DOI:10.26974/j.cnki.XBGC.2026.03.010

基于深度学习的西北地区临水混凝土病害智能识别

王笑言, 杨淑雁*
(宁夏大学 土木与水利工程学院,宁夏 银川 750021)

摘 要:针对西北地区临水钢筋混凝土结构病害检测存在的高成本、低效率及主观性强等问题,提出了一种基于卷积神经网络(CNN)的多阶段智能识别方法。通过融合像素级调整、空间域变换与几何变换的混合数据增强策略,将550张原始图像扩增至3 840张,有效缓解了实际工程中小样本数据的泛化问题。以EfficientNet-B3为主干网络,引入EnhancedDyConvSE模块中的增强型动态卷积与SE注意力机制,并结合混合权重方法,显著提升了对多样化病害形态特征的提取能力。结果表明,该模型在自建临水结构病害数据集上取得了98.18%的准确率与98.20%的精确率,能够准确识别盐结晶、坑蚀、剥落、露筋锈蚀和其他病害等5类典型病害,其性能显著优于ResNet-50、DenseNet-121等主流网络。

关键词:混凝土病害;图像扩增;卷积神经网络(CNN);图像识别;智能检测
中图分类号:TP183;TU528 **文献标志码**:A

西北地区临水钢筋混凝土(reinforced concrete,RC)柱在硫酸盐侵蚀与冻融循环耦合作用下易发生多种混凝土病害[1],其精准识别对该区域桥梁结构的健康监测具有重要意义。传统混凝土病害检测方法(如人工巡检、锤击法、超声波检测及钻芯取样)存在经济成本高、检测效率低的问题,且桥梁水下结构检测伴随较大安全风险。随着人工智能技术的发展,支持向量机(support vector machine,SVM)已在混凝土裂缝损伤识别中取得初步应用,准确率可达约80%[2]。此外,几何变换、高斯滤波降噪等图像预处理方法可有效提升图像信噪比,进而改善损伤识别效果[3-4]。卷积神经网络(convolutional neural network,CNN)凭借其自动化与高精度特性,为基础设施智能检测提供了可靠路径[5];深度卷积神经网络(deep convolutional neural network,DCNN)则在裂缝之外的多种混凝土表观病害识别中展现出广泛应用潜力[6]。近年来,基于注意力机制的Transformer架构在计算机视觉领域取得突破性进展[7]。尽管深度学习在混凝土裂缝、孔洞等表观病害检测中已取得一定成果,但针对硫酸盐侵蚀-冻融循环耦合作用下的混凝土损伤识别研究仍相对薄弱,现有文献涉及化学-物理耦合侵蚀的案例较为有限,且缺乏标准化的评估数据集。

本研究以西北地区临水RC柱为研究对象,通过对现场采集的混凝土病害图像进行预处理与多维度数据增强,提升小样本条件下的模型泛化能力。基于扩增后的数据集,系统对比ResNet-50、DenseNet-121、MobileNet-V1和EfficientNet-B3 4种CNN架构,筛选出参数量更少、识别精度更高的模型用于病害识别。针对模型在特定病害类别上识别精确率较低的问题,设计消融实验,引入增强型动态卷积压缩激励模块(enhanced dynamic convolution with squeeze-and-excitation,EnhancedDyConvSE)注意力机制,并结合类别权重优化策略,有效提升相应类别的识别性能。最终构建基于Python的检测框架,支持病害实时分类,以期实现西北地区临水RC柱混凝土病害准确识别。

收稿日期:2026-03-06
基金项目:宁夏自然科学基金重点项目(2025AAC020028)
作者简介:王笑言(2001—),女,硕士研究生,主要从事基于人工智能的硫酸盐腐蚀混凝土损伤识别研究(2635690658@qq.com)。
* **通信作者**:杨淑雁(1973—),女,副教授,博士,主要从事钢筋混凝土结构耐久性研究(yangshuyan@nxu.edu.cn)。
引用格式:王笑言,杨淑雁.基于深度学习的西北地区临水混凝土病害智能识别[J].西北工程技术学报(中英文),2026,25(3):266-276.

1　图像分类

图像是混凝土病害识别的基础,其质量直接关系到识别准确性。本文所采用的基于图像的混凝土病害识别整体框架如图 1 所示。主要包括以下环节:通过智能手机现场拍摄获取原始图像;对原始图像的尺寸、格式、亮度及角度进行调整,扩充图像库;图像标注环节采用 Label Studio 1.0.8 进行人工病害标注,同时引入 InceptionV3 模型对未标注图像进行预分类,最后人工修正分类错误的样本;使用扩增后的数据集训练并评估不同模型,选取最优模型后测试图像识别准确率。图像扩增的方法主要包括像素级操作(亮度调整、对比度调整、直方图均衡化及伽马变换)、空域操作(高斯滤波)以及几何变换(图像旋转)。

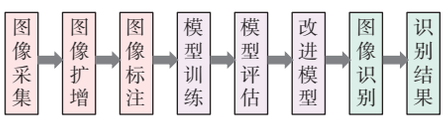


图 1　研究整体框架图

Fig. 1　Overall research framework

1.1　数据集划分

基于本课题组前期对宁夏全区临水 RC 柱混凝土病害调研所采集的图像,本文从中选取 550 张具有代表性的照片作为研究样本。前期研究已将该地区混凝土病害按严重程度划分为 4 个等级:①盐结晶(表面有白色沉淀物);②表面坑蚀(分布有圆形小坑);③保护层脱落(混凝土保护层成片剥落);④核心混凝土溃散(露筋锈蚀或结构偏移)。这些病害主要由硫酸盐物理结晶侵蚀与冻融循环耦合作用引起[1]。在图像分类过程中发现,不同等级的病害常同时出现,例如坑蚀伴随盐结晶、保护层脱落伴随盐结晶等。本研究借助 ImageJ 软件对每张图像进行损伤区域手动划分,并依据面积主导原则确定类别标签。此外,为避免因训练集标签“完全正确”导致模型泛化能力下降,增设“其他”类别以增强分类鲁棒性。最终基于 550 张原始图像构建数据集,划分为盐结晶、坑蚀、保护层脱落、露筋锈蚀和“其他”5 个类别,每类 110 张。

1.2　图像扩增

针对实际采集图像中存在的光照、角度与尺寸差异等问题,本研究采用图像增强方法对数据集进行扩增,以提升模型在小样本条件下的泛化能力。具体图像扩增方法如下。

1.2.1　像素级增强　①彩色变换。彩色变换通过调整图像亮度与对比度实现。亮度调整对每个像素的 RGB 通道进行线性缩放;对比度调整以中灰度为基准,采用线性插值方法进行缩放。②直方图均衡化。直方图均衡化[8]是一种广泛应用的灰度增强算法,其目的是使图像灰度分布趋于均匀,从而增强整体对比度。③伽马变换。伽马变换属于非线性图像处理方法,通过调整灰度相应曲线修正图像的对比度和亮度分布。

1.2.2　空间域滤波　高斯滤波[8]是一种线性平滑滤波方法,基于高斯正态分布函数计算滤波核各位置的权重,以实现图像噪声抑制与平滑效果。

1.2.3　几何变换　本研究采用绕 z 轴旋转的图像几何变换方法[8],以图像中心为原点,按设定角度旋转,以增强模型对方向变化的适应能力。

1.2.4　评价指标对比　为综合评价图像增强效果,本研究采用峰值信噪比(peak signal-to-noise ratio,PSNR)、结构相似性指数(structural similarity index ,SSIM)、均方误差(mean squared error,MSE)和归一化均方误差(normalized mean squared error,NMSE)4 项指标进行量化对比。针对盐结晶、坑蚀、保护层脱落、露筋锈蚀及其他病害等 5 类混凝土病害,将不同扩增方法处理后的图像评价指标对比结果列于表 1。各评价指标公式为

$$\mathrm{PSNR}=10\times\lg\left(\frac{I_{\max}^2}{\mathrm{MSE}}\right),\tag{1}$$

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)+(\sigma_x^2+\sigma_y^2+C_2)},\tag{2}$$

$$\mathrm{MSE}=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-K(i,j)]^2,\tag{3}$$

$$\mathrm{NMSE}=\frac{\sum_{i=1}^{T}(x_i-y_i)^2}{\sum_{i=1}^{T}x_i^2}。\tag{4}$$

式中:$I_{\max}$ 是原始图像 I 的像素最大值,设为 255;μ_x 和 μ_y 分别为图像 x 和 y 的局部均值;σ_x^2 和 σ_y^2 分别为图像 x 和 y 的局部方差;σ_{xy} 为协方差;C_1 和 C_2 为稳定常数;$I(i,j)$ 为原始图像在位置 (i,j) 处的像素值;$K(i,j)$ 为处理后图像在位置 (i,j) 处的像素值;M 是对应图像的行数(高度方向的像素数量),N 对应图像的列数(宽度方向的像素数量);T 为图像总像素数目,其中像素总数 $T=M\times N$;$\sum_{i=1}^{T}(x_i-y_i)^2$ 为预测误差的平方和;$\sum_{i=1}^{T}x_i^2$ 为原始图像所有像素值的平方和。

表1 不同图像扩增方法下的评价指标对比

Tab. 1 Comparison of evaluation metrics for different image augmentation methods

图像扩增方法	病害类型	峰值信噪比/dB	结构相似性指数	均方误差	归一化均方误差
调整亮度	盐结晶	17.02	0.80	1 290.32	0.05
	坑蚀	28.20	0.94	98.31	0.01
	保护层脱落	20.47	0.88	583.55	0.04
	露筋锈蚀	18.05	0.84	1 018.70	0.05
	其他	28.27	0.93	96.79	0.00
	求和	112.01	3.63	3 087.67	0.15
调整对比度	盐结晶	22.02	0.81	407.88	0.02
	坑蚀	25.43	0.86	186.16	0.01
	保护层脱落	31.65	0.89	44.45	0.00
	露筋锈蚀	20.67	0.73	557.59	0.03
	其他	28.39	0.93	94.11	0.00
	求和	128.16	4.22	1 290.19	0.06
直方图均衡化	盐结晶	17.05	0.78	1 282.45	0.05
	坑蚀	17.26	0.89	1 222.86	0.09
	保护层脱落	15.40	0.83	1 874.85	0.13
	露筋锈蚀	21.90	0.88	420.04	0.02
	其他	20.77	0.86	544.41	0.03
	求和	92.88	4.24	5 344.61	0.34
伽马变换	盐结晶	11.03	0.66	5 126.87	0.22
	坑蚀	11.86	0.46	4 236.44	0.31
	保护层脱落	11.74	0.51	4 355.25	0.31
	露筋锈蚀	11.66	0.56	4 431.61	0.23
	其他	11.59	0.56	4 510.59	0.24
	求和	57.88	2.75	22 660.76	1.31
高斯滤波	盐结晶	30.06	0.79	64.18	0.00
	坑蚀	31.38	0.89	47.33	0.00
	保护层脱落	28.33	0.79	95.60	0.01
	露筋锈蚀	26.16	0.72	157.50	0.01
	其他	26.00	0.82	163.27	0.01
	求和	141.93	4.01	527.88	0.03
几何变换	盐结晶	9.19	0.30	7 841.59	0.33
	坑蚀	11.02	0.39	5 139.70	0.37
	保护层脱落	10.71	0.32	5522.46	0.39
	露筋锈蚀	9.68	0.19	7 002.20	0.36
	其他	7.20	0.18	1 2392.08	0.66
	求和	47.80	1.38	37 898.03	2.11

峰值信噪比(PSNR):用于衡量增强后的图像相对于原始图像的失真程度,其值越大,表示图像质量越好,失真越小。结构相似性指数(SSIM):基于人眼视觉系统特性,从亮度、对比度和结构三方面评估两幅图像的相似性,取值范围为[0,1],值越大表示图像越相似。均方误差(MSE):计算两幅图像像素值之间的平均平方偏差,值越小说明图像间差异越小。归一化均方误差(NMSE):对均方误差进行归一化处理,反映误差能量相对于原始图像信号能量的比例,值越小表示误差所占比例越低。

鉴于各图像扩增方法数值差异较大,本文在每个评价指标下对 5 种病害类型的数值求和,以此对比不同扩增方法的整体性能。由表 1 可知:高斯滤波的 PSNR 求和值最大,表明其失真程度最小;而几何变换的 PSNR 求和值最小,失真程度最大。在结构相似性方面,直方图均衡化的 SSIM 求和值最高,图像相似性最好;几何变换的 SSIM 求和值最低,相似性最差。进一步从误差指标来看,高斯滤波的 MSE 和 NMSE 求和值均为最小,说明其重建误差最小、失真程度最低;相反,几何变换在这两个指标上求和值最大,表明其误差最大、失真最显著。

由表 1 还可知:同一图像扩增方法在不同病害类型上表现存在差异。以亮度调整为例,“其他”类别 PSNR 值最大,失真较小;而“盐结晶”类别 PSNR 值最小,失真较大。这主要源于不同病害在图像特征上的差异性。

1.2.5　*图像扩增实例*　对原始图像进行预处理,将所有图像统一缩放至 224 px×224 px。随后,针对 5 类病害图像,分别采用原图、亮度调整(系数 0.4)、对比度调整(系数 0.7)、直方图均衡化、伽马校正(γ=0.4)、高斯滤波及随机旋转(0°~30°)等方法进行图像扩增,结果如图 2 所示。图 2a 展示了临水柱表面盐结晶病害。由图 2a 可知,对比度调整和伽马校正后,损伤区域的视觉效果更为显著。图 2b 展示了坑蚀病害。由图 2b 可知,亮度调整和高斯滤波有效抑制了图像噪声,并增强了病害区域的对比度,从而显著提升了低质量图像中坑蚀特征的信噪比。图 2c 展示了保护层脱落病害。由图 2c 可知,直方图均衡化将图像转换为灰度形式,更直观地显示出缺失部分。图 2d 展示了露筋锈蚀病害。由图 2d 可知,经伽马变换后,钢筋与混凝土之间的对比度显著增强,同时图像旋转操作引入了不同角度的钢筋形态,有效增加了样本多样性。图 2e 展示了其他病害类型。病害类型包括桥梁、路肩等构筑物中出现的典型腐蚀病害。

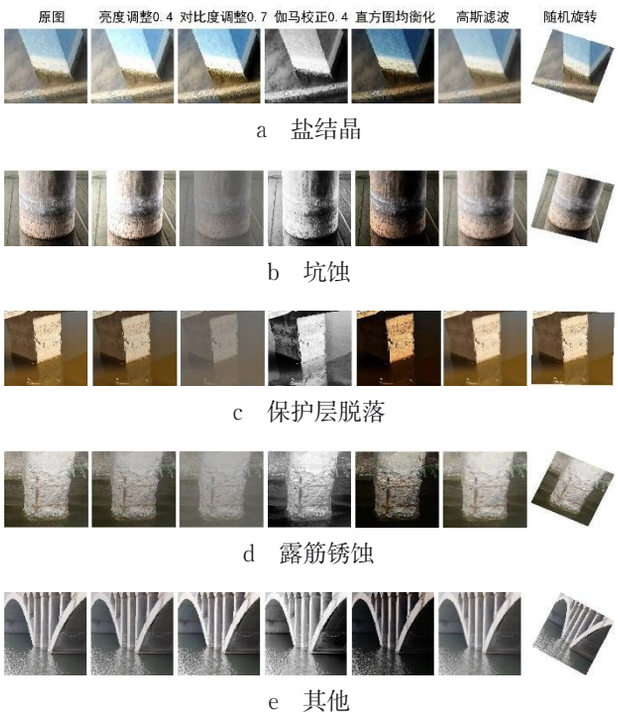


图 2　5 类混凝土病害的自建 Damage 数据库

Fig. 2　Self-constructed Damage dataset containing five types of concrete defects

经过图像扩增,图片总量由 550 张增加至 3 850 张,每类病害均为 770 张。为确保测试集和验证集的样本数量可被批量大小(32)整除,避免训练中出现不完整批次,删除 10 张图片,最终保留 3 840 张图像作为总样本。图像扩增完成后,基于 Label Studio (v1.0.8)对数据集进行标注。首先,对 550 张原始图像进行人工病害识别与标注,标注结果导出为 project.csv 文件。本地图像上传至 Label Studio 后,会被二次处理与存储,致使文件路径与扩增后数据集不一致,因此使用 Python 脚本将路径校正为统一命名格式,校正后的标注文件保存为 corrected_label.csv。随后,利用 InceptionV3 模型对剩余的 3 290 张扩增图像进行预标注,自动标注结果保存为 auto_label.csv。最后,通过 Python 脚本根据 CSV 标注文件自动将图像归类至相应病害类别文件夹,并对错误分类的样本进行人工修正。

2　模型训练

2.1　划分数据集及训练模型介绍

本文选取 4 种具有代表性的卷积神经网络(CNN)架构进行对比实验,包括 ResNet-50[9]、DenseNet-121[10]、MobileNet-V1[11]和 EfficientNet-

B3[12]。实验数据共 3 840 张图像,按 8∶1∶1 的比例划分为训练集、验证集和测试集[13],分别为 3 072、384、384 张。各模型均在上述数据集上进行训练与评估。

2.2 优化模型

在卷积神经网络(CNN)中,超参数调整是优化模型性能的关键环节。本实验涉及的主要超参数包括全连接层神经元数量、学习率、L2 正则化系数、Dropout 率及迭代轮数。经过调优后,4 种模型采用的超参数配置如下。

(1)ResNet-50[9]:全连接层包含 128 个神经元,学习率为 0.000 7,L2 正则化系数为 0.000 2,Dropout 率为 0.5,迭代轮数初始值为 5,最大值为 30。

(2)DenseNet-121[10]:全连接层包含 256 个神经元,学习率为 0.001,L2 正则化系数为 0.000 6,Dropout 率为 0.5,迭代轮数初始值为 5,最大值为 30。

(3)MobileNet-V1[11]:全连接层包含 256 个神经元,学习率为 0.000 1,L2 正则化系数为 0.000 2,Dropout 率为 0.5,迭代轮数初始值为 5,最大值为 30。

(4)EfficientNet-B3[12]:全连接层包含 128 个神经元,学习率为 0.001,L2 正则化系数为 0.000 2,Dropout 率为 0.4,迭代轮数初始值为 5,最大值为 30。

2.3 模型评估

2.3.1 不同模型架构的性能对比与效率分析 4 种模型架构在训练集和验证集上的性能评价如表 2 所示。模型参数量(params)直接影响模型存储大小;浮点运算次数(FLOPs)反映模型训练和推理过程中的计算复杂度;准确率(accuracy)是分类任务中最基本的性能指标,由 TensorFlow/Keras 框架的内置评估机制计算。由表 2 可知,ResNet-50 模型的参数量和浮点运算次数均为最大,MobileNet-V1 模型的值为最小。这主要源于两者基础架构的差异:ResNet-50 模型通过标准卷积堆叠与残差连接实现跨层恒等映射,有效缓解了梯度消失,但计算成本较高;MobileNet-V1 模型采用深度可分离卷积,将空间滤波与通道特征融合分离,大幅降低了计算负担。EfficientNet-B3 模型的参数量虽高于 DenseNet-121 模型,但由于采用复合缩放机制协调网络深度、宽度和分辨率,其浮点运算次数反而更低,实际训练和推理时间更短。从训练集和验证集的准确率来看,EfficientNet-B3 模型均是最高。

表 2 4 种模型架构在训练集和验证集中的性能评价

Tab. 2 Performance of four model architectures on the training and validation sets

模型	参数量/10^6个	浮点运算次数/10^9次	准确率/%	
			训练集	验证集
ResNet-50[9]	23.85	3.82	98.96	96.61
DenseNet-121[10]	7.30	2.64	99.48	92.45
MobileNet-V1[11]	3.49	0.47	99.35	94.01
EfficientNet-B3[12]	10.98	0.92	99.61	97.14

图 3 展示了验证集上不同模型的训练时间与准确率对比气泡图。图 3 采用多维可视化形式,在二维散点图基础上以气泡大小表示准确率,颜色区分不同模型。由图 3 可知:红色气泡准确率最高,绿色最低;在计算效率方面,绿色训练时间最长,黄色最短,橙色模型耗时适中且准确率较好。综合考虑,红色所代表的 EfficientNet-B3 模型整体表现最优。

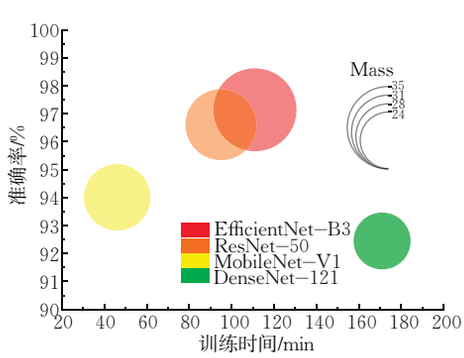


图 3 验证集上不同模型训练时间与准确率的对比

Fig. 3 Comparison of the training time and validation accuracy of different models

2.3.2 不同模型在测试集上的泛化性能评估 在深度学习中,模型评估是确保其性能、泛化能力和可靠性的关键环节。混淆矩阵[13]作为一种重要的分类模型评估工具,通过对比预测结果与真实标签,能够直观反映分类中的正确与错误情形。对于二分类任务,混淆矩阵包含 4 个基本指标:真正例(true positive,TP)、真反例(true negative,TN)、假正例(false positive,FP)和假反例(false negative,FN)。

其中,TP 指实际为正样本且被正确分类的数目;TN 指实际为负样本且被正确分类的数目;FP 指实际为负样本但被错误预测为正的数目;FN 指实际为正样本但被错误预测为负的数目。对于多分类任务(类别数 $C>2$),混淆矩阵扩展为 $C\times C$ 方阵,其行代表预测类别,列代表真实类别。该评估框架普遍适用于各类分类模型,如 SVM、CNN 等。基于混淆矩阵,可进一步计算多种分类评估指标,包括准确率(accuracy)、精确率(precision)、召回率(recall)和 F_1 分数(F_1-score)。准确率为正确预测的样本占总样本的比例;精确率为预测为正的样本中实际为正的比例;召回率为实际为正的样本中被正确预测的比例;F_1 分数为精确率和召回率的调和平均值。各指标计算公式为

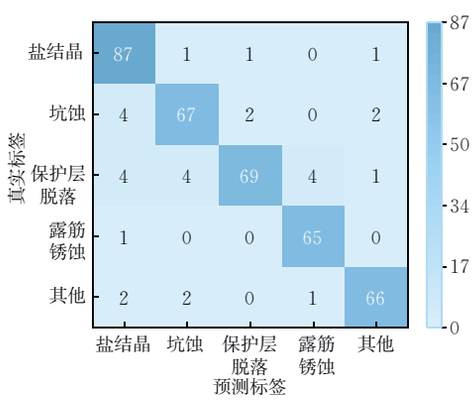
$$\text{accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}},\tag{5}$$

$$\text{precision}=\frac{\text{TP}}{\text{TP}+\text{FP}},\tag{6}$$

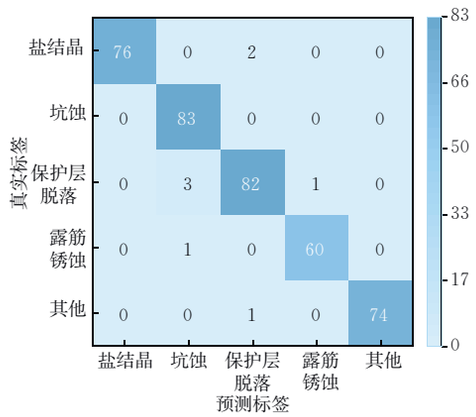
$$\text{recall}=\frac{\text{TP}}{\text{TP}+\text{FN}},\tag{7}$$

$$F_1\text{-score}=2\times\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}。\tag{8}$$

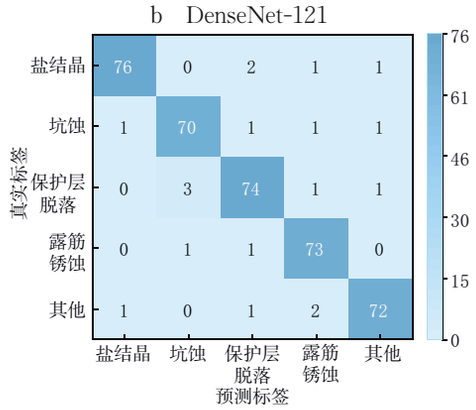
基于测试集,4 种模型训练所得的可视化混淆矩阵如图 4 所示。矩阵行表示预测标签,列表示真实标签,类别包括盐结晶、坑蚀、保护层脱落、露筋锈蚀以及其他。图 4 中纵轴对应真实类别,横轴对应预测类别。主对角线元素代表正确分类的样本数量,其余位置则表示错误分类的样本数量。图 4 中右侧附有颜色条,其数值范围从 0 到预测正确的最大样本数,颜色越深表示该位置对应的样本数量越多,反映预测类别的计数分布。从 4 个混淆矩阵的对角线数值可以看出:ResNet-50 和 DenseNet-121 模型的对角线值较为接近,分类准确率相当;MobileNet-V1 模型表现较好;EfficientNet-B3 模型的对角线值最高,分类准确率最优。



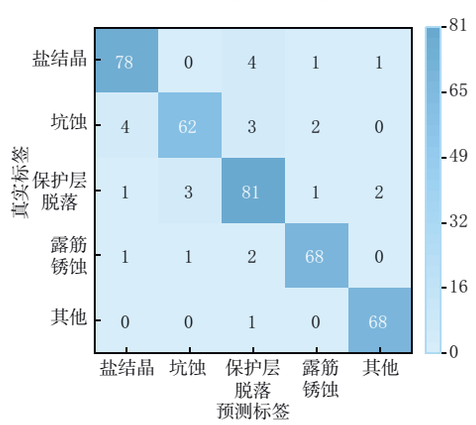
a　ResNet-50



b　DenseNet-121



c　MobileNet-V1



d　EfficientNet-B3

图 4　4 种模型在测试集上的混淆矩阵对比

Fig. 4　Comparison of the confusion matrices of four models on the test set

基于测试集混淆矩阵,通过相应公式计算每类病害的各项评价指标,4 种模型的性能对比结果如表 3 所示。从准确率、精确率、召回率和 F_1 分数等 4 项指标(表 3)综合来看,EfficientNet-B3 在所有模型中表现最优,在测试集上的准确率达到 97.14%,与验证集一致,表明该模型具有较强的泛化能力。

2.4　损伤评估

基于测试集数据,本文采用 EfficientNet-B3 模型对不同等级病害进行评估,各项指标如表 4 所示。

表 3 4种模型在测试集上的评估指标

Tab. 3 Evaluation metrics of four models on the test set

模型	准确率/%	精确率/%	召回率/%	F_1分数/%
ResNet-50[9]	94.79	93.00	92.97	92.95
DenseNet-121[10]	92.19	92.35	92.19	92.13
MobileNet-V1[11]	95.31	95.08	95.05	95.05
EfficientNet-B3[12]	97.14	97.69	97.66	97.66

以保护层脱落病害类别为例，precision=82/(2+0+82+0+1) =0.964 7，recall=82/(0+3+82+1+0)=0.953 5，F_1-score=0.959 1。由表 4 可知，5 种病害类别的精确率(precision)、召回率(recall)和 F_1分数(F_1-score)均超过 90%，其中盐结晶(efflorescence)、露筋锈蚀(disintegration)和其他(others)精确率较高，坑蚀(pitting)和保护层脱落(spalling)的精确率相对较低。精确率差异可能源于这两类病害在形态或表观特征上存在一定相似性，导致模型易对其产生混淆。此外，坑蚀(pitting)召回率最高，说明其漏检数量较少；其他(others)F_1分数最优，表明其精确率与召回率的综合性能最佳。

表 4 EfficientNet-B3模型在各病害类别上的评估指标

Tab. 4 Class-wide evaluation metrics of the EfficientNet-B3 model

损伤分类	精确率/%	召回率/%	F_1分数/%
盐结晶	98.70	97.44	98.07
坑蚀	95.40	100.00	97.65
保护层脱落	96.47	95.35	95.91
露筋锈蚀	98.36	98.36	98.36
其他	100.00	97.37	98.67

2.5 基于 EfficientNet-B3 的改进模型

2.5.1 *改进模型架构设计与训练性能分析* 由表 4 可知，EfficientNet-B3 模型在坑蚀和保护层脱落类别上预测精确率较低，主要归因于样本分布不均衡及模型对关键类别关注不足。为改善这一性能瓶颈，改进后的模型引入 EnhancedDyConv 中的增强型动态卷积[14]与 SE 注意力模块[15]，分别通过在 MBConv Stage 1(网络浅层)、MBConv Stage 3(网络中层)和 MBConv Stage 5(网络深层)实现，以强化混凝土病害关键特征提取；同时，结合混合权重损失函数(通过 DyConv 2 改进模块实现)来解决样本不均衡问题，并采用 Grad-CAM 实现病害区域可解释性，改进后的网络架构如图 5 所示。

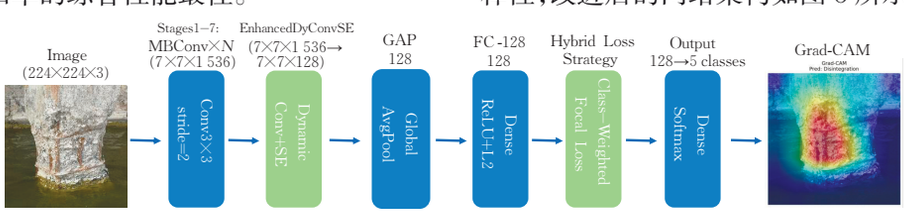


图 5 改进模型网络架构图

Fig. 5 Architecture of the improved model

混合权重损失函数融合了类别权重与 Focal Loss 损失函数[16]，一方面通过类别权重校正整体类别失衡，另一方面利用 Focal Loss 调制因子[17](γ=2.0)增强了模型对困难样本的关注。具体的损失函数[16]为

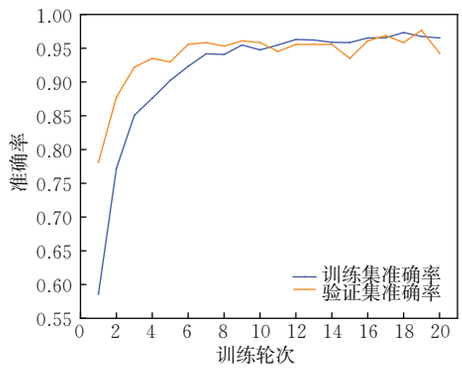
$$L=-\frac{1}{n}\sum_{i=1}^{n}w_{y_i}(1-P_i)^{\gamma}\ln(P_i)。\tag{9}$$

式中：P_i表示模型对第 i 个真实类别的预测概率；$(1-P_i)^{\gamma}$为 Focal Loss 调制因子(γ=2.0)，w_{y_i}为类别权重；n 为训练过程中 1 个批次的样本总数(120)。

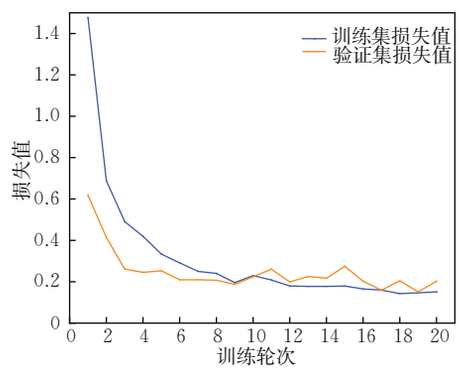
图 6 为改进模型的训练集和验证集收敛曲线。由图 6 可知：随着训练轮次增加，训练集与验证集准确率均快速上升，并最终稳定在 96% 以上；损失值同步下降并趋于收敛，且二者未出现显著分离。结果表明，改进模型训练过程稳定，未发生过拟合，具备良好的收敛性能与泛化能力。

改进模型对盐结晶、坑蚀、保护层脱落、露筋锈蚀和其他病害等 5 类病害在各模块下提取的特征如图 7 所示，从左到右依次为网络浅层、网络中层、网络深层及 DyConv 2。由图 7 可知，网络浅层模块主要提取图像边缘、轮廓与纹理等基础视觉特征，对混凝土表面细微变化与结构边界较为敏感；网络中层模块将浅层特征进一步抽象为局部结构特征，开始聚焦病害区域整体形态与空间分布；网络深层模块能够初步实现对病害关键区域的聚焦，抑制部分背景与无关区域的响应；DyConv 2 改进模块在低分辨率特征图上实现更精细的病害区域聚焦，桥墩、水面、阴影等无关背景区域响应值被显著降低。

2.5.2 *改进模型评估与消融实验分析* 改进模型



a　准确率



b　损失值

图6　改进模型的训练集和验证集收敛曲线

Fig. 6　Convergence curves of the improved model on the training and validation sets

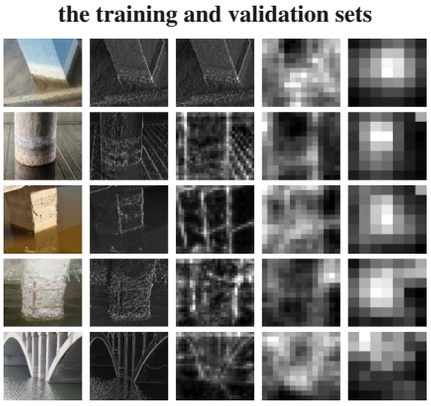
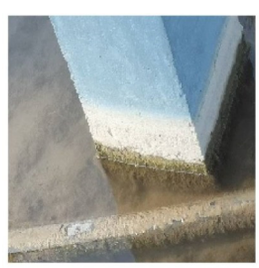


图7　改进模型对5种病害在各模块下提取的特征图

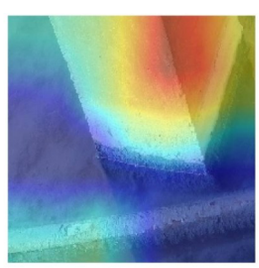
Fig. 7　Feature maps extracted by each module of the improved model for five types of defects

测试集整体准确率和精确率分别提高了 1.04% 和 0.51%;坑蚀和保护层脱落类别上的精确率分别提升了 0.94% 和 2.28%。为验证改进模型的可解释性,采用 Grad-CAM 方法[18]对模型决策依据进行可视化分析,结果如图 8 所示。图 8 中颜色越偏向红色、黄色,表明模型在分类过程中对该区域关注度越高。由图 8 可知,对于 5 类病害样本,模型高关注区域均聚焦于病害核心位置,证明模型能够有效捕捉不同病害的关键特征。部分样本热力图在水面、构件边缘等背景区域也出现一定程度关注,这是由于水面反光、构件轮廓等强边缘特征对模型造成干扰,但模型的最高关注仍集中于病害区域,未发生误判。这说明改进模型决策并非依赖无关噪声或全局背景,而是建立在对病害区域有效特征学习之上,进一步验证了模型的可解释性。

随后,本文设计了 1 组消融实验,通过逐步移除或调整模型组件评估各部分对整体性能的贡献。该实验包含 3 种注意力机制以增强特征表达能力,具体包括:高效通道注意力(efficient channel attention, ECA)[19],采用一维卷积构建轻量化的通道注意力模块,能够灵活嵌入现有 CNN 架构,在几乎不增加计算成本的情况下提升模型表征能力;无参注意力机制(simple improved attention mechanism,



a　盐结晶(原始)



b　盐结晶(Grad-CAM)



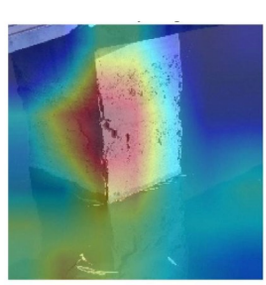
c　坑蚀(原始)



d　坑蚀(Grad-CAM)



e　保护层脱落(原始)



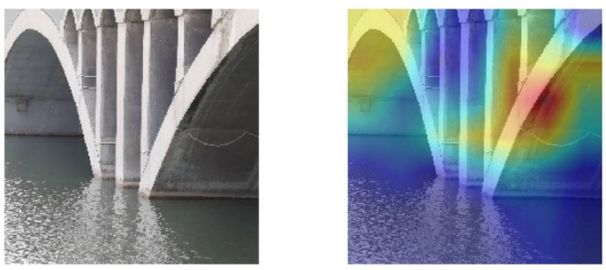
f　保护层脱落(Grad-CAM)



g　露筋锈蚀(原始)



h　露筋锈蚀(Grad-CAM)



i 其他(原始) j 其他(Grad-CAM)

图8 改进模型的Grad-CAM热力图可视化结果

Fig. 8 Grad-CAM heatmap visualizations of the improved model

SIMAM)[20],基于能量函数推导出 3D 注意力权重,无需引入额外参数即可为特征图中每个神经元分配重要性,兼具轻量化与抗过拟合特性;改进模型所采用的 EnhancedDyConvSE。实验结果如表 5 所示。

由表 5 可知,在 EfficientNet-B3 中引入 EnhancedDyConvSE 注意力机制效果最佳,准确率分别较 ECA 和 SIMAM 增加了 1.30%、1.83%,充分验证了本文方法的有效性。尽管该模型"参数量(params)"较大,但"运行时间(time)"却较短,其准确率达 98.18%,精确率达 98.20%,分别较引入前增加了 1.04% 和 0.51%。

本文进一步采用最优改进模型对不同混凝土病害进行评价,结果如表 6 所示。由表 6 可知,各项指标均显著提升:坑蚀精确率从 95.40% 提高至 96.34%,增加了 0.94%;保护层脱落精确率从 96.47% 提高至 98.75%,增加了 2.28%。

3 图像识别

本文将 EfficientNet-B3+EnhancedDyConvSE 模型在测试集上进行推理,预测结果保存为"EfficientNet_Best_Model.keras"文件。重新评估结果显示,模型准确率达 97.91%,精确率为 97.92%,与保存前性能差异均小于 1%。将该模型加载至 Python 部署环境后,对输入图像进行预测,仅需 24 s 即可返回预测类别及其对应置信度。置信度是指模型依据所学特征权重,对输入图像属于特定类别的确信程度。图 9 和图 10 分别展示了基于数据集外图像(宁夏地区及区外病害)的预测结果示例。由图 9~图 10 可知,模型对宁夏区外部分病害图像识别置信度较低,而在宁夏区域内表现良好,预测置信度较高,反映出其泛化能力仍受训练数据地域覆盖范围限制。该模型主要存在以下局限性:①模型训练依赖于原始数据集,尽管经数据扩增后样本量达 3 840 张,但核心来源仅为 550 张原始图像,多样性仍显不足;②图像采集条件(如拍摄距离、角度、光照及尺

表5 3种注意力机制的消融实验对比

Tab. 5 Comparison of ablation experiments on three attention mechanisms

模型	参数量/10^6个	时间/min	准确率/%	精确率/%	召回率/%	F_1分数/%
EfficientNet-B3	10.98	111	97.14	97.69	97.66	97.66
EfficientNet-B3+ECA	10.98	136	96.88	96.73	96.61	96.63
EfficientNet-B3+SIMAM	10.98	134	96.35	96.49	96.35	96.36
EfficientNet-B3+EnhancedDyConvSE	25.75	126	98.18	98.20	98.18	98.18

表6 最优改进模型在各病害类别上的评估指标

Tab. 6 Evaluation metrics of the optimal improved model for each defect category

损伤分类	精确率/%	召回率/%	F_1分数/%
盐结晶	97.62	98.80	98.21
坑蚀	96.34	100.00	98.14
保护层脱落	98.75	97.53	98.14
钢筋锈蚀	98.39	96.83	97.60
其他	100.00	97.44	98.70



图9 宁夏地区临水柱随机病害图像识别结果

Fig. 9 Recognition results for randomly selected images of defects in waterside columns in Ningxia

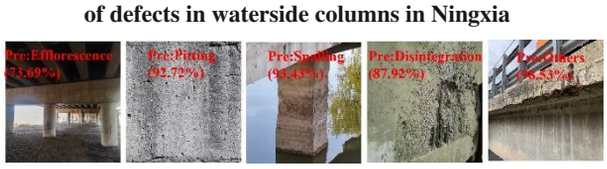


图10 宁夏区外临水结构随机病害图像识别结果

Fig. 10 Recognition results for randomly selected images of defects in waterside structures outside Ningxia

寸)对识别精度影响显著,需保持采集环境的一致性以提升模型稳定性。

4　结论

针对西北地区临水 RC 柱混凝土病害识别问题,本文基于前期采集的图像,采用图像增强技术对原始小样本数据集进行扩增,并选取 4 种 CNN 模型进行训练、验证与测试,依据评价指标筛选出最优模型。对于特定病害识别效果不佳的问题,进一步通过消融实验对优选模型进行改进,主要结论如下。

(1)综合采用亮度调整、对比度调整、直方图均衡化、伽马变换、高斯滤波及图像旋转等增强策略,将原始 550 张图像扩增至 3 840 张,使用 ResNet-50、DenseNet-121、MobileNet-V1 和 EfficientNet-B3 共 4 种 CNN 模型对扩增后数据集进行训练。结果表明,EfficientNet-B3 性能最优,训练集准确率为 99.61%,验证集准确率为 97.14%,测试集精确率、召回率和 F_1 分数分别为 97.69%、97.66% 和 97.66%。

(2)对于 EfficientNet-B3 在坑蚀和保护层脱落类别上精确率偏低的问题,设计消融实验对比 3 种注意力机制,发现引入 EnhancedDyConvSE 模块的 EfficientNet-B3 模型表现最佳,测试集的准确率和精确率分别提升至 98.18% 和 98.20%,较原模型提高了 1.04% 和 0.51%。坑蚀病害的精确率由 95.40% 提升至 96.34%,保护层脱落的精确率由 96.47% 提升至 98.75%,增幅分别为 0.94% 和 2.28%。

(3)EfficientNet-B3+EnhancedDyConvSE 模型测试完成后保存,在测试集上准确率达 97.91%,精确率为 97.92%,与保存前性能差异均小于 1%,展现出良好的泛化能力。部署于 Python 环境后,可实现对新输入图像中混凝土病害的快速识别,且输出置信度较高。

参考文献:

[1] 宋俊杰. 宁夏地区在役临水钢筋混凝土柱腐蚀现状调研和剩余极限承载力预测研究[D]. 银川: 宁夏大学, 2022.

[2] 杨扬, 王连发, 张宇峰, 等. 基于多特征融合的混凝土结构表面病害图像分类算法[J]. 长安大学学报(自然科学版), 2021, 41(3): 64-74.

[3] 唐钱龙, 谭园, 彭立敏, 等. 基于数字图像技术的隧道衬砌裂缝识别方法研究[J]. 铁道科学与工程学报, 2019, 16(12): 3041-3049.

[4] 谢明志, 樊丁萌, 蒋志鹏, 等. 基于计算机视觉的混凝土结构裂缝检测研究现状与展望[J/OL]. 西南交通大学学报, 1-20 [2026-07-24]. https://link.cnki.net/urlid/51.1277.u.20240913.1853.002.

[5] 刘鸿达, 孙旭辉, 李沂滨, 等. 基于卷积神经网络的图像分类深度学习模型综述[J]. 计算机工程与应用, 2025, 61(11): 1-21.

[6] Chen L K, Chen W X, Wang L, et al. Convolutional neural networks (CNNs)-based multi-category damage detection and recognition of high-speed rail (HSR) reinforced concrete (RC) bridges using test images[J]. Engineering Structures, 2023, 276: 115306.

[7] Han K, Wang Y H, Chen H T, et al. A survey on vision transformer [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (1): 87-110.

[8] 蔡体健, 刘伟. 数字图像处理: 基于Python[M]. 北京: 机械工业出版社, 2022: 54-130.

[9] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2016: 770-778.

[10] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 2261-2269.

[11] Howard A G, Zhu M L, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 1-9.

[12] Tan M X, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks [C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach. PMLR, 2019: 6105-6114.

[13] Feng D C, Yi X, Deger Z T, et al. Rapid post-earthquake damage assessment of building portfolios through deep learning-based component-level image recognition [J]. Journal of Building Engineering, 2024, 98: 111380.

[14] Chen Y P, Dai X Y, Liu M C, et al. Dynamic convolution: Attention over convolution kernels [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2020: 11027-11036.

[15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE,

2018: 7132-7141.

[16] Malarvannan S, V P, Venkatraman S, et al. A novel adaptive hybrid focal-entropy loss for enhancing diabetic retinopathy detection using convolutional neural networks [PP/OL]. V2. arXiv (2025-04-23) [2026-07-24]. https://doi. org/10.48550/arXiv. 2411. 10843.

[17] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]// IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 2999-3007.

[18] Aminudin M A I, Abdullah M N, Mustapha F, et al. Explainable deep learning framework for binary corrosion image classification using grad-CAM [J]. Sensors, 2025, 25(22): 7070.

[19] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2020: 11531-11539.

[20] Yang L X, Zhang R Y, Li L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning. Los Angeles. PMLR, 2021: 11863-11874.

Deep-Learning-Based Intelligent Recognition of Defects in Waterside Concrete Structures in Northwest China

WANG Xiaoyan, *YANG Shuyan**

(School of Civil and Hydraulic Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: To address the high cost, low efficiency, and strong subjectivity associated with defect detection in waterside reinforced-concrete structures in Northwest China, a multistage intelligent recognition method based on a convolutional neural network (CNN) was proposed. A hybrid data augmentation strategy integrating pixel-level adjustments, spatial-domain transformations, and geometric transformations was employed to expand the original dataset from 550 to 3, 840 images, effectively alleviating the poor generalization associated with small datasets in practical engineering applications. With EfficientNet-B3 as the backbone network, the EnhancedDyConvSE module, incorporating enhanced dynamic convolution and a squeeze-and-excitation (SE) attention mechanism, was introduced and combined with a hybrid weighting method. This approach significantly enhanced the model's ability to extract morphological features of diverse defects. The results showed that the model achieved an accuracy of 98.18% and a precision of 98.20% on a self-constructed dataset of defects in waterside structures. It accurately recognized five typical types of defects, including salt crystallization, pitting, spalling, and exposed and corroded reinforcement, and significantly outperformed mainstream networks such as ResNet-50 and DenseNet-121.

Keywords: concrete defects; image augmentation; convolutional neural network (CNN); image recognition; intelligent detection

(责任编辑 王德平)