

基于融合注意力机制和图像识别的路桥裂缝检测技术*

Road and Bridge Crack Detection Technology based on Fusion Attention Mechanism and Image Recognition

席恩伟

(1. 云南云岭高速公路工程咨询有限公司, 云南 昆明 650200;

2. 云南省数字交通重点实验室, 云南 昆明 650200)

摘要: 针对当前路桥检测领域面临的训练效率低下与检测性能不足的挑战, 设计了基于 Inception-Resnet-v2 的路桥裂缝检测模型。该模型结合其强大的特征学习能力与多尺度特征融合, 显著提升了对复杂路桥环境的裂缝检测精度。同时, GKA 聚类算法的应用有效减少了非必要区域的计算, 提高了检测效率。结果表明: 与 AlexNet 相比, 所提模型不仅在帧率 (FPS) 上实现了 8.67% 的提升, 确保了实时处理的潜力, 同时在准确率、精度、召回率及 F1 分数上分别取得了 3.19%、3.75%、1.34% 和 2.66% 的显著提升。该模型为提升路桥检测技术的智能化水平提供了有力支持, 并为该领域未来的研究与发展提供了参考与借鉴。

关键词: 路桥检测; 深度学习; 多尺度; 注意力机制; 聚类算法

中图分类号: TP393 文献标志码: A 文章编号: 1005-8249 (2024) 05-0155-07

DOI:10.19860/j.cnki.issn1005-8249.2024.05.028

XI Enwei

(1. Yunnan Yunling Expressway Engineering Consulting Co., Ltd., Kunming 650200, China;

2. Yunnan Key Laboratory of Digital Transportation, Kunming 650200, China)

Abstract: In response to the challenges of low training efficiency and insufficient detection performance in the current field of road and bridge detection, this paper designs a road and bridge crack detection model based on Inception Resnet-v2. This model combines its powerful feature learning ability with multi-scale feature fusion, significantly improving the accuracy of crack detection in complex road and bridge environments. Meanwhile, the application of GKA clustering algorithm effectively reduces the computation of unnecessary regions and improves detection efficiency. The experimental results show that compared with AlexNet, the proposed model not only achieves an 8.67% improvement in frame rate (FPS), ensuring the potential for real-time processing, but also achieves significant improvements in accuracy, precision, recall, and F1 score of 3.19%, 3.75%, 1.34%, and 2.66%, respectively. This model provides strong support for improving the intelligence level of road and bridge detection technology, and provides valuable reference and inspiration for future research and development in this field

Keywords: road and bridge inspection; deep learning; multi scale; attention mechanism; clustering algorithm

* 基金项目: 云南省数字交通重点实验室 (202205AG070008); 云南交投科技创新计划项目 (YCIC-YF-2021-11) 资助。

作者简介: 席恩伟 (1985—), 男, 硕士, 高级工程师, 研究方向: 数字交通, 智慧养护、工程结构健康监测。

收稿日期: 2023-08-26

研究与应用

0 引言

随着社会经济的快速发展,我国路桥产业实现了跨越式的进步,构建起庞大的基础设施网络^[1-2]。现代桥梁结构大多由混凝土制成,然而,随着使用年限的增加,混凝土表面不可避免地会出现形态各异、程度不一的裂缝,这些裂缝不仅是桥梁结构健康状态的直观反映,更潜藏着对桥梁整体安全性的重大威胁^[3]。由于传感器、计算机、通信、物联网等技术^[4]的发展,路桥相关数据得以高效采集与处理,为预防潜在灾难性事故的发生奠定数据支撑。因此,裂缝检测在桥梁健康监测和可靠性维护中发挥着重要作用。

传统的人工检查方法费时费力,无法广泛推广。近年来,机器学习和计算机视觉^[5]不断发展,可为路桥中裂纹检测带来便利。陈嘉锐等^[6]提出一种基于 YOLO v5s 模型的新型道路裂缝检测系统,可有效实现路面裂缝的分类与检测;隆涛等^[7]提出了一种基于注意力机制和可变形卷积的路面裂缝检测网络,使用密集连接机制强化特征信息,可高质量提取路面裂缝特征;景峰等^[8]提出一种改进的路面裂缝实时检测深度学习模型,在特征融合网络插入融合注意力模块,有效解决部分裂缝漏检问题。在实际道路环境中,桥梁裂缝检测面临诸多复杂挑战,这些挑战源自裂缝本身的多样性(包括形状、大小及分布的差异)以及外部环境因素的不可预测性(如光照不均、石油污染、恶劣天气等)。这些不利条件不仅增加了裂缝检测的难度,还可能影响检测结果的

准确性。虽然深度学习技术的广泛应用虽然为裂缝检测提供了新的解决路径,但其网络深度的增加和规模的扩大也导致计算成本上升和训练时间延长。这在一定程度上限制了深度学习模型在实时性要求较高的裂缝检测工作中的应用。

鉴于此,为了平衡模型性能与训练效率之间的关系,设计了一种基于融合注意力机制和图像识别的路桥裂缝检测模型。该模型由特征提取骨干网络、多尺度上下文信息融合模块、GKA 聚类算法模块和 Dropout 模块组成,综合了不同模块优点,可实现高效训练,准确性较高,可为路桥维护提供一定的借鉴作用。

1 模型建立

基于融合注意力机制和图像识别的路桥裂缝卷积神经网络模型结构图如图 1 所示。该网络由四个模块组成,包括基于 Inception-Resnet-v2 的特征提取骨干网络、多尺度上下文信息融合模块、GKA 聚类算法模块和 Dropout 模块。在图 1 所示的结构中,Inception-Resnet-v2 骨干网络用来提取路桥图像的裂缝特征,使用改进的 Inception 结构增加网络宽度,并引入残差网络防止梯度发散。多尺度上下文信息融合模块融合了 Stem 模块、Inception-Resnet-A 模块、Incept-Resnet-B 模块和 Inception-Resnet-C 模块,可有效提升小裂纹特征提取的敏感度。GKA 聚类算法模块可以准确识别目标区域,降低计算复杂度,提高网络检测率。Dropout 模块可有效地缓解过拟合问题,并在一定程度上达到了正则化效果。

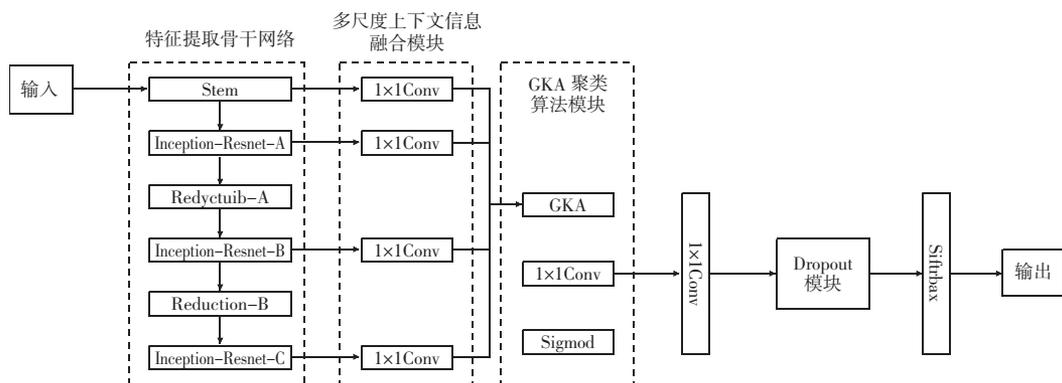


图 1 基于融合注意力机制和图像识别的路桥裂缝卷积神经网络模型结构图

1.1 特征提取骨干网络

Inception 网络结构包含不同大小的多个卷积核，可以增强网络的适应能力，从而提取更丰富的不同尺度特征。同时，通过采用 NIN 模型，Inception 网络结构可以大大减少模型的参数，使网络在不丢失模型特征表示的情况下尽可能减少卷积核的数量，从而降低模型的复杂度。

Resnet 残差网络结构如图 2 所示。Resnet 由一个骨干网络和两个分支网络组成。骨干网络用于识别不同分辨率的物体照片，两个分支网络训练高分辨率图像和目标图像，将其转换为特定分辨率的耦合图像。在 Resnet 中，输入 x_i 的特征图的数量可能与分支网络中特征图的数量不同，因此有必要使用 1×1 卷积来升级或降维输入。此时，残差运算表示为：

$$\begin{cases} F(x_i) = wx_i + \alpha \\ y_i = R(F) + h(x_i) \\ x_{i+1} = R(y_i) \end{cases} \quad (1)$$

式中： x_i 为输入； w 为权重； α 为偏置； y_i 为两个分支的和； $R(\cdot)$ 为激活函数； $F(\cdot)$ 为卷积运算； $h(\cdot)$ 为输入的变换函数； x_{i+1} 为残差模块的最终输出。一般情况下，Resnet 中激活函数为 Relu 函数，其形式为：

$$R(x) = \max(0, x) \quad (2)$$

式中：当 $x > 0$ 时， $R(x) = x$ ，其导数为 1；当 $x \leq 0$ 时， $R(x) = 0$ ，其导数为 0。在正向计算中，可以输入值 x 和阈值 0，并获得输出值；在反向计算中，梯度为 1 或 0。也就是说，梯度下降很小，或者没有发生变化。Relu 函数有利于梯度的扩展和防止梯度的发散，从而防止梯度在多层卷积之后变得非常衰减。与 Tanh 和 Sigmoid 等激活函数相比，Relu 函数计算简单，梯度下降较小，有利于深化网络。

引入残差网络学习单元的目的是避免在训练 Inception 网络模型时梯度完全消失的问题；同时，当网络模型的性能达到一定饱和时，可以对残差网络层进行相同的映射，使得训练网络更快、更容易收

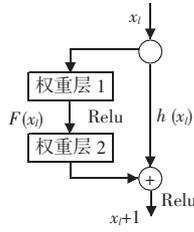


图 2 Resnet 残差网络结构
Fig. 2 Resnet residual network structure

敛。令 X_i 表示第 i 个残差单元的输入， X_n 表示第 n 个残差单元输入， $F(\cdot)$ 为残差函数，残差网络学习单元从第 i 层到第 n 层学习特性公式为：

$$\frac{\partial X_n}{\partial X_i} = \frac{\partial X_i + F(X_i, \omega_i, \alpha_i)}{\partial X_i} = 1 + \frac{\partial F(X_n, \omega_n, \alpha_n)}{\partial X_n} \quad (3)$$

因此无论网络层有多深，梯度永远不会接近 0。此外，传统 Inception 模块对提高网络性能的作用有限，且参数和计算的数量庞大，计算非常复杂。ResNet 模块加深了网络并提高了网络的分类精度，但也导致参数和计算的数量增加。同时，ResNet 中宽度和深度之间的不平衡导致特征提取的多样性比 Inception 模块更差。如果残差模块过于复杂，则跳跃连接带来的训练加速度弱于参数和计算次数急剧增加带来的训练减速，从而导致训练中断或梯度爆炸。

在某种程度上，Inception 模块和残差模块可以相互利用，提高检测精度，减少计算次数。

1.2 多尺度特征融合

深度卷积神经网络中浅层和深层提取的特征是不同的。浅层提取主要为中间特征，如边缘和纹理；而深层提取为超出人类直觉理解的高级语义特征。前者有利于获得目标位置，后者有利于目标检测。对于路桥裂缝图像中有许多小规模目标，其占整个图像的十分之几甚至百分之几。定位信息的缺乏导致小规模目标的定位出现较大偏差，将影响检测的整体性能。因此，一个好的深度卷积神经网络不仅要具备将目标与周围环境区分、消除干扰，还要能够实现良好的定位，确保检测精度。

随着网络的深入，不可避免地会丢失大量有效的定位信息，这降低了小规模目标检测结果的质量。因此，提出一种多尺度特征融合结构，使用具有三种不同采样率的 1×1 卷积核来获得多尺度特征信息。当采样率接近映射特征时， 3×3 卷积核无法有效地捕捉局部细节。因此，采用 1×1 卷积可提取路桥表面较小裂纹边缘的细节。

在 Inception-Resnet-v2 结构中，使用 Stem 模块、Inception-Resnet-A 模块、Incept-Resnet-B 模块和 Inception-Resnet-C 模块卷积的特征组合进行特征融合。多尺度特征融合结构如图 3 所示。

研究与应用

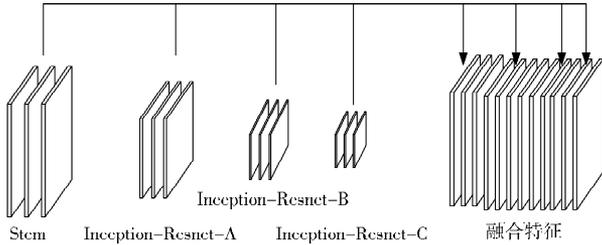


图3 多尺度特征融合结构图

Fig. 3 Multi scale feature fusion structure diagram

1.3 基于 GAK 的聚类算法

随着网络深度的加深, 参数数量变得庞大, 使得网络计算能力下降。为了解决这一问题, 提出混合遗传算数和 K-means 算法 (Genetic Arithmetic and K-means, GAK) 用来加快识别目标区域和降低计算复杂度。

K-means 聚类算法是基于划分的经典聚类算法之一。该算法以欧氏距离为相关测度, 找到相应的聚类中心向量进行最优分类, 从而使评价指标最小化。在 K-means 聚类算法中, 两个数据点之间的距离越小, 两个点之间的相关性就越大。K-means 算法具有鲁棒性强、计算简单、易于理解、易于实现等优点, 得到了广泛的推广和应用。然而, 其对聚类中心的选择要求很高, 并且容易收敛到局部最优解, 从而错过全局最优解。此外, K-means 聚类算法使用欧几里得距离作为数据点之间相关性的标准, 这可能会在处理数据点时造成距离失真, 极大地影响聚类结果。针对这一问题, 提出了改进了基于遗传算法的 K-means 聚类算法。

K-means 算法具有很强的局部搜索能力, 而遗传算法具有较强的全局搜索能力, 因此, 每一代进行遗传运算后, 引入 K-means 算法的运算步骤, 对新生群体中的每个个体进行优化, 优化后的个体进入下一代遗传运算。具体改进步骤如下:

(1) 编码 采用基于聚类中心的浮点编码方法进行编码。假设聚类中心为 m 维, 则 k 个聚类的每条染色体的长度为 $k \times m$ 。令染色体为 $\{x_1, x_2, \dots, x_k\}$, x_i 为 $(x_{i,1}, x_{i,2}, \dots, x_{i,m})$, 其中相应的染色体表示 k 簇中心的坐标。

(2) 初始化 从样本空间中选择 k 个个体, 每个个体代表一个初始聚类中心。聚类中心根据基本编码公式编码为染色体。染色体初始化重复 p_{size} 次以

产生初始群体, 之前 p_{size} 为种群的大小。

(3) 适应度函数 确定聚类中心后, 每个种群的适应度值 f 计算如下:

$$f = \frac{f_w(k)}{f_b(k)} \quad (4)$$

式中: $f_w(k)$ 为种群 k 的聚类中心距离; $f_b(k)$ 除 k 外其他种群的聚类中心距离。

(4) 选择操作 采用轮盘赌和最优保险策略相结合的混合选择算子实现基于适应度 $f(x_i)$ ($i=1, 2, \dots, p_{size}$) 计算个体选择的概率 $P(x_i)$:

$$P(x_i) = \frac{f(x_i)}{\sum_{i=1}^{p_{size}} f(x_i)} \quad (5)$$

根据计算出的选择概率, 采用轮盘赌选择个体参与交叉和突变操作, 生成新的种群。计算新群体中每个染色体的适应度值, 并使用上一代记录的适应度最高的个体来替换目前适应度最低的个体。

(5) 交叉操作 在两个个体 x_1 和 x_2 之间执行交叉操作, 新的个体计算如下:

$$x'_1 = \beta x_1 + (1 - \beta)x_2 \quad (6)$$

$$x'_2 = \beta x_2 + (1 - \beta)x_1 \quad (7)$$

式中: β 为常数。根据个体选择的概率 $P(x_i)$, 选择交叉点位置 j , 并通过交叉获得下一代个体 x'_1 和 x'_2 。

(6) 突变操作 指定各个编码串中的突变点, 并确定每个基因点的值范围 $[U_{min}, U_{max}]$ 。对于每个变异点, 从相应的基因值范围中取一个具有变异概率 $P(m)$ 的随机数来代替原始值。因此, 新的个体计算如下:

$$x_i = U_{min} + \theta(U_{max} - U_{min}) \quad (8)$$

式中: θ 为 0 到 1 之间的一个随机数。

当相同的最优个体适应度值连续出现并超过一定阈值时, 算法停止运行。算法结束, 并输出最终的聚类结果。基于 GAK 的聚类算法流程示意图如图 4 所示。

2 试验结果

2.1 数据集和试验设置

为验证所提基于融合注意力机制和图像识别的路桥裂缝卷积神经网络模型性能, 使用了两个路桥裂缝数据集作为输入样本: CCIC 数据集和 SDNET 数据集。CCIC 数据集中的图像是从各种混凝土建筑中

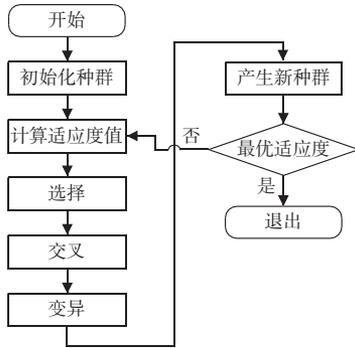


图4 基于 GAK 的聚类算法流程示意图
Fig.4 Schematic diagram of clustering algorithm based on GAK

收集的, 包括 40000 张分辨率为 227×227 像素的 RGB 图像, 分为负 (非裂纹) 和正 (裂纹) 两类。SDNET 数据集包含 56000 多张混凝土桥梁表面、墙壁和人行道的裂缝和非裂缝图像, 数据集包括 0.06 ~ 25 mm 宽的裂缝。该数据集还包括具有各种障碍物的图像, 包括阴影、表面粗糙度、缩放、边缘、孔洞和刻痕。此外, 为验证模型的泛化能力, 随机选择来自 CCIC 的 14000 张图像和 SDNET 的 18000 张图像, 组合成一个更大的路桥裂缝检测混合数据集。接下来, 将数据集按 7:3 分为训练集和测试集。数据集中部分样例如图 5 所示。

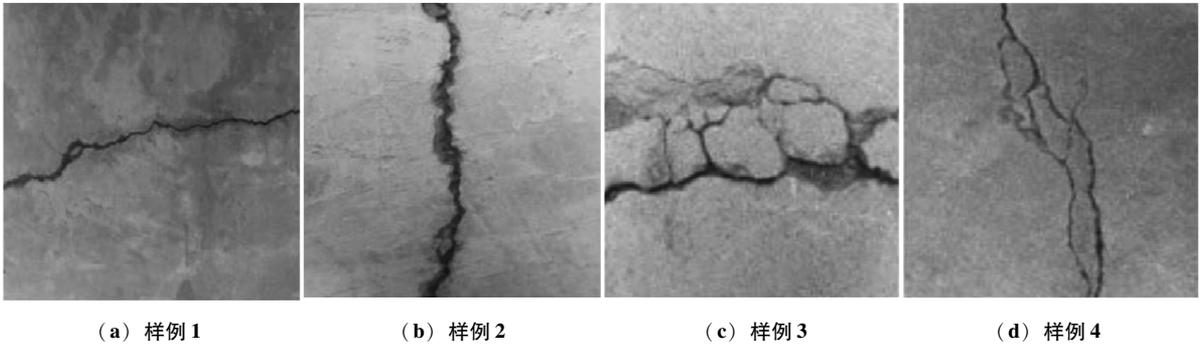


图5 数据集中部分样例
Fig.5 Partial examples in the dataset

试验时软件环境设置如下: python 3.8、cuda 11.6、Pytorch8.1 在 pycharm 软件进行了编译。硬件环境为 Ubuntu 18.04 操作系统, Intel R Core i7V6CPU @ 3.7GHz、64GB 内存和一个 GeForce RTX 3060 GPU 的浪潮服务器。

试验中将所提模型分别对比了 Xception、AlexNet、DenseNet、Inception-v4 等主流检测模型。同时, 选取准确性、召回率、精度、F 分数和 FPS 五个评估指标验证不同模型性能。准确度是指正确分类的图像在所有测试图像中所占的比例, 该指标可以反映所有测试图像的学习情况。准确性计算方法如下:

$$A_{cc} = \frac{T_p + T_n}{A_{ll}} \quad (9)$$

式中: A_{cc} 为准确率指标; T_p 为正确检测正样本分类的数量; T_n 为正确检测负样本分类的数量; A_{ll} 为所有样本的数量。召回率反映样本中有多少正样本被正确检测。召回率计算方法如下:

$$R_{ecall} = \frac{T_p}{T_p + F_n} \quad (10)$$

式中: R_{ecall} 为召回率; F_n 为错误检测为负样本的数量。精度表示预测为正样本中有多少是真正样本。精度计算方法如下:

$$P_{re} = \frac{T_p}{T_p + F_p} \quad (11)$$

式中: P_{re} 为精度指标; F_p 为二类错误, 即未检测为正样本, 实际为正样本的数量。

召回率和精度分别反映了正样本和负样本的学习情况。当网络对负样本的学习效果较差时, 大多数样本将被归类为正样本。在这种情况下, 召回率会很高, 但精度会很低, 反之亦然。为此, 可采用 F 分数指标更全面地反映网络的学习情况。F 分数计算方法如下:

$$F = \frac{2R_{ecall} \times P_{re}}{R_{ecall} + P_{re}} \quad (12)$$

FPS (一秒钟内检测到的帧数) 指标反映模型的检测速度, 即每个图像完成裂缝检测过程所需的时间。

网络训练时采用 SGD 优化器训练模型, 实验时

研究与应用

部分参数定义如下：批量大小设置为 32，最大迭代次数设置为 1000。

2.2 结果与分析

2.2.1 损失对比结果

图 6 所示为初始学习率对模型训练损失的影响对比结果，可以看出不同学习率选择对模型训练稳定性的关键作用。具体而言，当采用 0.05 作为初始学习率时，模型训练过程中的损失曲线呈现出显著的波动性，这直接反映了过高学习率导致的梯度下降不稳定现象，进而对模型的训练效果产生了不利影响。相反，当学习率设定为 0.0001 时，尽管在训练初期，损失下降的速度相较于 0.01 和 0.001 的学习率设置略显迟缓，但随着迭代次数的逐步累积，训练损失展现出了一种更为平稳且持续的下降趋势。最终，该学习率下模型达到了约 0.0432 的训练损失。

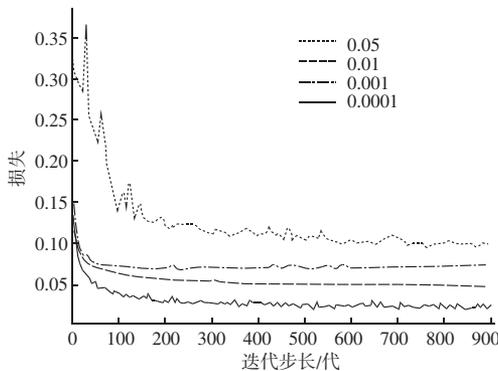


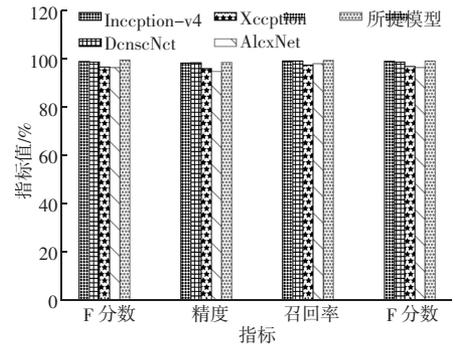
图 6 初始学习率对模型训练损失的影响对比结果
Fig. 6 Comparison results of the impact of initial learning rate on model training loss

综上所述，合理调整学习率是确保深度学习模型有效训练、提升模型性能的关键因素之一。为此，本研究取学习率为 0.0001。

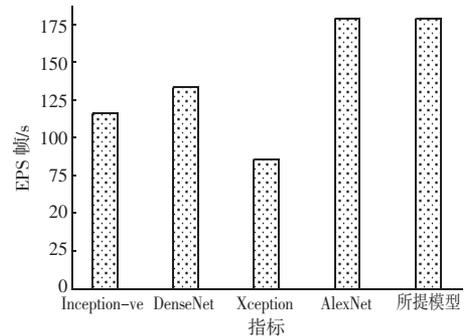
2.2.2 综合性能对比

不同模型综合性能对比结果如图 7 所示。在统一的训练框架下，Xception 与 AlexNet 模型在裂纹检测精度上表现相对逊色。深入剖析其原因，AlexNet 因其网络架构设计较为基础，特征提取层次有限，对复杂特征的捕捉能力较弱，加之网络深度不足，尽管在计算效率上占据优势，但难以适应高精度裂纹检测的需求。另一方面，Xception 模型尽管网络深度接近 DenseNet 201，其在裂纹检测上的不佳表现揭示了在卷积通道中分离通道相关性与空间相关性的策略，对于裂纹这一特定特征的提取可能并不高效，

进而影响了检测效果。



(a) 准确率、精度、召回率、F 分数对比结果



(b) FPS 对比结果

图 7 不同模型综合性能对比结果

Fig. 7 Comparison results of comprehensive performance of different models

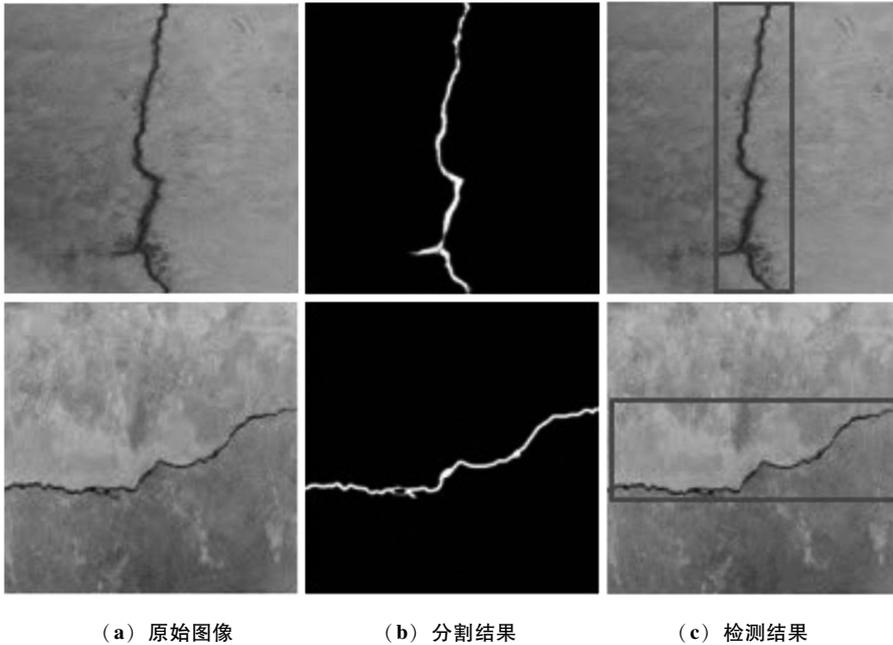
综合各项指标来看，所提模型在综合性能上展现出明显优势。与 AlexNet 相比，该方法不仅在帧率 (FPS) 上实现了 8.67% 的提升，确保了实时处理，同时在准确率、精度、召回率及 F1 分数上分别提升 3.19%、3.75%、1.34% 和 2.66%。所提模型具有提升检测效率与准确性方面的双重优势。

图 8 所示为复杂背景条件下部分裂纹检测结果。实验结果表明，该方法能够准确无误地识别并定位裂纹，展现出强大的鲁棒性和适应性。

3 结论

深入探讨了路桥裂缝检测与识别的关键问题，并据此构建了一种基于注意力机制与图像识别技术的检测模型，得到以下结论：

- (1) 采用注意力机制，可增强模型对关键信息的捕捉能力。
- (2) 基于 Inception-Resnet-v2 的特征提取骨干网络，为模型提供了强大的特征学习能力。
- (3) 提出多尺度上下文信息融合模块，通过整



(a) 原始图像

(b) 分割结果

(c) 检测结果

图 8 复杂背景条件下部分裂纹检测结果

Fig. 8 Partial crack detection results under complex background conditions

合不同尺度的特征信息，提升了模型对复杂裂缝形态的识别精度。

(4) 采用 GKA 聚类算法与 Dropout 技术的结合应用，有效优化了目标区域的识别效率并降低了过拟合风险。这些模块的协同工作，使得所提模型能够全面学习并理解路桥裂缝在不同尺度下的关键特征，从而实现了裂缝检测的准确性和可靠性的显著提升。

所提模型不仅丰富了路桥裂缝检测领域的技术手段，也为路桥的安全管理、使用状态监控以及维护决策提供了有力的技术支撑和参考借鉴。

参 考 文 献

[1] 李永杰. 路桥工程试验检测的现状与提高检测质量的措施 [J]. 工程建设与设计, 2023 (8): 210-212.

- [2] 陈惠水. 路桥过渡段沉降差控制方法探讨 [J]. 工程建设与设计, 2022 (20): 218-220.
- [3] 徐观亚. 基于 Faster R-CNN 与形态法的路桥表观病害识别 [D]. 长沙: 湖南大学, 2020.
- [4] 刘海莺. 基于物联网的皮革企业污水处理监测系统设计与实现 [J]. 中国皮革, 2023, 52 (3): 110-113.
- [5] PANERU B, BIST R, YANG X, et al. Tracking dustbathing behavior of cage-free laying hens with machine vision technologies [J]. Poultry Science, 2024, 103 (12): 104289-104289.
- [6] 陈嘉锐, 崔得龙, 邱泽环, 等. 基于 YOLOv5s 模型的新型道路裂缝检测系统 [J]. 现代电子技术, 2023, 46 (13): 62-66.
- [7] 隆涛, 董安国, 刘来君. 基于注意力机制和可变形卷积的路面裂缝检测 [J]. 计算机科学, 2023, 50 (增刊1): 402-407.
- [8] 景峰, 刘晓捷, 刘军, 等. 基于融合注意力和任务解耦的路面裂缝检测 [J]. 计算机工程与设计, 2023, 44 (5): 1565-1571.