改进的 YOLOv5 网络模型

2.1 轻量级网络 Xception

由于 YOLOv5 的主干网络 CSPDarknet 中含有较多的普通卷积层,这加大了计算量,影响了检测速度,因此该文使用 Xception 轻量化网络作为 YOLOv5 的主干特征提取网络。

Xception 网络识别性能优于 CSPDarknet,同时能减少卷积运算,并且在准确率、效率等方面已经超过了 MobileNet 等轻量化网络。Xception 是 Inception 网络的“极端”版本,该网络首先采用 1×1 卷积寻找跨通道相关性,然后分别映射每个输出通道的空间相关性,如图 2 所示。

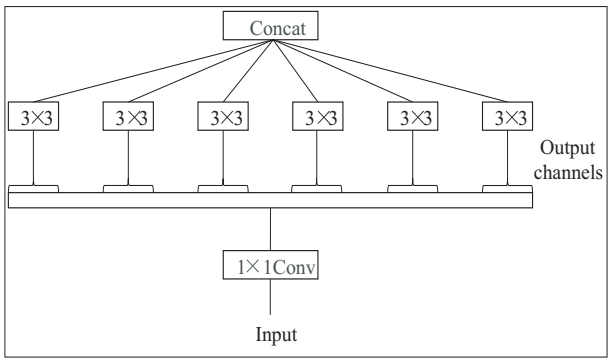

图 2 Xception 模块

Xception 网络由 3 个部分组成,输入图片的尺寸为 299×299×3,首先在 Entry flow 中经过两次普通卷积,然后使用深度可分离卷积和残差结构的组合得到 19×19×728 的特征图,所得特征图输入到 Middle flow

中继续使用深度可分离卷积进行特征提取,重复八次,最后输入到 Exit flow 中处理,最终得到 2 048 维度的向量。

为直观反映 Xception 和 CSPDarknet 的性能,在 YOLOv5s 网络中分别加入两种网络进行测试,两者的参数量对比如表 1 所示。可以看出 Xception 网络层数和参数量更少,计算速度更快,能够实现网络轻量化。

表 1　主干网络性能对比

模型	层数	参数量	GFLOPS
CSPDarknet	272	7 233 211	16.4
Xception	133	2 049 132	8.9

2.2　空洞空间金字塔池化 ASPP

YOLOv5 采用的是空间金字塔池化 SPP,它将输入的特征并行通过不同大小的池化核,变成固定大小的向量,再进行融合。

由于对主干网络进行轻量化,使得参数减少,虽然提高了检测速度,但导致检测的精度降低。为了加强主干网络特征提取能力,该文采用 ASPP(Atrous Spatial Pyramid Pooling)模块替换 SPP 模块。ASPP 是在 SPP 的基础上引入了空洞卷积,当提取的特征具有较大的感受野时,特征图的分辨率会降低,两者之间是矛盾的,空洞卷积很好地解决了这个矛盾。与普通卷积不同的是,空洞卷积引入了“扩张率(rate)”参数,它表示卷积核各个点之间的间隔数量,其原理如图 3 所示。

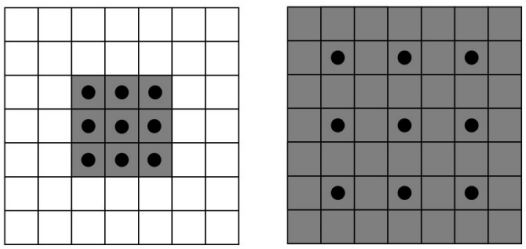

图 3　普通卷积与空洞卷积

图 3 左侧是普通卷积,扩张率为 1,感受野为 3×3,而右侧是空洞卷积,扩张率为 2,其感受野为 7×7,依此类推。空洞卷积的优势在于不进行池化操作,扩大了感受野,使得每个卷积的输出信息范围都较大,ASPP 原理如图 4 所示,与 SPP 相比,该模块使用多个并行的空洞卷积,单独处理不同大小的特征,然后再进行融合,在扩大了感受野的同时,提高了主干网络特征提取的能力,也提高了模型的检测速度。

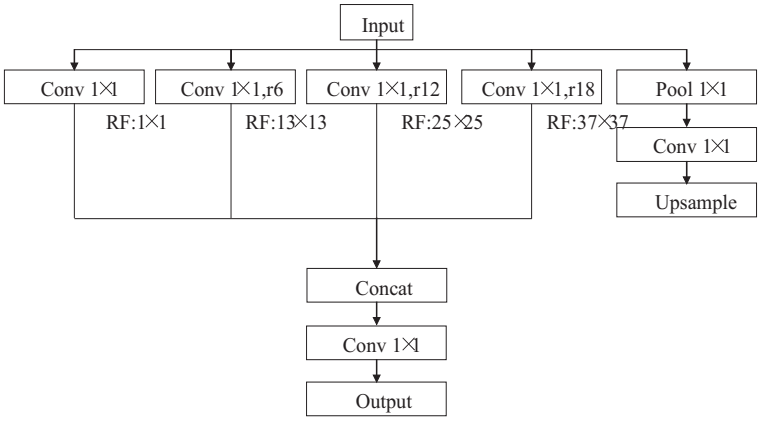

图 4　ASPP 模块

2.3　Shuffle attention 注意力机制

在深度学习中,注意力机制能够聚焦于基本特征而抑制不必要的特征,从而提高模型提取特征的能力,并且是即插即用,因此目前被广泛用于目标检测。为进一步增强模型的特征提取能力,提高裂缝检测精度,该文在 YOLOv5 的颈部添加 SA 注意力机制,与颈部的 C3 模块融合。

目前注意力机制主要分为通道注意力机制和空间注意力机制。这两种注意力机制只注意通道或空间一个方面,效率较低,如 SE[14]。此后,有学者将这两种注意力机制整合到一个模块中,如 CBAM[15],并且效果取得了显著改善,但加大了计算量,使得模型收敛困难。当然也有学者简化了通道或空间注意力,如 ECA[16]使用一维卷积简化了 SE 模块中权值的计算。而该文所采用的 SA 注意力机制以一种更轻但更有效的方式将两种注意力融合到一起,其整体结构如图 5 所示。

对于一个特征 X,SA 先沿着通道的维度将 $X \in R^{C\times H\times W}$ 分为 g 组,即 $X=[X_1,X_2,\cdots,X_g]$,每组特征 $X_k \in R^{C/g\times H\times W}$ 又被分成两个分支,一个分支根据通道之间的关系生成通道注意图,另一个分支根据空间信息生成空间注意图。其中,SA 内部的通道注意力和 SE 实现类似,而空间注意力的实现是通过 GN(GroupNorm)获取空间的维度信息,然后进行连接,最后使用通道洗牌操作,使得通道之间的信息在不同维度之间传递,通道注意力和空间注意力的最终输出分别如式(1)和式(2)所示。

$$X'_{k1}=\sigma(F_c(s))\cdot X_{k1}=\sigma(W_1 s+b_1)\cdot X_{k1} \quad (1)$$

$$X'_{k2} = \sigma(W_2 \cdot \mathrm{GN}(X_{k2}) + b_2) \cdot X_{k2} \tag{2}$$

式中，$s = F_{gp}(X_{k1}) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{k1}(i,j)$，表示的是通道统计量，$H$、$W$、$C$ 分别表示特征图的高、宽和通道数。这里 W_1、b_1、W_2、b_2 是引入的超参数，每个 SA 模块中，每个分支的通道数是 $C/2g$，因此总的参数为 $C/3g$，一般 g 为 32 或 64，这与百万参数相比，SA 是相当轻量级的。因此，相比于其它注意力机制，SA 同时考虑了通道和空间两种注意力，并且参数更少，在加强模型特征提取能力的同时也提高了检测速度。

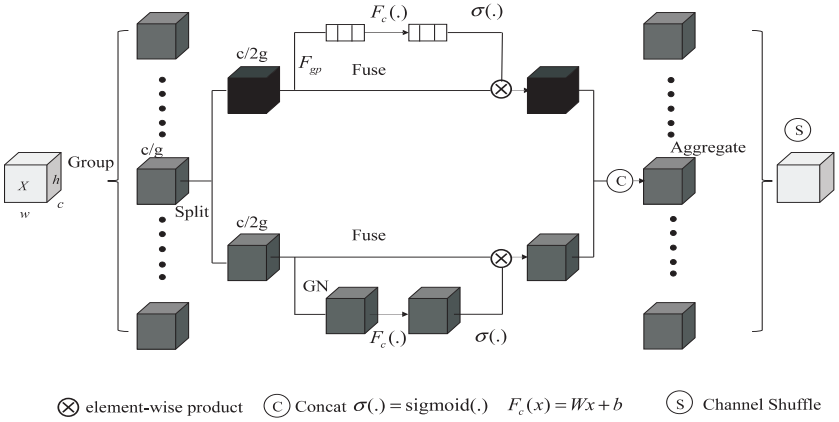

图 5 SA 原理

最终，改进的 YOLOv5 的网络结构如图 6 所示。当输入为 640×640×3 时，Entry flow 和 Middle flow 输出的特征图大小分别为 80×80×256 和 40×40×512，在颈部进行特征融合，而 Exit flow 输出特征图大小为 20×20×2 048，需经过一次卷积将通道数 2 048 调整为 1 024，再输入到 ASPP 模块中，之后进行特征融合，颈部 C3 模块与 SA 融合，提高检测精度，最后在 Head 端进行检测。

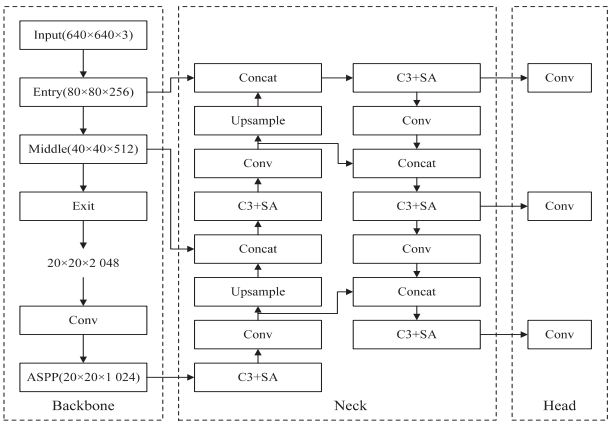

图 6 改进 YOLOv5 网络结构

3 实验及结果分析

3.1 数据集与实验环境

由于目前还没有统一的建筑裂缝检测的数据集，因此通过网络搜索和拍摄自行构建数据集，同时对采集到的图片采用 Mosaic 数据增强，最终得到约 6 000 张图片，并按照 8∶1∶1 比例划分训练集、验证集、测试集。

随后使用 labelimg 工具标注数据集，这里由于裂缝比较狭长，若直接采用一个整框标注会使得背景所占比例较高，无法充分提取特征，若采用小框标注，会加大训练难度。因此采用多个适中大小的框交替标注，覆盖整条裂缝，标注完成生成 YOLO 格式 txt 文件。

实验环境为 W10 系统、GPU 为 RTX3050，深度学习的框架为 Pytorch，设置初始学习率为 0.001，动量为 0.937，最大迭代次数为 600 轮。

3.2 模型训练与评价指标

将改进的 YOLOv5 模型在数据集上进行训练，其损失函数的曲线如图 7 所示。

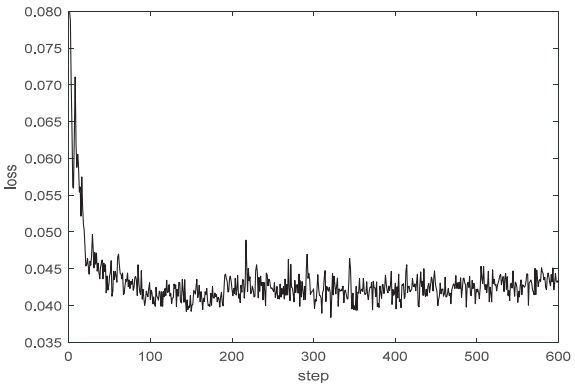

图 7 Loss 曲线

如图 7 所示，随着训练轮数增多，损失逐渐降低。前 100 轮衰减较快，第 200 轮迭代曲线有略微震荡，最终约 300 轮迭代时，曲线收敛。

深度学习模型评价指标一般采用精确率 P、召回率 R、平均精度 AP、平均精度均值 mAP、模型单位时间检测图片帧数 FPS 以及参数量等，相关评价指标的计算公式如式(3)～式(6)所示。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{AP} = \int_0^1 P \mathrm{d}R \tag{5}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{N} \mathrm{AP}_i}{N} \tag{6}$$

其中,TP 表示裂缝图片中正样本预测正确的数量,TN 表示负样本预测正确的数量,FP 为正样本预测错误的数量,FN 表示负样本预测错误的数量。

3.3　消融实验

为了验证所加模块对模型性能的影响,进行消融实验,其结果如表 2 所示。

表 2　消融实验

模型	mAP/%	参数/M	FPS/(f/s)
YOLOv5	89.6	7.2	48.1
YOLOv5+Xception	89.1	6.7	49.3
YOLOv5+ASPP	90.2	7.4	49.2
YOLOv5+SA	90.9	7.4	49.0
YOLOv5+Xception+ASPP+SA	91.2	7.2	50.8

由表 2 可以看出,首先,仅添加 Xception 模块,虽然检测速度得到了提高,但 mAP 降低了 0.5 百分点,因为对主干网络轻量化,使得参数量减少,造成了检测的精度降低。其次,仅添加 ASPP 模块,mAP 提高了 0.6 百分点,FPS 提高了 1.1 f/s,因为 ASPP 中引入了空洞卷积,扩大了感受野,加强了主干的特征提取能力。再者,仅加入 SA 注意力机制,mAP 提高了 1.3 百分点,同时由于 SA 模块的参数量较少,属于真正的轻量级模块,FPS 也得到了提高。最后,同时加入 3 种模块,提升效果最为明显,mAP 提高了 1.6 百分点,FPS 达到 50.8 f/s,且参数量也和原来一致,从而验证了所改进的模型在性能上优于原先的模型。

同时,为了验证 SA 注意力机制的有效性,在 YOLOv5 模型中分别加入目前比较常用注意力机制 SE、CBAM 以及 ECA 和 SA 进行对比,其实验结果如表 3 所示。

表 3　注意力对比实验

模型	mAP/%	参数/M	FPS/(f/s)
YOLOv5+SE	90.1	7.4	48.1
YOLOv5+CBAM	90.5	7.4	47.8
YOLOv5+ECA	90.5	7.2	48.7
YOLOv5+SA	90.9	7.2	49.0

从表 3 可以看出,由于 SE 只考虑通道之间的关系,未考虑空间信息,故 mAP 和 FPS 较低。CBAM 注意力机制结合了通道和空间注意力机制,虽然提高了检测精度,但是 CBAM 的计算量大,速度较慢。ECA 注意力机制虽然 mAP 和 CBAM 相同,但 FPS 较 SE、CBAM 有较大提高,因为 ECA 去掉了 SE 模块中的全连接层,在全局平均池化之后直接用一个 1 维卷积学习,提高了检测速度。而该文所加入的 SA 注意力机制,mAP 和 FPS 比其它 3 种注意力机制都高,因为 SA 的参数量最少,属于轻量化的模块,并且结合了通道和空间特征,既提高了检测的精度,也加快了检测速度,验证了 SA 注意力机制的有效性。

3.4　目标检测算法对比

为进一步对改进的模型进行评估,将改进算法与目前较典型的目标检测算法进行纵向对比,其结果如表 4 所示。

表 4　目标检测算法对比

模型	mAP/%	参数/M	FPS/(f/s)
Faster R-CNN	93.4	96.8	11.2
Mobile-SSD	80.6	23.7	39.4
YOLOv4-tiny[17]	85.4	24.5	48.7
文中算法	91.2	7.2	50.8

从表 4 可以看出,Faster R-CNN 作为二阶段的目标检测算法,虽然检测精度是所有算法里精度最高的,但是模型复杂,参数较多,检测速度最慢。Mobile-SSD、YOLOv4-tiny 等轻量化模型,虽然检测速度快于 Faster R-CNN,但检测精度不高。而文中算法虽然精度没有 Faster-RCNN 高,但 mAP 仍达到了 91.2%,且检测速度是最快的。因此,综合检测的精度和速度两方面来看,文中算法性能是最优的。

3.5　检测效果对比

最后,为了验证改进模型的检测效果,选取 3 组典型的裂缝图片,其检测效果对比如图 8 所示。

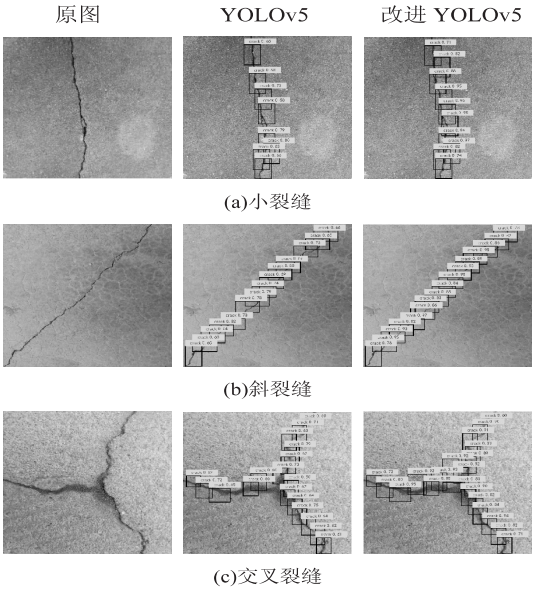

图 8　检测效果对比

从图 8 可以看到,检测以上几种裂缝,改进之后的算法检测裂缝的置信度都大于原算法,除此之外,原算法的检测框出现了不连续的情况,没有完整反映出裂缝的走势,而改进之后的算法的检测框都是连续的,覆盖整个裂缝,验证了改进算法的有效性。

4 结束语

提出了一种改进的 YOLOv5 建筑裂缝检测算法,主要创新点在于使用 Xception 网络对主干网络轻量化,并采用 ASPP 模块替换 SPP 模块,最后在颈部添加 SA 注意力机制。通过实验证明,改进算法满足了裂缝检测精度和实时性的要求,但由于检测框较多,导致模型的召回率较低,有待进一步优化。

参考文献:

[1] 丁 威,俞 珂,舒江鹏. 基于深度学习和无人机的混凝土结构裂缝检测方法[J]. 土木工程学报,2021,54(S1):1-12.

[2] 郑正南. 基于深度学习和图像处理技术的建筑裂缝检测与测量方法研究[D]. 南京:南京理工大学,2021.

[3] 夏 坚,周利君,张 伟. 基于迁移学习与 VGG16 深度神经网络的建筑物裂缝检测方法[J]. 福建建设科技,2022(1):19-22.

[4] LIU W,ANGUELOV D,ERHAN D,et al. SSD:single shot multibox detector[C]//European conference on computer vision. [s. l.]:Springer,2016:21-37.

[5] 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. Las Vegas:IEEE,2016:779-788.

[6] 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(10):23-34.

[7] 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[J]. 华南理工大学学报:自然科学版,2020,48(2):84-93.

[8] 李鹏程,孙立双,谢志伟,等. 基于改进 MobileNet-SSD 的路面裂缝图像检测算法[J]. 激光杂志,2022,43(7):123-127.

[9] 杨富强,余 波,赵嘉彬,等. 基于改进 YOLOv3 的桥梁底部裂缝目标检测方法[J]. 中国科技论文,2022,17(3):252-259.

[10] 郝巨鸣,杨景玉,韩淑梅,等. 引入 Ghost 模块和 ECA 的 YOLOv4 公路路面裂缝检测方法[J/OL]. 计算机应用:1-7[2022-11-21]. http://kns.cnki.net/kcms/detail/51.1307.TP.20220904.1507.002.html.

[11] CHOLLET F. Xception:deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Hawaii:IEEE,2017:1251-1258.

[12] CHEN L C,PAPANDREOU G,KOKKINOS I,et al. Deeplab semantic image segmentation with deep convolutional nets,atrous convolution,and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,40(4):834-848.

[13] ZHANG Q L,YANG Y B. SA-Net:shuffle attention for deep convolutional neural networks[C]//ICASSP 2021-2021 IEEE international conference on acoustics,speech and signal processing. Toronto:IEEE,2021:2235-2239.

[14] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:7132-7141.

[15] WOO S,PARK J,LEE J Y,et al. CBAM:convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Munich:Springer,2018:3-19.

[16] WANG Q,WU B,ZHU P,et al. ECA-Net:efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF conference on computer vision and pattern recognition. Seattle:IEEE,2020:13024-13036.

[17] WANG C Y,BOCHKOVSKIY A,LIAO H Y M. Scaled-yolov4:scaling cross stage partial network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle:IEEE,2021:13029-13038.