

基于GS-MobileNet的轻量级道路异常检测算法

余见涛, 龙传煜, 罗帆, 郑仁胜, 田雨雨

贵州乌江水电开发有限责任公司乌江渡发电厂, 贵州 遵义

收稿日期: 2026年3月26日; 录用日期: 2026年4月23日; 发布日期: 2026年4月30日

摘要

针对传统路面缺陷检测方法在复杂场景下实时性差、误检率高, 以及现有深度学习模型难以兼顾精度与边缘部署效率的问题, 本文构建了轻量化的路面缺陷检测算法GS-MobileNet。基于MobileNetV4-Ghost的主干网络架构。通过引入Ghost模块的廉价线性变换操作, 在保证特征表达能力的同时显著降低模型参数量。创新设计的局部-全局协同注意力机制, 在浅层和中间层网络嵌入BoT模块强化裂纹边缘连续性建模, 深层采用SimAM无参注意力动态抑制背景噪声。进一步通过BiFPN动态加权多尺度融合策略, 有效改善微小裂纹的检出能力, 并结合分组查询解码器实现分类-定位任务的高效对齐。最终构建GS-MobileNet模型以3.12 M参数量取得91.3%的精确率与92.1%的mAP。

关键词

路面缺陷检测, 深度学习, 轻量化, GS-MobileNet

A Lightweight Road Anomaly Detection Algorithm Based on GS-MobileNet

Jiantao Yu, Chuanyu Long, Fan Luo, Rensheng Zheng, Yuyu Tian

Wujiangdu Hydropower Plant, Guizhou Wujiang Hydropower Development Co., Ltd., Zunyi Guizhou

Received: March 26, 2026; accepted: April 23, 2026; published: April 30, 2026

Abstract

Addressing the issues of poor real-time performance and high false positive rates of traditional road defect detection methods in complex scenarios, as well as the difficulty of existing deep learning models in balancing accuracy and edge deployment efficiency, this paper constructs a lightweight road defect detection algorithm named GS-MobileNet. The algorithm is based on the MobileNetV4-Ghost backbone network architecture. By introducing the Ghost module's cheap linear transformation operations, it significantly reduces the number of model parameters while ensuring feature

representation capability. The innovatively designed local-global cooperative attention mechanism embeds the BoT module in shallow and middle layers to enhance crack edge continuity modeling, while the deep layers adopt the SimAM parameter-free attention to dynamically suppress background noise. Furthermore, the BiFPN dynamic weighted multi-scale fusion strategy effectively improves the detection capability for tiny cracks, and combined with the group query decoder, it achieves efficient alignment of classification and localization tasks. The final GS-MobileNet model, with only 3.12 M parameters, achieves a precision of 91.3% and an mAP of 92.1%.

Keywords

Road Defect Detection, Deep Learning, Lightweight, GS-MobileNet

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

交通基础设施的大力发展,促使中国公路、铁路和高铁建设事业以令人惊叹的速度飞速成长,也促使公路等基础设施在中国分布区域越来越广泛。公路投入使用的时间越长就越容易产生裂缝、坑洼损伤变形,这些损伤变形严重影响了车辆正常行驶[1]。准确检测路面缺陷并作出响应,可以真实地反映工程实体的质量,也可以减少经济损失,防止事故的发生。然而,路面缺陷的原因复杂且不确定,导致缺陷的形状、位置和宽度是随机和不规则的。路面极易受到周围环境的干扰,因此路面缺陷检测的研究一直很困难[2]。在早期,我国公路路面病害检测主要是由专业人员通过步行或驾驶车辆,沿道路进行实地检查,依靠肉眼观察和简单工具来识别和记录道路异常的方法。这种方法可以直接观察道路状况,对异常类型和严重程度有直观判断。但这种方法仍然存在不少问题,人工巡检速度慢,难以覆盖大面积道路网络,尤其是高速公路或偏远地区,检测结果依赖巡检人员的经验和判断,可能存在漏检或误检,并且巡检人员需要在车流中工作,存在一定的安全风险。所以亟需一种提升检测速度、提高检测精度的新技术[3],以减少检测成本、降低人工方法存在的主观性偏差。

在过去十年中,机器学习的发展与缺陷和损伤检测研究的兴趣有关。它被很好地应用于研究自动路面裂缝检测问题。此外,由于无需手动调整阈值参数,机器学习有助于开发参数选择。基于机器学习,产生了许多有关表面裂纹和坑洼损伤检测的方法,如聚类分析算法、支持向量机和随机森林。传统机器学习方法在路面缺陷检测中可解释性强、计算资源需求低,但其特征设计依赖经验、泛化能力有限也较为明显。随着深度学习技术的发展,传统机器学习方法逐渐被取代,但在特定场景(如小规模数据集或资源受限环境)中仍具有一定的应用价值。

深度学习的迅猛发展推动了计算机视觉技术的突破。图像分类、目标检测、语义分割等领域取得了重大进展。图像分类用于确定图像中出现的对象是什么。目标检测则旨在确定物体的位置和类别。语义分割的目标是为每个像素分配类别标签。将物体检测技术应用于路面缺陷检测中,能够在一定程度上消除施工过程中的安全隐患。与传统的人工巡检法相比,利用机器视觉和深度学习可以实现对道路表面缺陷的自动、精确辨识[4],保证了识别的准确性,同时提高了检测效率,显著降低了时间和人工成本。

因此利用深度学习技术对公路进行实时监测是未来的发展趋势,而道路异常检测作为公路监测的重要任务,其关键技术已经成为研究的重点。随着深度学习的发展,目标检测、语义分割等在计算机视觉任务中大放异彩,将其运用到道路异常检测,实现对公路实时监测信息自动化、智能化处理,得到更准

确的道路异常检测与评估结果，具有重要的工程意义。

2. GS-MobileNet 轻量级道路异常检测算法

面向嵌入式边缘计算平台的路面表观缺陷检测需求，骨干网络的轻量化重构成为衡量算法效能与部署可行性的关键技术路径。本文采用 MobileNetV4 [5] 结合 Ghost [6] 模块为骨干网络，通过引入改进特征金字塔网络、RTDETRDecoder 等模型策略，可在降低网络计算密度与内存占用的同时，实现边缘端毫秒级实时推理。针对轻量化网络固有的表征能力瓶颈，尽管基准模型缺陷检测精度较低，但通过局部 Transformer 与无参注意力特征校准的协同优化，可实现检测效能损失的可控性补偿，满足工业场景下精度-速度-成本的最优。通过以上技术路径，本文构建了轻量级路面缺陷自动检测算法 GS-MobileNet。

GS-MobileNet 的搭建过程主要包括特征提取网络、特征融合层、特征预测和损失计算四个部分。其网络结构如图 1 所示。首先，输入的道路缺陷图像通过特征提取网络层(Backbone)进行特征提取，随后将提取到的相关特征输入到特征融合网络层(Neck)进行融合，最后在预测层(Head)完成道路缺陷图像的分类预测。

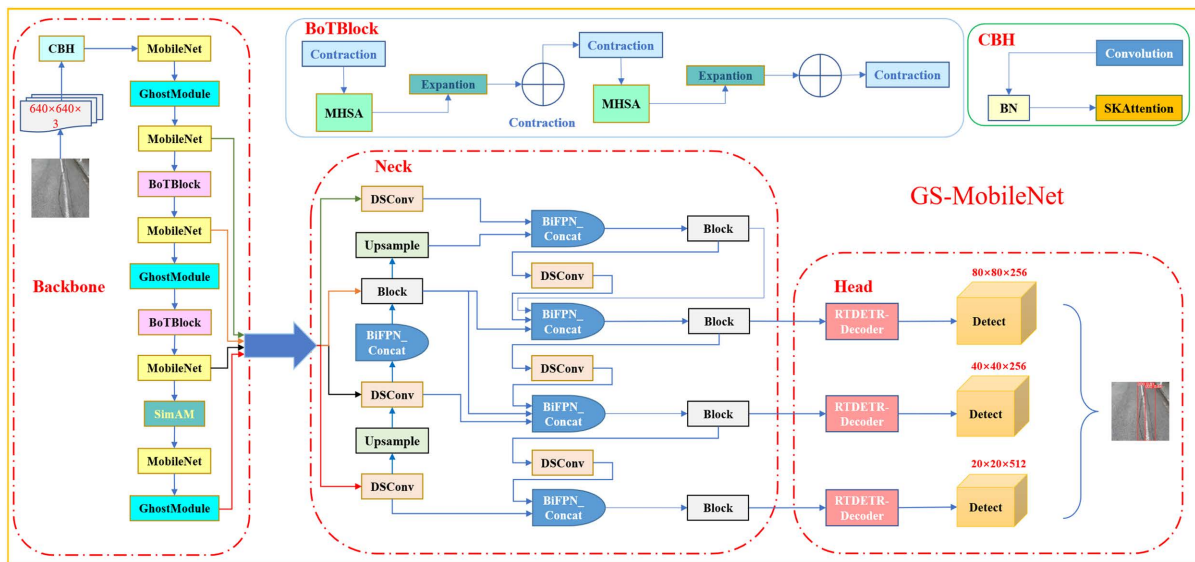


Figure 1. GS-MobileNet model architecture

图 1. GS-MobileNet 模型架构

2.1. 自适应多头注意力的轻量化特征提取网络

在本研究的目标检测算法中，特征提取网络选用了 MobileNetV4 与 Ghost 模块相结合的结构作为骨干网络。MobileNetV4 依托神经架构搜索(NAS)技术，在深度可分离卷积与倒残差结构的基础上实现了较优的轻量化性能。然而，这种轻量设计在提升效率的同时，可能存在浅层特征表达不够丰富、通道间信息冗余度不足等问题，进而限制了对微小裂纹等细粒度纹理信息的捕捉能力。为解决上述不足，本研究在 MobileNetV4 的瓶颈结构中引入 Ghost 模块，替代计算开销较大的 3×3 标准卷积。

MobileNetV4 结合 Ghost 模块依赖卷积的局部感受野，难以捕捉长程空间依赖，导致复杂背景下路面缺陷漏检率偏高。并且在路面缺陷检测模型的轻量化设计中，深层特征语义抽象与复杂背景干扰是核心挑战，本研究选择第 4 层与第 7 层嵌入 Bottleneck Transformers [7] 模块，并于特征提取第 9 层引入 SimAM [8] 无参注意力机制。该设计通过分层级特征增强策略，兼顾浅层细节保留与深层语义优化，同时避免计

算开销的显著增加,添加过程如图2所示。特征提取第4层处于 MobileNetV4 的浅层过渡阶段,特征图分辨率较高,主要提取边缘、纹理等局部细节;第7层定位于中层特征深化阶段,特征图分辨率适中,开始融合局部特征与初级语义信息。在浅层引入 BoTBlock,利用其局部窗口注意力机制,增强对路面缺陷微纹理的捕捉能力,通过局部窗口内像素关联建模,弥补传统卷积对非连续纹理的敏感度不足问题,使裂纹边缘响应强度提升。在中层特征中引入 BoTBlock,采用多头自注意力,在较低计算成本下建立跨区域语义关联。此位置的特征图仍保留一定空间细节,注意力机制可有效区分真实裂纹与背景伪影,降低误检率。

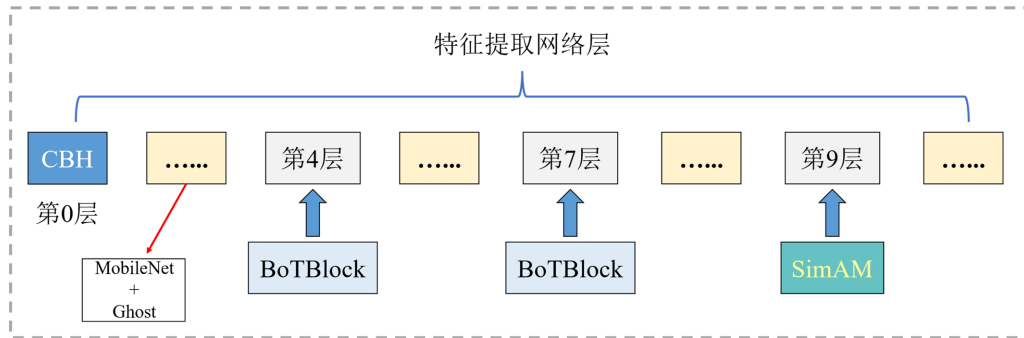


Figure 2. The process of adding BoT and SimAM
图2. BoT 和 SimAM 添加过程

特征提取第9层整合多尺度语义信息后,易受路面表面纹理噪声干扰。SimAM 模块基于能量函数理论,通过统计特征图能量分布动态聚焦关键区域进行动态校准,并且相比 SE、CBAM 等模块,SimAM 无需额外参数,避免在轻量化模型中引入存储负担。SimAM 注意力步骤如图3所示,其能量函数如式(2.1)所示。

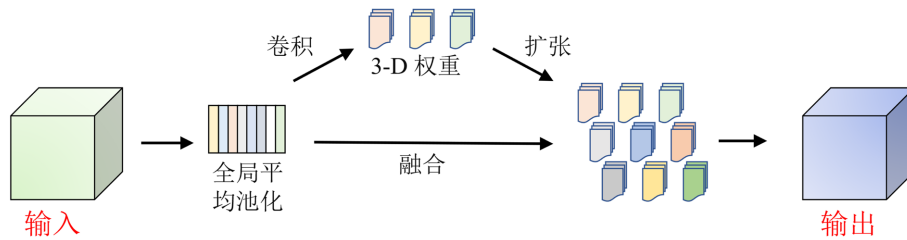


Figure 3. SimAM attention step
图3. SimAM 注意力步骤

$$e_i(w_i, b_i, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_i x_i + b_i))^2 + (1 - (w_i t + b_i))^2 + \lambda w_i^2 \quad (2.1)$$

式中, w_i 和 b_i 分别表示加权和偏置的变化, t 表示输入特征中单个通道的目标神经元, x_i 代表该通道内其他神经元, i 表示空间维度的索引, M 表示通道中神经元的数量。

自适应多头注意力的轻量化特征提取网络融合了局部细节与全局上下文信息,使模型在保持轻量化特性的同时,提升微观与宏观缺陷的检测精度,尤其改善了复杂背景下多尺度裂纹的共存检出率,为边缘端实时检测提供高性价比的特征增强方案。然而,在复杂环境中,例如阴影交错、光照变化或背景杂乱等因素的影响下,现有的检测方法仍然容易产生误检。这些复杂背景往往会干扰缺陷的准确定位,降

低检测模型的鲁棒性。因此，为了应对这些挑战，本研究提出了分布式双向加权特征金字塔网络。

2.2. 分布式双向加权特征金字塔网络

在本研究的目标检测算法中,特征融合网络采用了改进的 BiFPN [9]。BiFPN 网络结构图如图 4 所示。BiFPN 是一种高效的多尺度特征融合方法,通过双向信息流动来有效整合来自不同层次的特征图,从而提高了对不同尺寸目标的检测能力,但其核心依赖标准卷积的固有缺陷限制了实际应用效果。首先,标准卷积的固定几何采样模式难以适应跨尺度特征的空间偏移问题:在道路异常检测场景中,高层特征图因下采样操作易导致细长裂纹端点或小尺寸坑洼的位置漂移,而 BiFPN 融合不同分辨率特征时,标准卷积的刚性采样网格无法动态校正几何对齐误差,致使融合后特征存在语义错位,直接影响检测头的定位精度。其次, BiFPN 的堆叠结构加重计算负担:以 EfficientDet 为代表的模型通过重复堆叠 BiFPN 层提升性能,但每层标准卷积的高参数量导致计算开销随通道成倍增长,难以满足边缘设备的实时性需求。针对上述问题,本研究引入分布移位卷积 DSConv [10]对 BiFPN 进行了改进,改进后的 BiFPN 见图 1 Neck 部分。

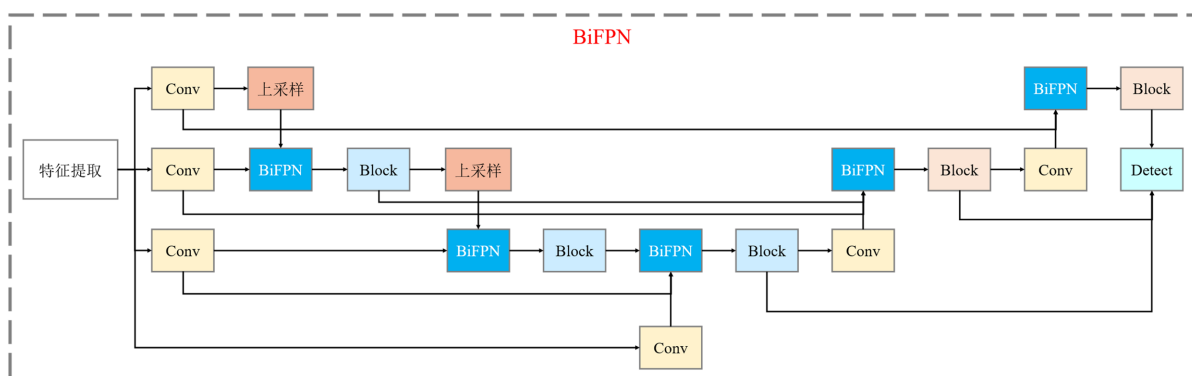


Figure 4. Original BiFPN network structure diagram

图 4. 原始 BiFPN 网络结构图

在道路异常检测任务中,本研究所提出的分布式双向加权特征金字塔网络具有以下优势:动态偏移补偿提升了跨尺度裂纹端点与坑洼边界的定位精度;参数量压缩与硬件友好计算使多级特征融合满足车载工控机的实时推理需求;而偏移场驱动的隐性注意力机制则有效抑制了复杂环境(如阴影交错)下的误检率。

为了进一步优化部署效率,本研究采用 RT-DETR Decoder 作为模型的检测头。RT-DETR Decoder 是一种轻量级、硬件友好的解码器,能够通过高效的特征解码和精确的目标定位机制,进一步加速推理过程,同时保持较高的检测精度。

Head 部分是 GS-MobileNet 用于最终目标检测的网络结构,经过特征融合后的特征图传入 Head 层进行预测,本研究使用 RT-DETR Decoder 作为模型的检测头。RT-DETR [11]解码器是专为实时目标检测任务优化的 Transformer 解码架构,其通过动态查询生成与分层特征交互设计,显著提升了检测效率与精度。

3. 实验设计与结果分析

3.1. 道路异常数据集

本文研究对象是普通公路中的道路异常,经过调查分析,路面裂纹和坑洼对公路使用具有显著危害。路面坑洼直接破坏路面平整性,降低行车舒适性,导致车辆颠簸甚至失控,还加剧车辆损耗,推高用户

维修支出。路面裂纹根据其形状又分为横向裂纹、纵向裂纹和鳄鱼裂纹，其中横向裂纹和纵向裂纹如果未及时修复，裂纹会扩展为大面积破损，大幅增加养护成本；鳄鱼裂纹蓄积雨水后增加打滑，严重时会引起交通事故。因此，本研究使用手机设备收集了纵向裂纹、横向裂纹、鳄鱼裂纹、路面坑洼四类作为检测对象，各类路面异常的示例如下图 5 所示。本研究对收集的图像数据进行了图像点处理、直方图处理，并利用生成模型生成裂纹图像。



Figure 5. Pavement anomaly type
图 5. 路面异常类型

图像点处理是一种数据增强方法，通过翻转、旋转、裁剪、变形、缩放等操作增加样本数量，同时不改变图像的原始内容。点处理方式如图 6 所示。

图像直方图是用于表示数字图像亮度分布的统计工具，在一个坐标轴内，横坐标代表 255 个灰度阶，纵坐标代表灰度阶出现的次数，如图 7 所示，一个像素为 9 的图像，该图像存在 1、2、3、251 四个灰度阶，出现的次数分别为 3、1、2、3。在绘制直方图时，将灰度级作为 x 轴处理，该灰度级出现的次数作为 y 轴处理，那么这个像素为 9 的图像可以绘制出如图 7 所示的柱状图，称为直方图。在道路异常检测任务中，图像直方图处理是提升数据质量的关键步骤之一。通过对图像亮度分布进行调整，可以有效增强目标区域的可见性，同时减少光照变化对模型性能的影响。

本研究对原始图像分别进行直方图均衡化、直方图拉伸、直方图匹配和直方图归一化处理，直方图均衡化：显著提升了低对比度区域的细节(如裂纹、坑洼)，但部分图像出现过亮或过暗现象；直方图拉伸：在保留图像整体亮度分布的同时，增强了全局对比度，适用于光照不均匀的场景；直方图匹配：将图像

亮度分布调整为高斯分布，模拟了不同光照条件下的道路场景，增强了数据多样性；直方图归一化：将图像亮度范围标准化到[0, 255]，消除了图像间的亮度差异，为模型提供了统一的输入。图像处理结果如图 8 所示。

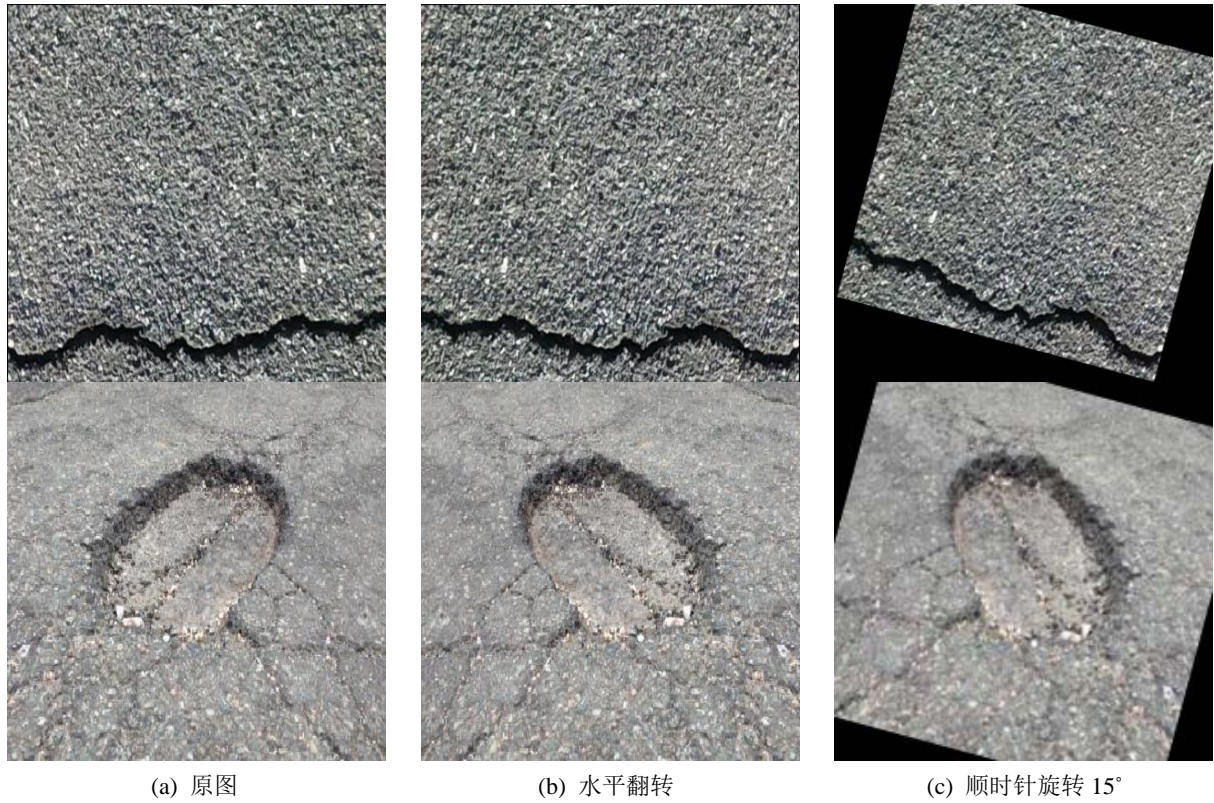


Figure 6. Image point processing
图 6. 图像点处理

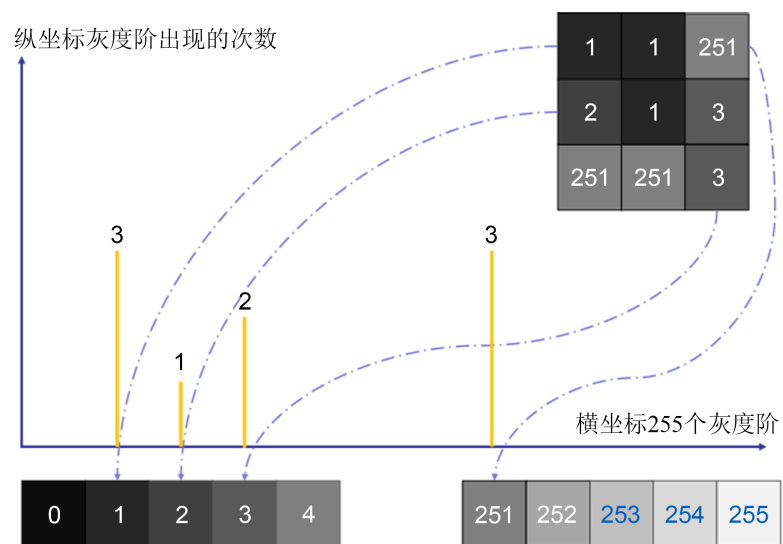


Figure 7. Example image of grayscale statistics
图 7. 灰度统计示例图

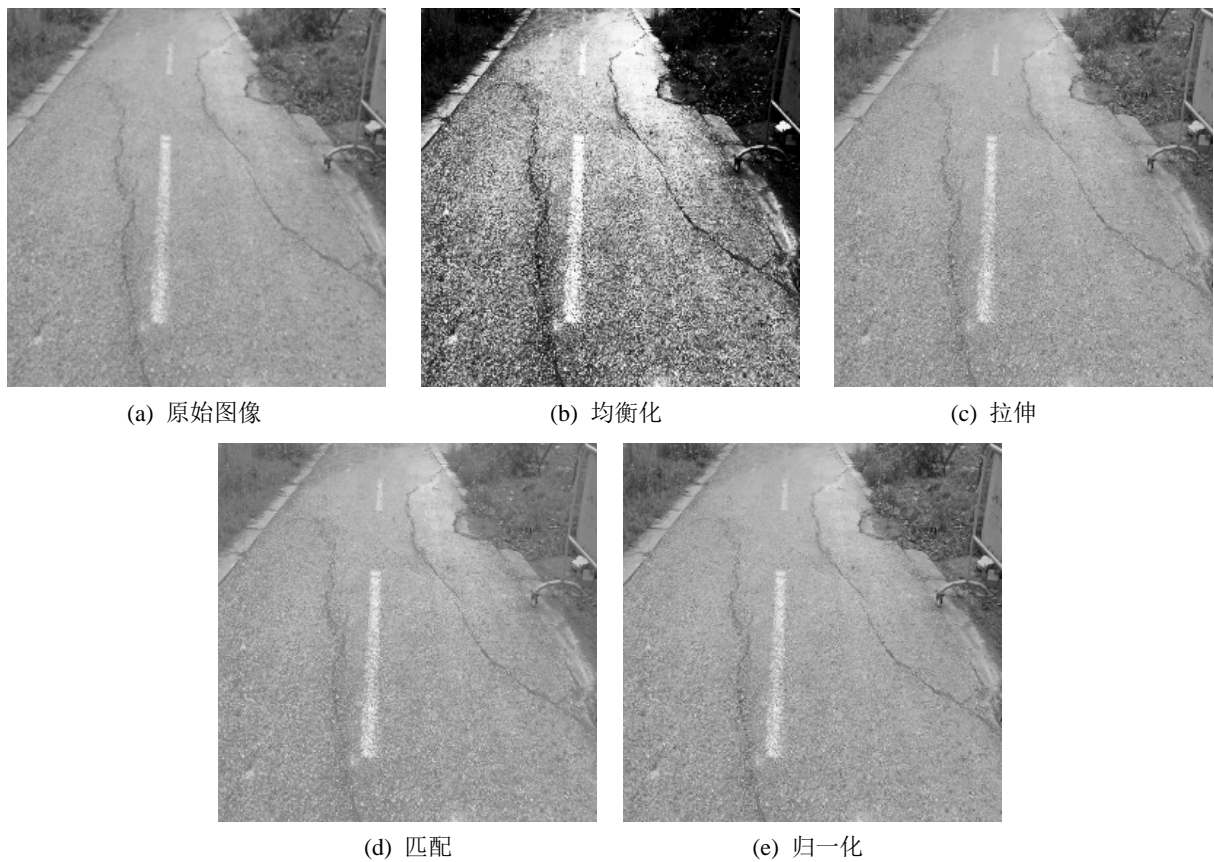


Figure 8. Image histogram processing result
图 8. 图像直方图处理结果

针对裂纹样本的形态多样性与空间分布复杂性，传统数据增强方法难以充分覆盖真实场景中裂纹的随机拓扑结构变化，尤其在样本量有限时易导致模型过拟合。为此，本研究采用生成对抗网络 DCGAN 对收集的路面裂纹图像进行训练，在对抗性训练范式中，生成器致力于合成高仿真裂纹样本以混淆鉴别器的判别能力，而鉴别器通过特征提取网络辨识输入样本的真伪属性。通过这种动态博弈机制，DCGAN 成功构建了具有多尺度特征的裂纹生成模型，其输出结果如图 9 所示，在纹理复杂度、边缘不规则性等形态学特征上与真实样本高度一致，主观视觉评估表明二者难以有效甄别。

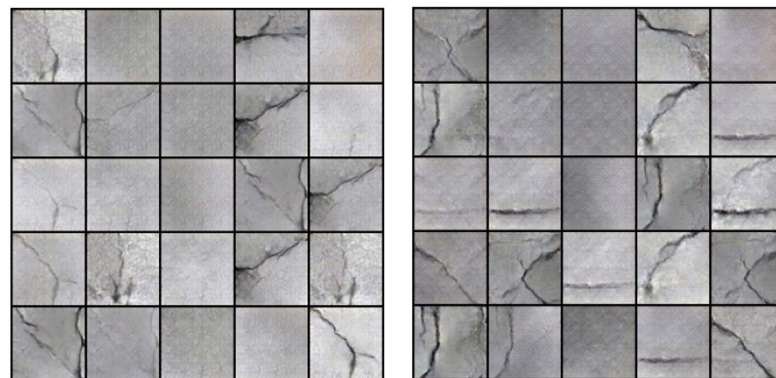


Figure 9. Crack samples generated by DCGAN
图 9. DCGAN 生成的裂纹样本

经过数据处理后，裂纹类别图像共计 4506 张，坑洼图像 2500 张。为了考虑模型结构和内存开销，在数据标注前，本研究使用无损尺寸变换工具将所有图像统一转换为 640×640 像素。

3.2. 实验环境

本研究的训练在个人计算机上进行，常见的深度学习框架有：Tensor Flow、NCNN 和 PyTorch 等。本研究所提出的 GS-MobileNet 以 PyTorch 作为深度学习框架，相关个人计算机配置和软件配置信息如表 1 所示。

Table 1. System resulting data of standard experiment
表 1. 相关配置信息

配置	属性
系统	Windows 11
CPU	Intel Core i9-10900k
GPU	NVIDIA RTX 3080
深度学习框架	PyTorch1.9
编程语言	Python3.8
CUDA	11.1
辅助库	Numpy、cv2 等

3.3. 参数设置

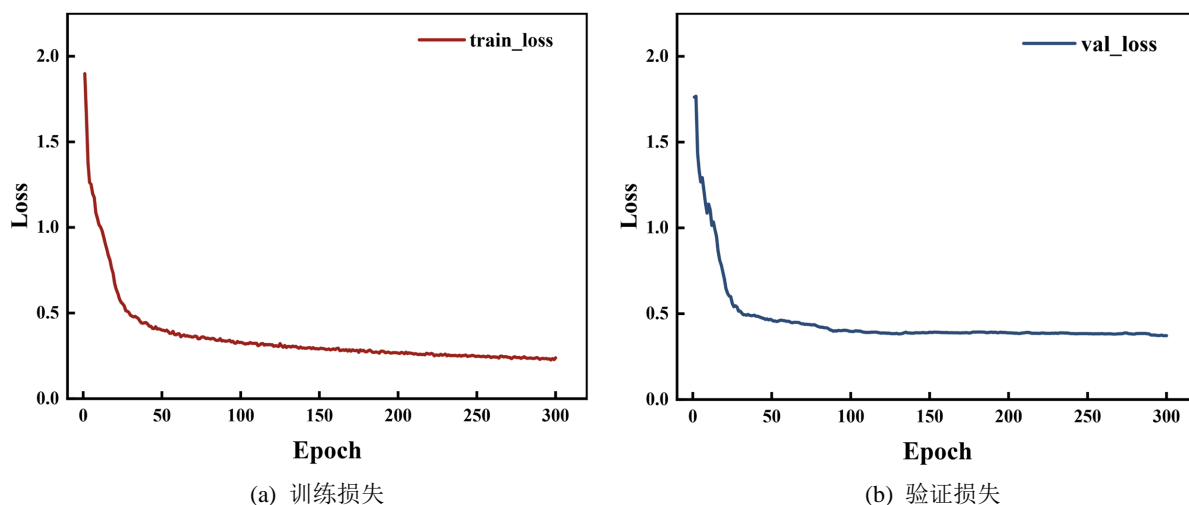


Figure 10. Training loss curve
图 10. 训练损失曲线

在模型训练过程中，选择合适的超参数对于优化性能至关重要。根据数据集的大小和复杂性，训练的总 epoch 通常在几十到几百之间。为了防止过拟合，本研究采用了早期停止策略，监控验证集的损失，当验证损失不再改善时，便停止训练。对于批处理大小，通常选择 16 或 32，以平衡内存限制和训练效

率。初始学习率一般设置为 10^{-3} 或 10^{-4} ，并结合学习率调度策略，动态调整学习率以提高模型性能。为了防止过拟合并增强模型的泛化能力，权重衰减通常设置为 10^{-4} 或 5×10^{-4} 。此外，最小学习率通常设置为初始学习率的 0.01 或 0.1，以便在训练后期进行微调。

本研究中，模型训练的 epoch 设置为 300，批量处理大小为 16。初始学习率设定为 10^{-3} ，权重衰减为 10^{-4} ，最大学习率为 0.01，最小学习率为 10^{-4} ，并采用 Adam 优化策略。考虑到初始学习率较小，本研究还引入了余弦退火学习率衰减策略，以在训练初期逐步减小学习率，直到网络参数趋于稳定后加速训练。随着训练的进行，模型逐渐逼近最优解，此时学习率逐步降低，以实现更精细的更新，从而使训练结果无限接近最优解。如图 10 所示，验证损失最终达到 0.372，训练损失为 0.238。

3.4. 评估指标

在视觉计算领域的性能评估中，检测框与标注框的空间重叠度是衡量定位精度的重要指标，通常通过交并比(Intersection over Union, IoU)来量化。如图 11 所示，IoU 值是通过计算标注框(实线)与预测框(虚线)的交集面积与并集面积之比来得出的。根据预设的临界值，IoU 值通常用于二分类判定：当检测框的 IoU 超过阈值时，判定为有效检测；否则，视为定位偏差。

鉴于检测任务包含目标与背景的二分类特性，评估指标通过检测结果与真实标注之间的对应关系，将样本划分为四类：真阳性(TP)、假阳性(FP)、假阴性(FN)和真阴性(TN)。这些类别的含义如图 12 所示，构成了完整的混淆矩阵评价体系。

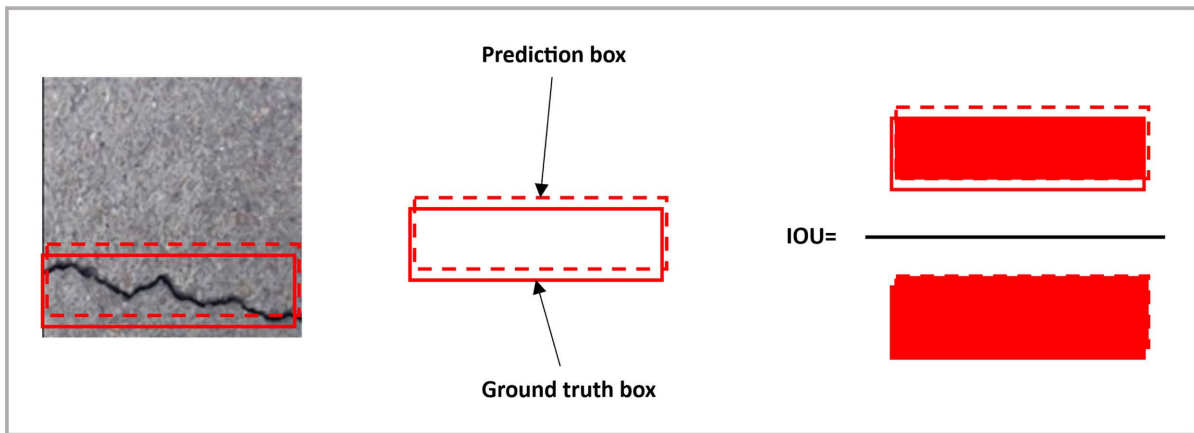


Figure 11. The calculation process of IoU
图 11. IoU 的计算过程

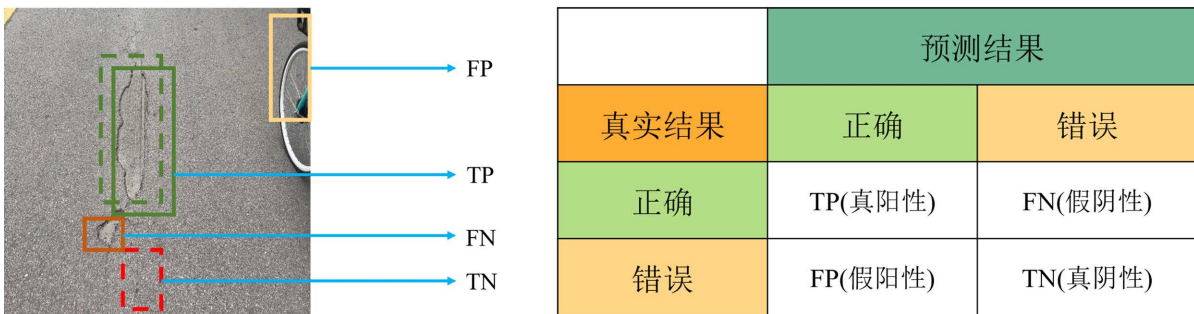


Figure 12. Confusion matrix
图 12. 混淆矩阵

目标检测模型的评价通常依赖于一系列标准。本研究也采用了这些常用的标准来评估模型性能，主要使用精确率(Precision)、召回率(Recall)和平均精度(mAP)作为评估指标。精确率反映了模型正确分类样本的概率，召回率则衡量了模型识别出所有相关样本的能力。精确率和召回率的计算公式分别如式(3.1)、式(3.2)所示。为了验证本研究提出的模型足够轻量化，能够满足边缘设备的部署要求，本研究还采用了参数量(Params)和计算成本(GFLOPs)作为额外评估标准。参数量直接影响模型的大小以及对计算资源和存储的需求；而 GFLOPs 表示每秒十亿次浮点运算，用于衡量算法或模型的计算复杂度。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.2)$$

3.5. 结果分析

为了平衡模型的评价指标、检测速度和参数，本研究进行了大量实验，不仅验证了各模块的有效性，还考虑了边缘设备部署所需的轻量化和高精度检测需求。实验主要验证了模型的关键模块，包括 Bottleneck Transformers 模块、SimAM 无参注意力机制，以及改进后的 Neck 和 Head。实验首先以 MobileNetV4-Ghost 特征提取网络、传统 PANet 特征融合网络和普通预测头作为基线模型进行验证，随后逐步加入其他模块。消融实验的结果见表 2。

Table 2. Ablation experiment results

表 2. 消融实验结果

编号	模型	Precision (%)	Recall (%)	mAP (%)	Para. (M)	GFLOPs
1	M4 + Ghost (backbone)	79.8	78.4	83.2	1.85	2.9
2	A + BoT (backbone)	82.3	80.2	85.6	2.38	3.5
3	B + Sim (backbone)	85.4	81.6	87.8	2.67	3.8
4	C + 改进 BiFPN (Neck)	88.2	82.4	89.6	2.95	4.1
5	D + RTDETRDecoder (Head)	91.3	83.3	92.1	3.12	4.3

本研究通过渐进式模块集成策略，系统验证了各改进组件在路面缺陷检测任务中的有效性。实验基于自建路面缺陷数据集，所有模型训练均采用相同超参数，MobileNetV4-Ghost 作为轻量化主干，以 1.85 M 参数量与 2.9 GFLOPs 的计算成本，实现基线性能：精确率 $P = 79.8\%$ ，召回率 $R = 78.4\%$ ， $mAP = 83.2\%$ 。在主干网络嵌入 BoT (Bottleneck Transformers) 模块后，模型参数量增至 2.38 M，计算量提升至 3.5 GFLOPs，但检测性能显著优化：精确率提升至 82.3%， mAP 提高至 85.6%。引入 SimAM 无参注意力机制后，模型以 0.29 M 的参数量增量，将精确率提升至 85.4%， mAP 达 87.8%。采用 BiFPN 结合 DConv 替代传统 PANet 进行多尺度特征融合后，模型在参数量与计算量小幅增加的条件下，精确率提升至 88.2%， mAP 突破 89.6%。最终模型搭载分组查询解码器，以 3.12 M 参数量与 4.3 GFLOPs 计算量，实现精确率 $P = 91.3\%$ 、 $mAP = 92.1\%$ 的性能突破。

上述的消融实验只能验证各个模块在模型中的有效性，但无法证明 GS-MobileNet 优于其他模型。因此，本研究进行了系列对比实验，以评估其相对表现。对比的模型主要包括目前在目标检测领域应用广

泛的一些端到端的单阶段目标检测模型，如 YOLOv5s [12]、YOLOv7 [13]、SSD 和 RT-DETR。这些模型在道路异常检测中能够提供快速响应，并且许多研究者在此基础上进行了创新与改进，将其广泛应用于各类检测任务。实验数据详见表 3。

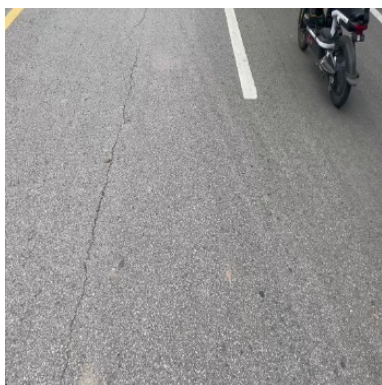
Table 3. Comparative test results
表 3. 对比试验结果

编号	模型	Precision (%)	Recall (%)	mAP (%)	Para. (M)	GFLOPs
3	YOLOv5s	83.2	80.5	0.861	7.1	16.3
4	YOLOv7	84.6	81.2	0.878	36.7	29.5
5	SSD	72.1	69.4	0.752	26.4	31.6
6	RT-DETR	92.9	84.7	0.942	32.8	98.6
7	GS-MobileNet	91.3	83.3	0.921	3.12	4.3

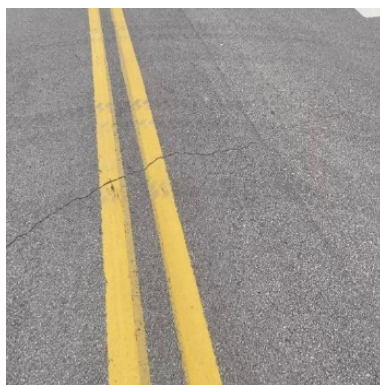
GS-MobileNet 以 0.921 mAP 位列第二，但其参数量仅为 RT-DETR 9.5% 的情况下，精度超越 YOLOv7 等经典模型，显示出轻量化设计的潜力。RT-DETR 凭借 Transformer 架构的全局建模能力，以 32.8 M 参数量与 98.6 GFLOPs 计算成本取得最高检测精度，但其庞大的计算密度导致边缘端部署可行性极低。传统 YOLO 系列模型呈现显著的性能与效率分化：YOLOv5s 以 7.1 M 参数量实现 0.861 mAP，其轻量化设计在精度与效率间取得平衡；YOLOv7 通过重参数化与动态标签分配策略，mAP 提升至 0.878，但参数量与计算量仍难以适配资源受限场景。早期模型 YOLOv3、YOLOv4 与 SSD 因架构冗余性问题，性能显著落后，且计算效率低下，暴露了传统检测范式的局限性。

对比实验表明，本研究提出的 GS-MobileNet 兼顾了检测精度与轻量化，在使得模型足够简单和轻量的同时，满足了道路异常实时检测精度的要求。

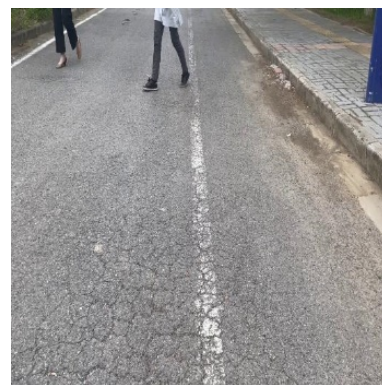
本文使用在路面上现场拍摄的图像，用于评估所提出的 GS-MobileNet 算法的运行性能。为了防止出现漏检或者错误检测情况，在本实验中，将 IoU 阈值设定为 0.5，模型检测结果的预测框得分值大于 0.5 则表示该预测框有效。在边缘设备上进行检测的部分原图如图 13 所示，图 13(a)为纵向裂纹图像，图 13(b)为横向裂纹图像，图 13(c)为鳄鱼裂纹图像，图 13(d)为积水的路面坑洼图像，图 13(e)为阴影遮挡的路面坑洼图像，图 13(f)图像为包含异物的裂纹图像。



(a)



(b)



(c)

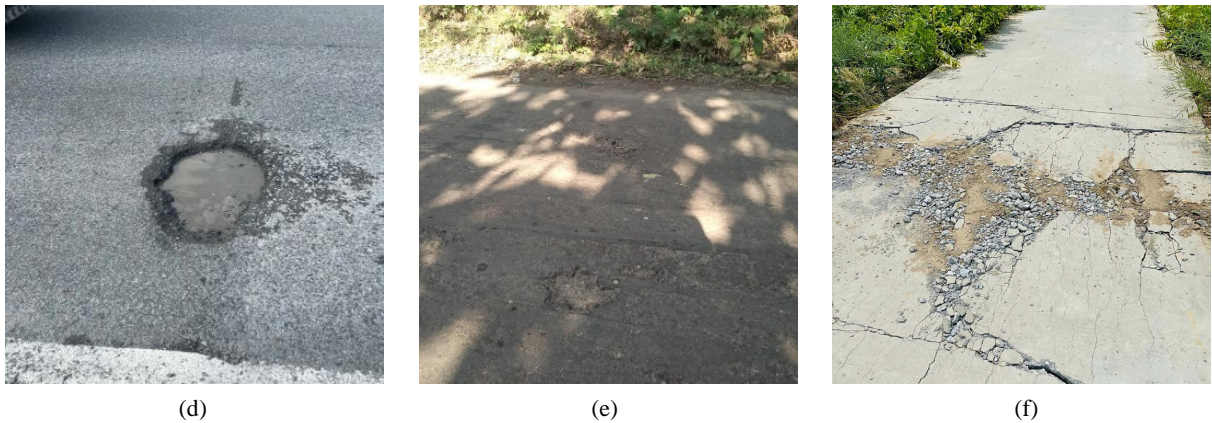


Figure 13. Detect original image
图 13. 检测原始图像

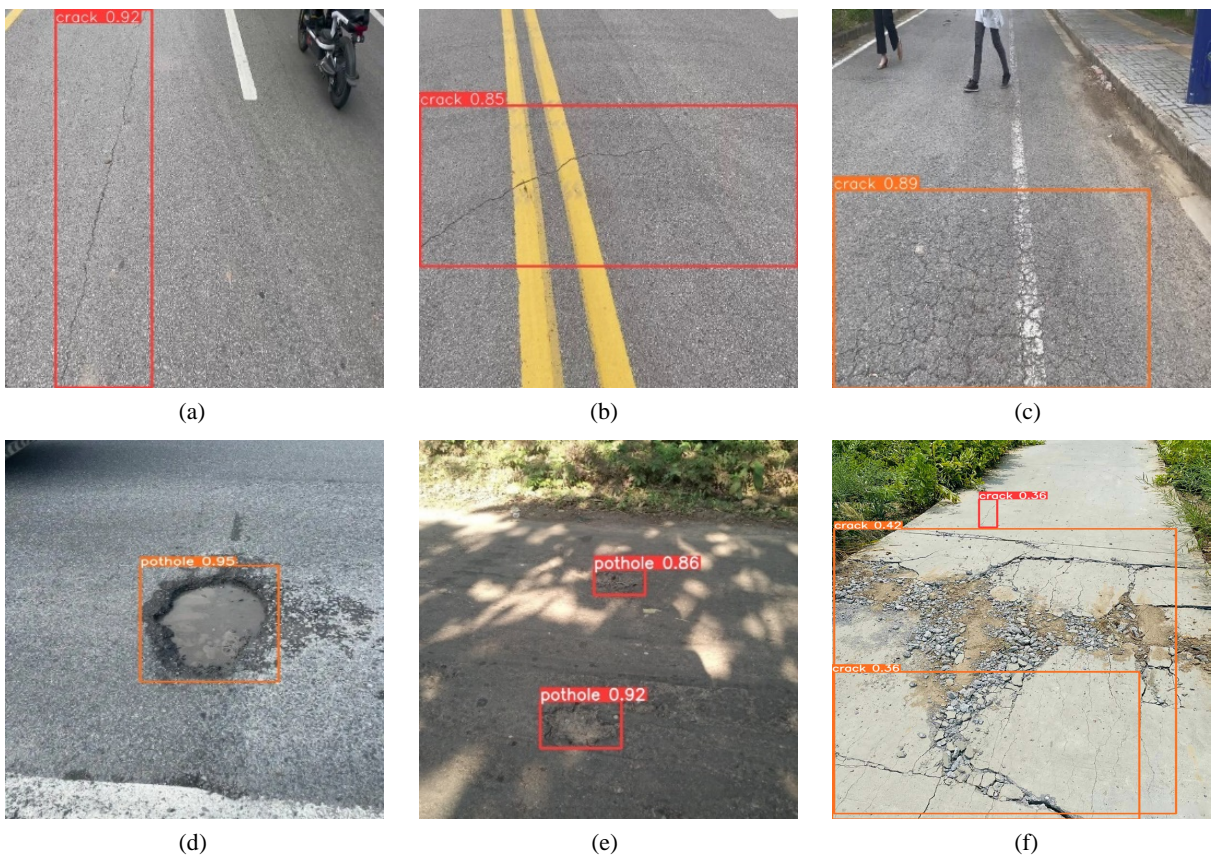


Figure 14. Test result image
图 14. 检测结果图

测试结果如图 14 所示。即使输入图像未经过预处理，GS-MobileNet 仍能准确识别图像中的缺陷类别，且各检测框的置信度得分均处于较高水平。从 GS-MobileNet 的可视化结果可以看出：Bottleneck Transformers 通过窗口注意力机制有效捕捉了裂纹的连续性拓扑结构；SimAM 无参注意力机制降低了长程裂纹的漏检率，尤其在复杂背景下，提升了多尺度裂纹的共存检出能力；改进后的 BiFPN 有效抑制了阴影区域的误检现象；RTDETRDecoder 则优化了分类置信度与定位精度之间的对齐关系，进一步提升了模型

的检测精度。因此, 尽管在图 14(b)和图 14(c)中存在标线干扰, 或图 14(e)中存在阴影干扰, GS-MobileNet 依然保持了稳定的检测效果。然而, 图 14(f)的检测效果较差, 表明当前数据集的类别多样性不足, 缺乏针对“负样本”或“干扰项”的系统性构建, 导致模型在遇到分布外样本时泛化能力下降。

4. 结语

本文针对道路场景实时检测需求, 提出基于 MobileNetV4-Ghost 的主干网络架构。通过引入 Ghost 模块的廉价线性变换操作, 在保证特征表达能力的同时显著降低模型参数量。创新设计的局部-全局协同注意力机制, 在浅层和中层网络嵌入 BoT 模块强化裂纹边缘连续性建模, 深层采用 SimAM 无参注意力动态抑制背景噪声。进一步通过 BiFPN 动态加权多尺度融合策略, 有效改善微小裂纹的检出能力, 并结合分组查询解码器实现分类-定位任务的高效对齐。最终构建的 GS-MobileNet 模型以 3.12 M 参数量取得 91.3%的精确率与 92.1%的 mAP。与 YOLO 和 RT-DETR 系列算法的对比验证了改进方法的有效性和优越性, 尤其在精度和实时性方面具备明显优势, 具有一定的理论意义与实际应用价值。

当前研究聚焦于异常检测环节, 后续可延伸至道路健康状态评估与养护决策支持。通过融合检测结果与路面材料特性、交通载荷等多元信息, 建立基于深度强化学习的养护决策模型。同时可开发可视化道路健康数字孪生系统, 为基础设施全生命周期管理提供技术支撑。

参考文献

- [1] Kirschke, K.R. and Velinsky, S.A. (1992) Histogram-Based Approach for Automated Pavement-Crack Sensing. *Journal of Transportation Engineering*, **118**, 700-710. [https://doi.org/10.1061/\(asce\)0733-947x\(1992\)118:5\(700\)](https://doi.org/10.1061/(asce)0733-947x(1992)118:5(700))
- [2] Li, P., Wang, C., Li, S. and Feng, B. (2015) Research on Crack Detection Method of Airport Runway Based on Twice-Threshold Segmentation. 2015 *Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*, Qinhuangdao, 18-20 September 2015, 1716-1720. <https://doi.org/10.1109/imccc.2015.364>
- [3] Oliveira, H. and Correia, P.L. (2009) Automatic Road Crack Segmentation Using Entropy and Image Dynamic Thresholding. *European Signal Processing Conference (EUSIPCO)*, Glasgow, 24-28 August 2009, 622-626.
- [4] 周强, 脱羚, 王莹, 等. 基于多特征信息融合的陶瓷墙地缺陷在线检测[J]. 硅酸盐通报, 2019, 38(1): 259-264.
- [5] Qin, D., Lechner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., et al. (2024) Mobilenetv4: Universal Models for the Mobile Ecosystem. In: Leonardis, A., et al., Eds., *Computer Vision—ECCV 2024*, Springer, 78-96. https://doi.org/10.1007/978-3-031-73661-2_5
- [6] Tang, Y., Han, K., Guo, J., et al. (2022) GhostNetV2: Enhance Cheap Operation with Long-Range Attention. *Proceedings of the 36th International Conference on Neural Information Processing Systems*, New Orleans, 28 November-9 December 2022, 9969-9982.
- [7] Srinivas, A., Lin, T.Y., Parmar, N., et al. (2021) Bottleneck Transformers for Visual Recognition. <http://arxiv.org/abs/2101.11605>
- [8] Yang, L., Zhang, R.Y., Li, L., et al. (2021) SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Proceedings of the 38th International Conference on Machine Learning, PMLR*, Vol. 139, 11863-11874.
- [9] Tan, M., Pang, R. and Le, Q.V. (2020) Efficientdet: Scalable and Efficient Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 10778-10787. <https://doi.org/10.1109/cvpr42600.2020.01079>
- [10] Gennari, M., Fawcett, R. and Prisacariu, V.A. (2019) DSConv: Efficient Convolution Operator. 2019 *IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019*, Seoul, 27-28 October 2019, 5148-5157. <http://arxiv.org/abs/1901.01928>
- [11] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al. (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 16965-16974. <https://doi.org/10.1109/cvpr52733.2024.01605>
- [12] Jocher, G. (2020) YOLOv5 by Ultralytics. <https://github.com/ultralytics/yolov5>
- [13] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) Yolov7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/cvpr52729.2023.00721>