

DOI:10.13409/j.cnki.jdpme.20240304004

基于Mask R-CNN的多类建筑物损伤识别方法*

杨敬松^{1,2,3}, 王煜鑫¹, 李智涛⁴, 卢泽葳¹, 彭福民¹

(1. 防灾科技学院资源与环境学院, 河北 三河 065201; 2. 河北省地震灾害仪器与监测技术重点实验室, 河北 三河 065201; 3. 廊坊市精密主动震源重点实验室, 河北 三河 065201; 4. 应急管理部国家自然灾害防治研究院, 北京 100085)

摘要: 地震发生后快速对建筑物损伤进行识别, 可以提高灾害损失评估的效率, 并为救援提供有效地决策支持。针对因背景干扰带来的重要特征表达能力弱的问题, 提出一种基于深度学习框架Mask R-CNN的多建筑物损伤识别方法。首先, 对样本图像进行预处理, 克服复杂环境背景因素干扰, 并进行多途径扩增, 得到用于深度学习的扩增样本数据集。其次, 优化特征提取网络, 采用嵌入注意力机制模块SE的MobileNetv3网络作为主干网络, 增加模型对建筑物损伤空间及语义信息的提取, 有效避免背景对模型性能的影响, 改进损失函数, 避免遗漏类别和类别错分现象, 同时引入迁移学习, 降低训练成本; 最后, 采用定性分析和定量评估相结合的手段, 多维度评估模型泛化能力和鲁棒性。改进后的Mask R-CNN模型的平均精度达到了84.34%, 相对于原始的Mask R-CNN模型, 精度提高了9.12%。结果表明, 改进后的模型在识别含有多种损伤特征和噪声背景的建筑物损伤图像方面表现良好, 可以为地震后建筑物的损伤评估提供有效地技术支持。

关键词: 人工智能; 建筑物损伤识别; Mask R-CNN; 实例分割

中图分类号: TU375 **文献标识码:** A **文章编号:** 1672-2132(2025)03-0562-09

Building Damage Identification Method based on Mask R-CNN

YANG Jingsong^{1,2,3}, WANG Yuxin¹, LI Zhitao⁴, LU Zewei¹, PENG Fumin¹

(1. College of Resource and Environment, Institute of Disaster Prevention, Sanhe 065201, China;
2. Hebei Key Laboratory of Seismic Hazard Instrumentation and Monitoring Technology, Sanhe 065201, China;
3. Langfang Key Laboratory of Precision Active Source, Sanhe 065201, China;
4. National Institute of Natural Hazards, Ministry of Emergency Management, Beijing 100085, China)

Abstract: After an earthquake, quickly identifying building damage can improve the efficiency of disaster loss assessment and provide effective decision-making support for rescue efforts. Background interference can weaken the expression of important features. To address this issue, a method based on the deep learning framework Mask R-CNN was proposed for identifying multiple types of building damage. First, the sample images underwent preprocessing to overcome interference from complex environmental backgrounds. Multiple augmentation techniques were applied to generate a dataset suitable for deep learning. Next, the feature extraction network was optimized using the MobileNetv3 net-

* 收稿日期:2024-03-04;修回日期:2024-08-05

基金项目: 国家自然科学基金项目(41874219)、2024年廊坊市科学技术研究与发展计划自筹经费项目(2024011022)资助

作者简介: 杨敬松(1975—),女,教授,博士。主要从事地震预警技术和优化算法研究。E-mail: yangjingsong@cidp.edu.cn

通信作者: 李智涛(1978—),女,副研究员,博士。主要从事地震烈度速报等相关领域的基础理论与应用研究。

E-mail: zhitaoli1027@163.com

work with an embedded SE (Squeeze-and-Excitation) attention mechanism module as the backbone. This design enhanced the model's ability to extract spatial and semantic information related to building damage and effectively reduced the negative impact of background interference. Furthermore, the loss function was improved to avoid category omissions and misclassifications, while transfer learning was introduced to reduce training costs. Finally, a combination of qualitative analysis and quantitative evaluation was employed to assess the model's generalization ability and robustness. The results demonstrated that the improved Mask R-CNN model achieved an average precision of 84.34%, which was a 9.12% improvement over the original Mask R-CNN model. The improved model performs well in identifying building damage images with various damage features and noisy backgrounds, providing effective technical support for post-earthquake building damage assessment.

Keywords: artificial intelligence; building damage detection; Mask R-CNN; instance segmentation

0 引言

地震发生后,由于强烈的地面震动及伴生的地面裂缝和变形,建筑物可能出现裂纹、倒塌等损伤。快速准确地识别建筑物损伤,可以评估建筑物的结构完整性和安全性,为灾害损失评估提供准确的数据来源,并为灾后建筑物维修提供决策支持。

传统的建筑物损伤识别主要依赖于灾后的社会调查,对调查和评估人员的专业能力要求高,需要消耗大量的人力和时间成本,且灾后环境复杂,危险系数大,因此对建筑损伤图像进行自动识别和分类,可以提高建筑物损伤评估的效率和准确性,减少人力和物力成本。近年来,深度学习在图像识别领域取得了显著的发展^[1],并被广泛应用于各种应用场景^[2-4],包括从卫星遥感和无人机航拍影像中提取建筑物的震害信息,如陈梦^[5]采用全卷积神经网络和基于编码解码的神经网络两种方法进行建筑物震害提取的研究分析,表明深度学习技术在提升建筑物倒塌遥感识别自动化程度方面有极大的潜力。如今,天网监控系统逐渐完善,视频监控网络密度大、分布相对均匀,在城镇和乡村都有广泛分布,图像获取变得更加容易,人们把目光转向了通过从摄像头获取的图像来进行损伤检查工作^[6-9]。Y. J. Cha等^[10]使用深度 CNN 检测混凝土裂缝;A. Vetrivel等^[11]将深度 CNN 和三维点云技术结合起来检测建筑物的立面和屋顶损伤;C. M. Yeum等^[12]将牛津大学计算机视觉组(visual geometry group, VGG)CNN架构的一种浅层模型 VGG-F 应用于从大比例尺影像提取视觉信息,进行建筑物倒塌分类,但是其应用的基础卷积神经网络,存在重要特征表达能力弱、准确度不高的问题。视频监控

网络获取的图像可以提供更多样化和实时的数据,为灾害应对和决策提供有力支持,但需要解决因背景干扰等原因带来的识别准确度低的问题。因此,本研究针对背景干扰导致重要特征表达能力不足的问题,以灾后建筑物裂纹、剥落、钢筋暴露、部分倒塌和完全倒塌五类损伤为研究对象,基于较为成熟的 Mask R-CNN 深度学习网络模型,提出了一种自动识别建筑物多处损伤的 CNN 模型。

1 建筑物损伤识别实验数据集构建

1.1 数据来源

当前公开的建筑物损伤数据集相对有限,本文通过多种方式获取图像数据,构建建筑物损伤数据集:(1)从已公开的数据集中获取图像:使用了 PHI^[13]的开放数据作为主要数据源,该数据集包含 8 个子数据集,通过筛选和处理,将其中包含建筑物损伤的图像提取出来,用于构建建筑物损伤数据集;(2)利用 Python 爬虫脚本从网络上收集建筑物损伤图像;(3)从各大视频网站获取视屏数据,截取视频图像帧作为图像数据。实验筛选使用了其中与建筑物裂纹、剥落、钢筋暴露、部分倒塌和完全倒塌等模式有关的部分数据集;其中裂纹 231 张、剥落 476 张、钢筋暴露 370 张、部分倒塌 203 张和完全倒塌 202 张,数据样例如图 1 所示。

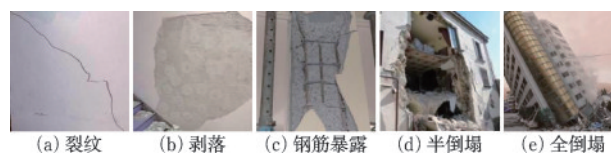


图 1 建筑物损伤数据集样例

Fig.1 Examples from building damage dataset

1.2 数据集构建

1.2.1 数据清洗与扩充

在删除数据集中标签错误的图像后,考虑到样本数量有限,为了提高训练样本的多样性,更好地提取建筑物损伤特征并提高模型泛化能力,避免模型在训练过程中发生过拟合现象,采用数据增强技术对数据集进行样本扩增:运用几何变换中的翻转操作,模拟不同拍摄方向和角度的变化;对图像进行亮度变换,模拟不同光照条件下的建筑物损伤;对图像进行锐度变换,模拟相机聚焦效果的变化;对图像进行模糊处理,模拟相机运动或焦点失准造成的模糊效果。

1.2.2 多边形标注

Mask R-CNN 模型,在训练前需要对建筑物损伤图像进行标注构建训练集。采用 Labelme 中的工具,沿着损伤轮廓取点,对裂纹、剥落、钢筋暴露、部分倒塌和完全倒塌等损伤进行多边形标注,避免复杂背景的影响,标注后生成的文件如图 2 所示。

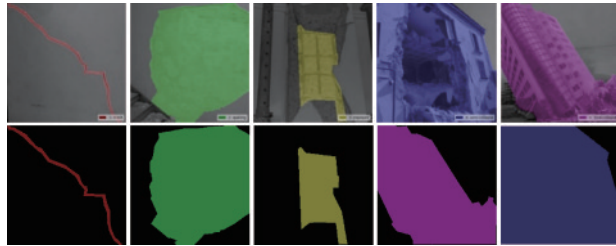


图 2 Json 标签图和 mask 图

Fig.2 JSON label and mask diagrams

1.2.3 数据集划分

考虑到建筑物损伤形态特征多样、噪声干扰多、语义信息丰富的特点,为增强模型的鲁棒性,按照不同比例分别扩增裂纹、剥落、钢筋暴露、部分倒塌和完全倒塌图像。将数据集按 7:2:1 划分为训练集、验证集和测试集,以支持模型的训练、调整和评估。数据集划分结果见表 1。

表 1 建筑物损伤数据集

Table 1 Building damage dataset

损伤种类	标注标签	原始数目	扩增后数目		
			训练集	测试集	验证集
裂纹	crack	231	1 026	40	20
剥落	spalling	476	1 304	100	50
钢筋暴露	exposure	370	1 400	60	30
半倒塌	semi-collapse	203	858	40	20
全倒塌	Total-collapse	202	852	40	20

2 基于改进 Mask R-CNN 的建筑物损伤识别模型

2.1 Mask R-CNN 基本原理

Mask R-CNN 如今被广泛应用于目标检测和实例分割。在 Faster R-CNN 的基础上,通过引入实例分割分支,以实现更精细的目标分割。并替换了 ROI Pooling 层,将 ROI 区域内的特征图与固定大小的特征图对齐,以便进行后续的目标分类和掩码预测,如图 3 所示。通过卷积神经网络进行特征提取和候选框生成,Mask R-CNN 能够准确地定位和分类图像中的目标,并为每个目标生成高质量的掩码。这使得它在许多计算机视觉任务中表现出色。

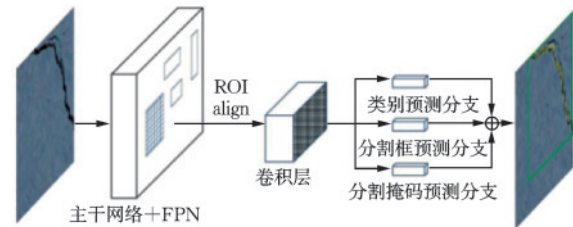


图 3 Mask R-CNN 网络结构

Fig.3 Mask R-CNN network architecture

2.2 Mask R-CNN 模型改进

2.2.1 MobileNetv3 模块

本文借鉴 MobileNetv3 中的 Inverted-Bottleneck 结构和特征金字塔网络(FPN)^[14]的思路,对网络结构进行改进,旨在提高模型的效率、速度和对不同建筑物损伤的识别能力。这些改进有助于使模型更适用于实际场景中的建筑物损伤识别任务。改进 Mask R-CNN 的建筑物损伤识别模型架构如图 4 所示。

MobileNetv3 网络分为 MobileNetV3-Large 和 MobileNetV3-small 两种,本文使用其中的轻量级网络 MobileNetV3-Large 替换 ResNet50 作为 Mask R-CNN 的主干网络,可以在保持较高准确性的同时显著减少模型的参数量和计算复杂度。MobileNetv3-Large 引入深度可分离卷积,减少参数量,加快计算速度;引入倒残差结构,缓解梯度消失问题,提高模型训练效果;引入基于 squeeze and excitation 结构的轻量级注意力模型,提高重要特征的代表能力。MobileNetv3-Large 网络结构见表 2。

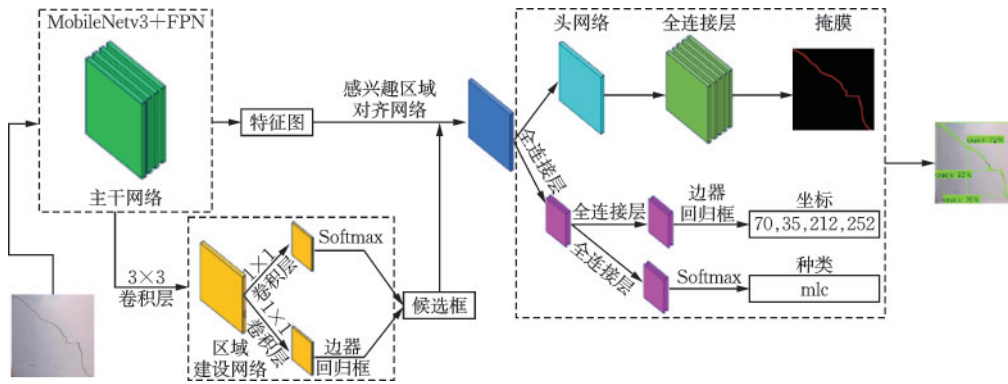


图4 改进后 Mask R-CNN 模型架构

Fig.4 Improved Mask R-CNN model architecture

表2 MobileNetV3-Large 网络结构

Table 2 Structure of MobileNetV3-Large

运算单元	输出通道数	输入通道数	步长	激活函数	是否使用SE
Conv2d	16	3	2	Hard-swish	×
Inverted-bneck-shortcut	16	16	1	ReLU	×
Inverted-bneck	24	16	2	ReLU	×
Inverted-bneck-shortcut	24	24	1	ReLU	×
Inverted-bneck	40	24	2	ReLU	✓
Inverted-bneck-shortcut	40	40	1	ReLU	✓
Inverted-bneck-shortcut	40	40	1	ReLU	✓
Inverted-bneck	80	40	2	Hard-swish	×
Inverted-bneck-shortcut	80	80	1	Hard-swish	×
Inverted-bneck-shortcut	80	80	1	Hard-swish	×
Inverted-bneck-shortcut	80	80	1	Hard-swish	×
Inverted-bneck	112	80	1	Hard-swish	✓
Inverted-bneck-shortcut	112	112	1	Hard-swish	✓
Inverted-bneck	160	112	2	Hard-swish	✓
Inverted-bneck-shortcut	160	160	1	Hard-swish	✓
Inverted-bneck-shortcut	160	160	1	Hard-swish	✓
Conv2d	960	160	1	Hard-swish	×
Pool	—	960	1	—	×
Conv2d	1 280	960	1	Hard-swish	×
Conv2d	—	1 280	1	—	×

表2中: Conv2d, 普通卷积层; Pool, 池化层; Inverted-bneck, 由两个 1×1 卷积层和一个 3×3 深度卷积层组成的复合结构, 增加非线性表示能力和参数共享; Inverted-bneck-shortcut, 在 Inverted-bneck 结构中加入了 shortcut 连接, 将卷积层的输出与输入进行融合, 消除网络退化的影响, 提高模型的表达能力和准确性。

(1) 深度可分离卷积: MobileNetV3-Large 利用了 MobileNetV1 中的深度可分离卷积结构, 该结构

能够减少计算量并保持一定的准确性。该结构由深度卷积和逐点卷积两个步骤组成。首先使用深度卷积核每个输入通道进行卷积操作, 然后使用 1×1 逐点卷积核对通道间的特征进行卷积操作。这种方式将空间信息和通道信息的处理分开, 从而减少参数数量。相比传统的卷积操作, 深度可分离卷积可以极大地减少参数数量, 同时在一定程度上保持模型的准确性和计算效率, 结构如图5所示。

(2) 倒残差结构: MobileNetV3-Large 引入了

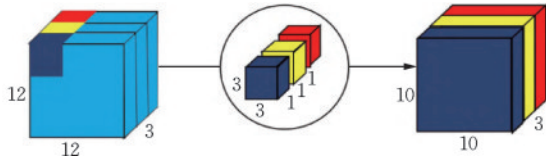


图5 深度可分离卷积

Fig.5 Depthwise separable convolution

MobileNetV2的倒残差结构并进行了改变。倒残差结构包括一个扩张卷积层、一个深度可分离卷积层和一个逐点卷积层。倒残差结构的引入在MobileNetV3-Large中能够提升模型的非线性表达能力,并通过深度可分离卷积减少参数数量。同时,跳跃连接的使用有助于保留和传递更丰富的特征信息,从而提高模型的准确性和计算效率,结构如图6所示。

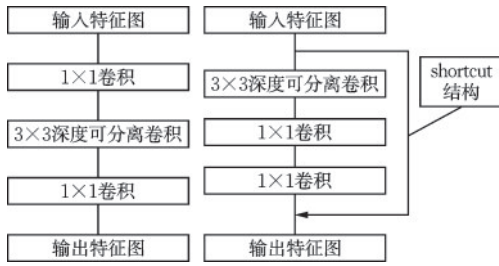


图6 倒残差结构

Fig.6 Inverted residual structure

(3) 轻量级注意力结构: MobileNetV3-Large引入了Squeeze-and-Excitation^[15]结构作为一种轻量级注意力模型。这种模型旨在通过学习通道间的权重关系,增强模型对重要特征的关注程度,从而提高模型的表达能力和准确性。注意力机制如图7所示。

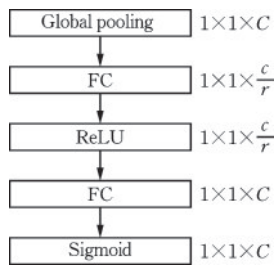


图7 注意力机制

Fig.7 Attention mechanism

(4) 通过使用hard_swish激活函数,模型可以在保持计算速度较快的同时,具备一定的非线性表达能力,而hard_swish激活函数是通过hard_sigmoid实现的。

$$\text{hard_sigmoid}(\beta x) = \frac{\text{ReLU6}(x + 3)}{6} \quad (1)$$

$$\text{hard_swish} = \text{hard_sigmoid}(\beta x) \quad (2)$$

2.2.2 损失函数

Mask R-CNN的损失函数 L_{loss} 由三部分损失组成,其表达式为:

$$L_{\text{loss}} = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}} \quad (3)$$

式中, L_{cls} 为分类损失值; L_{box} 为边界框回归损失值; L_{mask} 为分割损失值。

其中, L_{cls} 采用平均二值交叉熵损失函数,在训练建筑物损伤数据集过程中,存在遗漏类别和类别错分现象,针对上述两个问题本文提出了正负样本权重注意力机制和类别权重注意力机制的交叉熵损失函数来代替原论文提到的交叉熵损失函数。具体做法是:

(1) 针对正负样本不平衡,根据正负样本数量占总样本的比例分配权重,即样本数量越多分配的权重越小,样本的数量越少分配的权重越大。

$$\text{Weight}_{\text{pn}} = [C_p / (C_p + C_n), C_n / (C_p + C_n)] \quad (4)$$

其中, $\text{Weight}_{\text{pn}}$ 表示正负样本注意力权重; C_p 表示正样本的数量; C_n 表示负样本的数量; $C_p / (C_p + C_n)$ 表示负样本的权重; $C_n / (C_p + C_n)$ 表示正样本的权重。

(2) 针对不同类别样本难易程度不平衡,采用不同类别提取精度比例和给它们分配的权重成反比,即类别提取精度越高分配的权重越小,类别提取的精度越低分配的权重越大。

$$\text{Weight}_{\text{cls}} = \text{Sigmoid} [(1 - acc_1) / C_1, (1 - acc_2) / C_2, \dots, (1 - acc_n) / C_n] \quad (5)$$

其中, $\text{Weight}_{\text{cls}}$ 表示类别权重注意力机制; $C_1 \sim C_n$ 表示每个类别的数量; $acc_1 \sim acc_n$ 表示每个类别的精度; Sigmoid 表示将比例映射到(0, 1)范围内。

(3) 由式(4)、(5)可得注意力机制的交叉熵函数为:

$$L_{\text{cls}} = L_{\text{cls}} * \text{Weight}_{\text{pn}} * \text{Weight}_{\text{cls}} \quad (6)$$

(4) 由式(4)、(5)、(6)得mask_rcnn的损失函数为:

$$L_{\text{loss}} = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}} \quad (7)$$

2.2.3 迁移学习

迁移学习是一种机器学习技术,通过将从一个任务或领域学到的知识应用到另一个相关任务或领域上,从而加快学习速度,提升泛化能力,提高数据效率。为了更好地训练模型,本文采用迁移学习对模型进行训练,将通过实验对比特征提取网络采用迁移学习和非迁移学习策略训练的模型性能。

3 实验与结果分析

3.1 试验环境及超参数设置

本次试验采用的硬件环境为：NVIDIA GeForce GTX 1050Ti GPU、Intel i5-7300 CPU 处理器、内存 24 GB。软件环境为：Windows 10 操作系统、Python3.7、CUDA11.3、CUDNN8.9.1、pytorch1.10.0。设置单 GPU 同时处理图像数为 2，每幅显微图像实例分割的类别为裂纹、剥落、钢筋暴露、部分倒塌和完全倒塌 5 类（不包含背景）。通过比较模型在相同环境、不同参数下的性能，确定的超参数设置见表 3。

表 3 模型超参数设置

参数	数值
主干网络	MobileNetv3
初始学习率	0.004
最小图像维度	224
最大图像维度	1 024
锚框	(16, 35, 64, 128, 256)
动量因子	0.9
置信度	0.5
1 迭代轮数	100

3.2 试验评价指标

不同于一般的研究对比中，仅以各类精度指标对模型性能进行评估。考虑到工程应用的实际落地需求，本文选择从以下几个方面进行多方位的评估。

(1) 训练时长：对基于深度学习的方法来说，训练时间的长短决定了模型更新的速度。这意味着训练耗时越短，模型对于新类的学习效率更高。因此，文中通过记录识别模型的训练时间来评估其在面对工业制造领域中复杂多变的情形时所具有的适应性。

(2) 模型大小：在实际的工业应用中，相应技术的配套成本应尽可能的低，以迎合硬件设施条件。而模型大小作为从实验阶段迁移至工业环境的直接对象，其影响着应用端的兼容性。所以，文中采用训练后模型权重参数的内存占用大小作为评估该项指标的重要依据。

(3) 精度：作为体现实验结果优劣中必不可少的评估指标之一，精度可展示出模型应对建筑物损伤数据集的直接效能。对此，文中采用平均准确率 AP 以及平均准确率均值 mAP 作为评价指标相对全面地评判模型是否达到预期的性能。AP 表示 P-R 曲线与坐标轴围成的面积，通常以召回率为横轴，精确率为纵轴，AP 值同样也是建立在精确率和召回率两个指标下的评价指标，AP 的高低代表着该模型的性能优劣。具体的计算公式如下：

$$AP = \int_0^1 P(r) dr \quad (8)$$

而 mAP 则为所有类别 AP 的平均值。

(4) 推理速度：为了契合实际的生产节奏，推理速度是评估模型性能又一个不可忽视的重要指标。因此，文中使用 FPS (Frames Per Second) 来评价模型在推理过程中的运行速度即每秒可处理的损伤图像张数。

(5) 损失函数：用于反映模型的性能，衡量预测值与真实值之间的差距，也显示模型收敛情况，较小的损失函数值通常表示更好的检测效果。

3.3 定性实验

根据图 8 所示的结果，可以明显看出本文提出的改进 Mask R-CNN 模型在识别建筑物损伤方面表现出色。模型能够准确地识别出裂纹、剥落、钢筋暴露、部分倒塌和完全倒塌等不同类型的损伤，并且能够完整地分割出损伤区域的边缘信息。这些结果充分展示了本文提出的改进算法具有鲁棒性和泛化能力，且能够有效地应用于实际场景中，对于损伤类型的识别和分割任务具有较高的准确性。

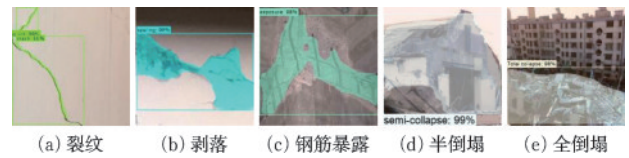


图 8 定性试验识别结果

Fig.8 Qualitative identification results from experiments

3.4 定量分析实验

3.4.1 实验结果

在完成建筑物损伤数据集的构建、Mask R-CNN 网络的改进后，本节将对 5 类损伤的模型识别和分割效果进行实验分析。将数据增强后的 5 860

张建筑物损伤图像样本作为实验的数据集,对本文设计的网络模型进行训练并测试识别和分割性能,各类损伤的识别效果见表4。

表4 模型损伤识别效果

损伤类别	裂纹	剥落	钢筋暴露	半倒塌	全倒塌
AP/%	82.47	88.71	92.09	90.14	99.76

从表4可以看出,每一类的识别精度均在80%以上。在5类建筑物损伤中,全倒塌的识别效果最好,识别精度达到99.76%,而裂纹、剥落的识别效果相对较差,其中裂纹的识别精度为5类损伤中最低的,82.47%。综合来看,改进后的模型对整个建筑物损伤数据集都有相对较高的识别精确度和较好的分类性能。

各类损伤的分割效果见表5。从表5可以看出,除裂纹之外,每一类的分割精度均在85%以上。在5类建筑物损伤中,全倒塌的分割效果最好,分割精度达到99.76%,而裂纹的分割效果较差。原因有二:一是图片分辨率的限制;二是裂纹较窄,在剥落附近的较小裂纹不容易被检测分割。综合来看,改进后的模型对整个建筑物损伤数据集都有相对较高的分割精确度。

表5 模型损伤分割效果

损伤类别	裂纹	剥落	钢筋暴露	半倒塌	全倒塌
AP/%	52.38	90.38	87.99	91.18	99.76

为验证数据增强对于模型识别效果的影响,对数据增强前后的数据集进行识别效果对比,结果如图9所示。可知数据集扩充后,之前部分损伤如裂纹,全倒塌,由于本身的数据量较少,增强后的识别

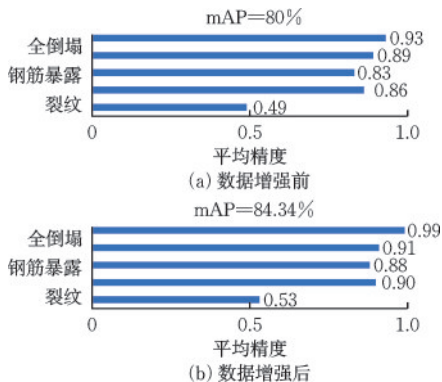


图9 模型在数据增强前后分割效果对比

Fig.9 Comparison of model segmentation performance before and after data augmentation

效果都有相应的提升。而通过进一步增强数据,发现识别精度不再有明显提升。

对于本研究中裂纹分割精度只有52.38%的问题,分析可能的其他原因是由于本研究中使用的Mask R-CNN网络模型,开发模型时主要基于的COCO数据集,包括了20个类别,其中对象多为较大且规则的物体,相比之下,建筑物损伤,特别是裂纹损伤细小、细长、分支多、边界模糊且不规则;剥落中也常带有钢筋暴露或者裂纹,这对模型的分割能力提出了更高的要求。针对这些问题,后续可以考虑根据损伤特征来调整RPN的大小和深度、优化掩模分支并对原始图像进行数据增强与预处理^[16]。

3.4.2 对比分析

试验使用表1数据集分配标准进行训练、验证及测试。为了评估本文提出的改进Mask R-CNN模型在建筑物损伤识别任务中的性能,选择了原Mask R-CNN模型、SSD模型、YOLOv3作为对比模型,进行了模型对比实验。对比结果见表6。

表6 与其他流行方法的整体性能对比

方法	主干网络	训练时长/h	模型大小/MB	mAP/%	Fps/(帧·S ⁻¹)
SSD	VGG16	4.76	100	71.91	26
YOLOv3	DarkNet-53	4.28	236	62.82	23
Mask RCNN	ResNet50	9.26	335	75.22	15.78
本文算法	MobileNetV3	7.67	169	84.34	18.55

表6展示了本文方法与现有主流方法(包括两阶段与单阶段)的整体效果对比。可以看出,在各个评估方面,本文模型都具备较好的性能,具体分析如下。

(1) 训练时长:由于两阶段方法将候选框的生成与筛选阶段与后续的分类与回归阶段分开处理,导致它的训练过程相对较长。但是,经过一系列的轻量化操作,本文所提出的方法在训练时长上相比同为两阶段方法的Mask RCNN减少了1.59 h。这一优势意味着本文提出的模型具有更高的学习效率,能够更快地适应新类别缺陷的识别与定位任务。

(2) 模型大小:尽管精度的提升需要依赖网络容量的增加,但本文方法在保持高精度的同时,有效降低了网络复杂度。相较于最小模型SSD,本文

方法虽然参数量多了 69 M,但与精度次优的 Mask RCNN 相比,内存占用大小仅为近一半。这表明本文的模型在保持高精度的同时,也有效降低了网络复杂度,从而降低了配置成本,更容易部署在工业应用端。

(3) 定位精度:在定位精度方面,本文方法展现出显著优势。从表 6 中可以看出,本文方法在这一重要检测指标上高于其他方法。具体而言,相较于训练时间较短的 YOLOv3 模型,本文方法高出约 21.52% 的 mAP;与参数量最小的模型 SSD 相比,高出约 12.43%;即使与精度次优的 Mask RCNN 相比,仍具有 9.12% 的精度优势。

(4) 推理速度:虽然功能性模块的引入在一定程度上影响了模型的推理速度,但本文方法在保证精度优势的同时,仍具有一定的速度优势。具体的,虽然推理速度慢于单阶段算法,但与同为两阶段方法的 Mask RCNN 相比,本文方法具有近 3% 的优势,以及 18.55 FPS 的推理速度,这足以满足实际的任务需求。

由图 10 可知,改进后的模型在训练过程中表现出了更好的收敛性能和更高的准确率。模型逐渐收敛,振荡越来越微弱,改进前模型最终收敛在 0.158 附近,改进后模型最终收敛在 0.146 附近,表明模型模型训练充分,收敛良好,能够更准确地预测建筑物损伤图像的分割结果,进一步证明了模型对于建筑物损伤图像分割任务的有效性和可行性。

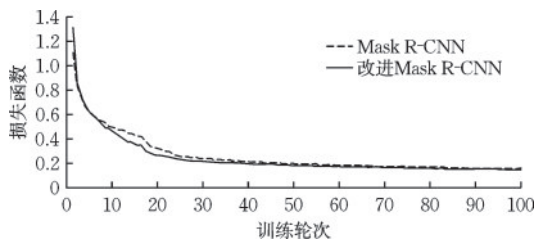


图 10 改进前后损失函数对比

Fig.10 Comparison of loss function performance before and after improvement

消融实验有助于深入了解模型的内部机制,验证各个模块结构的有效性。为评估模型的不同模块对整体性能的影响,将改进模型分别与使用数据增强后训练数据的算法 A、采用 MobileNetv3 作为 Backbone 并改进损失函数的算法 B、Mask R-CNN 原模型三者进行对比试验,结果见表 7。

表 7 各个模块对网络性能的影响

Table 7 Impact of each module on network performance

Mask R-CNN	改进模块			平均精度/%
	数据增强	MobileNetv3	损失函数	
✓				75.22
✓	✓			79.53
✓		✓		78.10
✓			✓	79.86
✓	✓	✓	✓	84.34

表 7 中,用不同的改进方法对 Mask R-CNN 模型进行了消融实验,并评估了这些改进对建筑物损伤识别模型性能的影响。得出以下几点总结性认识:

(1) 仅使用数据增强技术,模型的平均精度均值提升了 4.31%。这表明通过增加训练数据的多样性和数量,可以改善模型的泛化能力,从而提高建筑物损伤识别模型的性能。

(2) 使用 MobileNetv3 网络替代 ResNet50 网络,模型的平均精度均值提升了 2.88%。这表明 MobileNetv3 网络在建筑物损伤识别任务中具有更好的性能,可能由于其轻量级的结构和高效的特征提取能力。

(3) 通过改进损失函数,模型的平均精度均值提升了 4.64%。这表明优化损失函数可以提高模型对于建筑物损伤的识别能力,使其更好地逼近真实标签。

(4) 同时融合数据增强、MobileNetv3 网络和改进损失函数,模型的平均精度均值提升了 9.12%,达到 84.34%。这进一步验证了各种改进方法的互补效果,并证明了综合应用这些改进能够显著提升建筑物损伤识别模型的性能,对于提高建筑物损伤识别模型的准确性和鲁棒性具有重要意义。

4 结 语

基于改进 Mask R-CNN 的基本框架,针对建筑物损伤数据集训练过程中存在的遗漏类别和类别错分问题,以及重要特征表达能力弱、识别准确率低的问题,提出了两个改进措施。对数据集进行数据增强,采用 MobileNetv3 网络作为改进模型的骨干特征提取网络,提高模型的推理速度和对建筑物损伤的提取能力;同时,针对建筑物损伤数据集训练过程中存在遗漏类别和类别错分现象,改进了损失函数,避免建筑物损伤掩膜边缘缺失及定位不准

确问题,实现对建筑物损伤精准有效识别。

在自定义数据集上,本文提出的建筑物损伤识别模型的平均识别精度为84.34%,相较于改进前模型,平均精度均值提升9.12%;消融实验中表明该模型设计合理,能充分发挥各模块的优势,可为建筑物损伤精准识别提供理论和技术支持。

参考文献:

- [1] Goodfellow I, Bengio Y, Courville A. Deep learning [M]. Cambridge: MIT Press, 2016.
- [2] Koch C, Georgieva K, Kasireddy V, et al. A review of computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure[J]. Advanced Engineering Informatic. 2015, 29(2): 196-210.
- [3] Wu W, Qurishee M A, Owino J, et al. Coupling deep learning and UAV for infrastructure condition assessment automation [C]//Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2). Kansas City, MO, USA: [s.n.], 2018.
- [4] Griffiths D, Boehm J. Rapid object detection systems, utilising deep learning and unmanned aerial systems (UAS) for civil engineering applications[J]. The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, (ISPRS), 2018, 42: 391-398.
- [5] 陈梦. 基于深度学习的建筑物震害遥感识别研究[D]. 北京:中国地震局地震预测研究所,2019.
Chen M. Research on identification of building damage information from remote sensing image based on deep learning[D]. Beijing: Institute of Earthquake Forecasting, China Earthquake Administration, 2019. (in Chinese)
- [6] Lee Y I, Kim B, Cho S. Image-based spalling detection of concrete structures using deep learning[J]. Journal of the Korea Concrete Institute, 2018, 30: 91-99.
- [7] 陈鹏,汪本康,高飒,等. 利用 ResNet 进行建筑物倒塌评估[J]. 武汉大学学报(信息科学版),2020, 45(8): 1179-1184.
Chen P, Wang B K, Gao S, et al. Building collapse assessment with ResNet[J]. Geomatics and Information Science of Wuhan University, 2020, 45(8): 1179-1184. (in Chinese)
- [8] Hori M, Sutoh A, Saitoh Y. Strong motion measurement using security video camera[J]. Journal of Structural Mechanics and Earthquake Engineering, Japan Society of Civil Engineers, 2000, 647(51):57-66.
- [9] Arimitsu Yokota, Takayuki Hamamoto, Hisashi Koga. Estimation of earthquake ground motion by image analysis of sliding objects taken with a fixed camera [C]// 21st International Conference on Pattern Recognition (ICPR 2012). Tsukuba, Japan: [s.n.], 2012.
- [10] Cha Y J, Choi W. Deep learning-based crack damage detection using convolutional neural networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.
- [11] Vetrivel A, Gerke M, Kerle N, et al. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-Kernel-learning [J]. Photogrammetry and Remote Sensing, 2018, 140: 45-59.
- [12] Yeum C M, Dyke S J, Ramire Z, et al. Big visual data analysis for damage evaluation in civil engineering[C]// The International Conference on Smart Infrastructure and Construction. Cambridge, UK: [s.n.], 2016.
- [13] Gao Y Q, Mosalam K M. Deep transfer learning for image-based structural damage recognition[J]. Computer-Aided Civil and Infrastructure Engineering, 2018, 33(9): 748-768.
- [14] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [15] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [16] Cai Z, Vasconcelos C R. High quality object detection and instance segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1483-1498.

(本文编辑:赵霞)