

DOI: 10.20040/j.cnki.1000-7709.2023.20221615

# 基于深度学习的水工混凝土结构表面缺陷检测

曹国金<sup>1</sup>, 苏超<sup>2</sup>, 王文君<sup>2</sup>

(1. 广州市流溪河灌区管理中心, 广东 广州 510920; 2. 河海大学水利水电学院, 江苏 南京 210024)

**摘要:** 早期和及时的缺陷检测对于保证水工混凝土结构的安全运行至关重要。基于深度学习的计算机视觉方法无需复杂的手工特征工程, 可以在远程图像中自动判定结构缺陷类别, 克服了传统人工视觉劳动强度大、主观性强且易出错的缺点。受此启发, 提出了一种基于深度学习的缺陷检测方法, 在 ResNeXt50 网络中引入注意力机制, 以自适应地重新校准通道级特征响应, 使模型更加关注图像中的缺陷信息, 增强特征提取能力。测试结果表明, 所提方法可以实现 88.0% 的 F1 分数, 对于常见混凝土缺陷可以实现较好的分类效果。

**关键词:** 水工混凝土结构; 缺陷检测; 深度学习; 卷积神经网络; 注意力机制

**中图分类号:** TV743

**文献标志码:** A

**文章编号:** 1000-7709(2023)06-0137-05

## 1 引言

受环境和人为因素的影响, 水工混凝土结构随着服役时间的增长会不断积累材料退化、长期疲劳和表面腐蚀等缺陷。这可能会导致灾难性的经济损失, 甚至生命损失<sup>[1]</sup>, 因此结构缺陷检测在结构健康监测领域越来越受到重视。传统的缺陷检测以人工视觉检查为主, 对专业人员具有很强的依赖性, 并且效率低, 检测结果具有很强的主观性<sup>[2]</sup>。因此, 迫切需要开发出可靠、高效的结构缺陷检测方法以促进水利基础设施中表面缺陷的视觉检查。基于计算机视觉的自动化缺陷检测可以解决以人为导向的视觉检查的局限性, 并已引起众多关注, 已提出很多基于计算机视觉的缺陷检测算法。当前计算机视觉方法可分为传统的手动特征提取和端到端的深度学习。传统视觉方法(如边缘检测、阈值分割、小波变换)的缺点在于每张图像的特征不尽相同, 不同类别的对象需要用不同种类的特征来描述, 而且手工设计的特征对于多样性的变化没有很好的鲁棒性。深度学习算法堆叠在一个复杂、抽象的层次结构中, 模仿人类获取某些类型知识的方式, 并且每个抽象层次都是用从层次结构的前一层获得的知识创建的。这一特点使其避免了复杂的特征工程设计, 并为深度学习在计算机视觉中的应用提供了巨大潜

力<sup>[3,4]</sup>。鉴此, 本文提出了一种基于深度学习的混凝土表面缺陷分类模型, 通过向模型中输入原始 RGB 图像, 提取特征信息帮助做出最终的预测, 还引入了注意力机制来突出有益特征信息, 以进一步提高模型的性能。

## 2 卷积神经网络原理

### 2.1 卷积层

卷积层是神经网络的基础部分, 主要用来提取特征。深度为 1 的卷积计算公式为:

$$a_{i,j} = f\left(\sum_{m=0}^2 \sum_{n=0}^2 \omega_{m,n} x_{i+m,j+n} + \omega_b\right) \quad (1)$$

式中,  $a_{i,j}$  为特征图第  $i$  行第  $j$  列元素;  $f$  为激活函数;  $m, n$  分别为行、列序列;  $x_{i,j}$  为第  $i$  行第  $j$  列元素;  $\omega_{m,n}$  为第  $m$  行第  $n$  列权重;  $x_{i+m,j+n}$  为第  $i+m$  行第  $j+n$  列元素;  $\omega_b$  为偏置项。

深度大于 1 时卷积计算公式为:

$$a_{d,i,j} = f\left(\sum_{d=0}^{D-1} \sum_{m=0}^{F-1} \sum_{n=0}^{F-1} \omega_{d,m,n} x_{d,i+m,j+n} + \omega_b\right) \quad (2)$$

式中,  $D$  为深度;  $F$  为滤波器的大小(宽度或高度, 两者相同);  $\omega_{d,m,n}$  为第  $d$  层第  $m$  行第  $n$  列权重;  $x_{d,i,j}$  为第  $d$  层第  $i$  行第  $j$  列元素。

### 2.2 池化层

池化层也称下采样层, 参数矩阵通过池化操作可有效缩小矩阵尺寸、减少参数数量、提升模型

收稿日期: 2022-08-05, 修回日期: 2022-09-06

作者简介: 曹国金(1973-), 男, 高级工程师, 研究方向为水利工程技术管理, E-mail: 1819860784@qq.com

通讯作者: 苏超(1960-), 男, 博士、教授、博导, 研究方向为水工结构缺陷检测与仿真计算, E-mail: csu\_hhu@126.com

运算速度、缩短训练时间。同时池化层对防止过拟合也具有重要作用。池化层的参数不需要通过学习获得,使用时只需指出池化类型、池化操作核的大小、池化操作的步长等超参数即可。

### 2.3 全连接层

全连接层在卷积神经网络中的主要作用是将学习到的“分布式特征表示”进行映射,对图像进行分类。全连接层通常是卷积神经网络的最后几层,经过几个卷积和池化层后,神经网络中的高级推理过程通过全连接层完成。在全连接层中,神经元通过权重矩阵对输入向量进行线性变换。然后通过激活函数实现模型的非线性。因此,全连接层中的每个神经元将随时间捕获输入数据的元素,这些元素随后被用来预测 Softmax 层处的正确值。

### 2.4 激活函数

在神经网络中,神经元节点将上层神经元的输出结果作为该神经元的输入值,并继续被传向下一层,依此类推,直至找到原始问题所需解决的方案。多层神经网络中上下层节点之间的输出、输入值之间关系可用激活函数来表示。激活函数是深度神经网络中不可或缺的关键模块,当前在深度卷积神经网络中常用的激活函数有 Sigmoid 型函数、修正线性单元(ReLU)和指数化线性单元(ELU)。

## 3 混凝土缺陷分类模型

### 3.1 分类模型特征提取网络 Se-ResNeXt50

为实现混凝土缺陷的准确分类,提出了一种基于深度学习的自动化分类模型 Se-ResNeXt50。模型的整体结构见图 1。由图 1 可看出,输入的 RGB 图像首先经过卷积核尺寸为  $7 \times 7$ ,步长为 2 的卷积操作和池化核尺寸为  $3 \times 3$ ,步长为 2 的最大池化操作来快速降低图像分辨率至原分辨率的  $1/4$ ,同时大幅增加特征图的通道数以此进行初步图像特征信息的提取。然后将得到的特征图依次

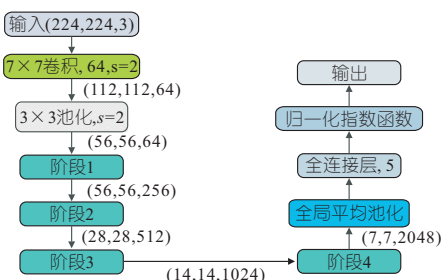


图 1 Se-ResNeXt50 整体结构图

Fig. 1 Overall structure of Se-ResNeXt50

输入到堆叠的 ResNeXt 模块<sup>[5]</sup>和 SE 模块中提取深层特征。ResNeXt 模块采用了分离—变换—聚合的计算策略,在增加准确率的同时基本不改变模型的复杂度,具有较强的扩展性。ResNeXt 模块的结构见图 2。ResNeXt 模块内的路径数定义为基数(图 2 中基数为 32),并且所有的路径均采用相同的卷积参数。输入特征图首先通过  $1 \times 1$  卷积操作来降低通道数,然后经过  $3 \times 3$  卷积操作来提取图像特征,最后再通过  $1 \times 1$  卷积来恢复通道数。与 ResNet 网络相同,ResNeXt 也采用残差连接,通过在网络中增加直连通道使神经网络可以不用学习整个模型的输出,而是学习上一个网络输出的残差值,进而缓解由深度学习模型深度增加带来的退化问题。

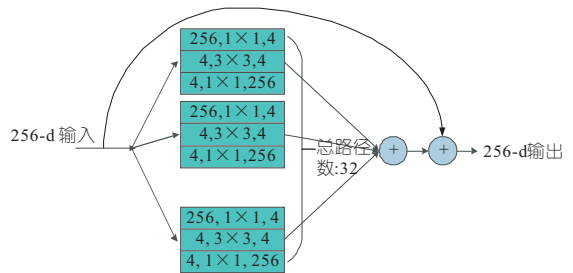


图 2 ResNeXt 模块结构图

Fig. 2 ResNeXt module structure diagram

### 3.2 SE 注意力机制

Se-ResNeXt50 在传统深度学习模型基础上引入了注意力机制来构建不同特征通道之间的相关性,计算图像中像素点之间的相似度来增强或削弱每个预测像素值,自适应地重新加权每个通道,使模型更加关注有助于判断的重要信息。

Se-ResNeXt50 使用的 SE 注意力机制结构见图 3。SE 注意力机制<sup>[6]</sup>模块使网络能够动态调整各通道的权重,重新校准特征来提高网络的表示能力。通过这个操作,使网络分配给有效特征图更高的权重,进而提升模型的性能。如图 3 所示,对于给定的输入特征,首先进行 Squeeze 操作,通过全局平均池化操作顺着空间维度进行特

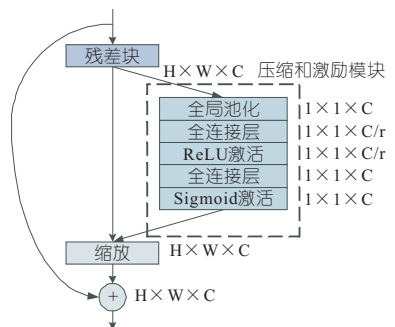


图 3 SE 注意力机制

Fig. 3 SE attention mechanism

征压缩,将每个二维的特征通道变成一个实数,这个实数某种程度上具有全局的感受野,并且输出的维度和输入的特征通道数相匹配。其次是 Excitation 操作,用 2 个全连接操作来实现,第一个全连接用来进行通道压缩,降低参数量,后面紧跟 ReLU 激活函数;第二个全连接用来恢复通道数,并紧跟 Sigmoid 激活函数为每个特征通道生成权重。

### 3.3 全局平均池化和 Softmax 分类

输入图像经过逐层的特征提取,最后通过全局平均池化和全连接层得到一个维度等于类别数的特征向量,然后直接输入到 Softmax 层做出预测。全局平均池化整合了特征图的全局空间信息,使模型对属于图像的空间变换更加鲁棒。此外,由于维度数等于类别数,可以更好地将类别与最后一个卷积层的特征图对应起来。

## 4 数据集与损失函数

### 4.1 数据集

试验中,由 HÜTHWOHL P 等<sup>[7]</sup>提出的混凝土缺陷数据集被用来训练提出的缺陷多分类网络 Se-ResNeXt50。数据集包含裂缝、剥落、风化、一般缺陷和无缺陷 5 个类别。深度学习模型的可靠性主要体现在训练好的模型在新数据集上是否依旧可以实现较好的预测精度。为了对模型的可靠性进行公平测试,试验前将混凝土缺陷数据集按照 7:1:2 的比例随机划分为训练集、验证集和测试集。训练集是为模型权重的学习而提供的数据集,用以学习数据之间的潜在模式和关系。验证集用来了解模型在不同模型类型和超参数选择的性能。测试集用于在比较或选择最终模型时对模型的性能进行无偏估计,仅用于测试已训练模型的泛化能力。表 1 总结了划分后的缺陷数据集的组成。图 4 为每类缺陷的原始样本图像。

表 1 划分后的缺陷数据集组成

Tab. 1 Composition of the divided defect dataset

类别	训练集	验证集	测试集	总计
裂缝	513	117	159	789
风化	202	46	63	311
剥落	278	63	86	427
一般缺陷	172	38	54	264
无缺陷	294	67	91	452
总计	1 459	331	453	2 243



图 4 缺陷数据集样示意图

Fig. 4 Schematic diagram of a sample defect dataset

深度神经网络的训练以大量的输入数据为基础。数据增强指的是通过一系列的随机变化生成与原始图像相似,但不相同的样本加入到训练过程。通过随机改变训练样本可以一定程度上减少神经网络模型对某些属性的依赖,增强模型的鲁棒性和泛化能力。在训练过程中,随机翻转、随机裁剪、随机旋转的增强方法被用来减少模型对于对象位置的依赖,颜色、明亮度改变用于降低模型对颜色的敏感性。

### 4.2 损失函数

监督学习本质是通过给定一系列训练样本  $(x_i, y_i)$ ,模型尝试学习  $x \rightarrow y$  的映射关系,使得给定一个  $x$ ,即便  $x$  不在训练样本中,也能输出  $\hat{y}$ ,尽量与真实的  $y$  接近。损失函数的核心是评估模型对数据集的建模效果。如果模型预测完全错误,损失函数将输出更高的数字。如果预测结果拟合的非常好,则会输出一个较低的数字。在研究中,采用交叉熵损失函数来指导多分类模型的训练。交叉熵损失函数  $L_{\text{Loss}}$  的定义为:

$$L_{\text{Loss}} = - \sum_{i=0}^{C-1} y_i \log p_i \quad (3)$$

式中, $C$  为类别数目; $y = [y_0, \dots, y_{C-1}]$  为样本标签的 one-hot 编码表示,当样本属于该类别时  $y_i = 1$ ,否则  $y_i = 0$ ; $p = [p_0, \dots, p_{C-1}]$  为一个概率分布; $p_i$  为样本属于第  $i$  类的概率。

## 5 训练设置与迁移学习

所有训练和测试试验均在配置有 Windows10 操作系统、Intel Core i7-8700 CPU、16GB 内存和 NVIDIA GTX1080Ti GPU 的电脑端进行。试验研究采用 Pytorch 深度学习框架。GPU 加速的软件环境为 CUDA-10.2、CUDNN-8.1、Pytorch-1.6 和 Python-3.7。Adam 优化器<sup>[8]</sup>被用来调整影响模型训练和输出的网络参数,使其接近最优值。训练的批量大小为 8,权重衰减为  $1 \times 10^{-3}$ ,训练 100 个周期。初始学习率设置为  $1 \times 10^{-3}$ ,在训练过程中应用 poly 学习率策略来动态调整学习率,即初始学习率乘以  $(1 - e_{\text{epoch}} / \max e_{\text{epoch}})^{0.9}$  ( $e_{\text{epoch}}$  为当前训练周期数)。

传统机器学习模型需要从头开始训练,计算量大,需要大量数据才能达到高性能。迁移学习是一种高效的深度学习训练方法,通过重用预先训练的模型作为新任务模型的起点,实现用相对较少的数据训练深度神经网络的目标。为了减少训练数据量对分类模型性能的影响,试验中采用

迁移学习的方法来提高模型收敛的效率。

## 6 结果与分析

### 6.1 评估指标

目前用于衡量分类任务准确性的最流行的分类指标为精确率  $J$ 、召回率  $Z$  和 F1 分数  $F_1$ 。其中准确率为正确预测的正类数占预测的正类总数比例；召回率为正确预测的正类数占实际正类总数比例；F1 分数同时考虑假阳性和假阴性，为精确率与召回率的调和均值。

$$J = T_P / T_P + F_P \quad (4)$$

$$Z = T_P / (T_P + F_N) \quad (5)$$

$$F_1 = 2[JZ / (J + Z)] \quad (6)$$

式中,  $T_P$  为将正类预测为正类的数量;  $F_P$  为将负类预测为正类的数量;  $F_N$  为将正类预测为负类的数量。

### 6.2 试验结果

在相同的混凝土缺陷数据集上将提出的 Se-ResNeXt 分类网络与其他两个卷积神经网络 DenseNet121、MobileNet V3 进行比较, 结果见表 2。由表 2 可看出, 所提的分类网络在测试模型中性能最好, 其平均精确率、召回率和 F1 分数分别为 0.887、0.875、0.880。与排名第二的分类模型 DenseNet121 相比, 所提模型将 F1 分数从 0.840 提高到 0.880。此外, 由表 2 还可看出, 所提模型对于缺陷类别裂缝、风化、剥落和无

表 2 DenseNet121、MobileNet V3 及所提模型测试结果

Tab. 2 Test results of DenseNet121, MobileNet V3 and proposed model

分类网络	数据集类别	精确率	召回率	F1 分数
DenseNet121	裂缝	0.938	0.943	0.940
	风化	0.898	0.841	0.869
	剥落	0.809	0.884	0.845
	一般缺陷	0.780	0.593	0.674
	无缺陷	0.838	0.912	0.873
	平均值	0.853	0.835	0.840
	MobileNet V3	裂缝	0.919	0.862
风化		0.797	0.937	0.861
剥落		0.881	0.860	0.870
一般缺陷		0.707	0.537	0.610
无缺陷		0.800	0.923	0.857
平均值		0.821	0.828	0.818
本文所提模型		裂缝	0.903	0.937
	风化	0.903	0.889	0.896
	剥落	0.886	0.907	0.896
	一般缺陷	0.833	0.741	0.784
	无缺陷	0.911	0.901	0.906
	平均值	0.887	0.875	0.880

缺陷可以达到 85% 以上的 F1 分数。一般缺陷定义为既不属于无缺陷类别又不属于任何特定缺陷类别的缺陷。因此, 很难实现泛化。尽管如此, 所提模型仍达到 78.4% 的 F1 分数, 相较其他两种模型, 对于一般缺陷的检测准确度有明显提高。

### 6.3 注意力模块消融研究

为了验证 SE 注意力模块对混凝土缺陷分类模型性能的影响, 使用相同的数据集分别对有和无注意力模块的分类模型进行训练和测试, 结果见表 3。由表 3 可看出, 由于加入了 SE 模块, 分类网络可以更加有效地利用全局信息选择性地增强对缺陷分类有益的信息的特征, 同时抑制那些无用冗余特征来提高模型性能。与无 SE 模块的分类模型相比, 精确率、召回率和 F1 分数的值分别提高了 3.7%、2.4%、3.1%。

表 3 有无注意力机制模块条件下模型分类性能

Tab. 3 Classification performance with and without attention mechanism module

有无 SE 注意力模块	精确率	召回率	F1 分数
有	0.887	0.875	0.880
无	0.850	0.851	0.849

### 6.4 卷积神经网络的视觉解释

基于深度学习的混凝土缺陷检测模型虽然可以提供准确的分类结果, 但对网络在输入图像中的位置、前向传递中激活神经元系列及网络如何到达其最终输出缺乏合理的解释。为了在实际工程中应用深度学习模型, 必须建立对它们的信任并使其更加透明。试验中使用 Grad-CAM 对 Se-ResNeXt50 模型的预测过程进行视觉解释, 创建模型中最后一个卷积层的热力图。Grad-CAM 实施过程具体为先将图像作为输入并创建一个模型, 该模型在要创建 Grad-CAM 热力图的层被切断, 并附加全连接层进行类别预测。然后运行模型, 获取层输出和损失值。接着找到所需的模型某一层输出的梯度和模型损失值。在该处可以获取有助于预测、缩小、调整大小和重新缩放的梯度部分, 以便热力图与原始图像重叠。图 5 为 Se-ResNeXt50 对混凝土缺陷的热力图可视化结

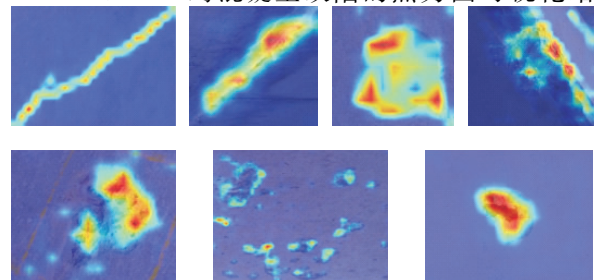


图 5 热力图可视化结果

Fig. 5 Heatmap visualization results

果,热力图中的高亮深色区域表示模型在做出预测时重点关注的区域。由图 5 可看出,Se-ResNeXt50 在判断当前图像所属缺陷类别时重点关注的区域是合理的,预测的依据是根据其内在特征,而不是图像中的随机区域。

## 7 结论

a. 探索了深度学习在混凝土缺陷分类任务中的直接应用。为了验证所提模型的性能,与其他两个卷积神经分类网络进行对比分析。结果显示所提模型在混凝土缺陷分类数据集上优于现有模型,有助于实现更具成本效益和客观的混凝土结构表面缺陷检查。此外,基于 Grad-CAM 的热力图可视化结果表明,所提模型对于缺陷类别的预测是可靠的。

b. 所提模型在缺陷分类任务中效果良好,但在混凝土结构的自动检测中仍存在许多挑战。首先,模型性能受数据集的限制,每个缺陷类别的检测精度还存在差距,另一个挑战是将所提分类网络用作检测和分割任务的主干特征提取网络,以便为检测器和解码器提供更好的特征。

### 参考文献:

[1] 孟庆成,万达,吴浩杰,等. 基于卷积神经网络的混

凝土裂缝图像识别方法[J]. 沈阳建筑大学学报(自然科学版),2021,37(5):832-840.

[2] 郎洪,温添,陆键,等. 基于深度学习的三维路面裂缝类病害检测方法[J]. 东南大学学报(自然科学版),2021,51(1):53-60.

[3] 张珂,冯晓晗,郭玉荣,等. 图像分类的深度卷积神经网络模型综述[J]. 中国图象图形学报,2021,26(10):2305-2325.

[4] 蒋文波,罗秋容,张晓华. 基于数字图像的混凝土道路裂缝检测方法综述[J]. 西华大学学报(自然科学版),2018,37(1):75-84.

[5] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 1492-1500.

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

[7] HÜTHWOHL P, LU R, BRILAKIS I. Multi-classifier for reinforced concrete bridge defects[J]. Automat constr, 2019, 105:102824.

[8] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv: 1412.6980, 2014.

## Surface Defect Detection of Hydraulic Concrete Structures Based on Deep Learning

CAO Guo-jin<sup>1</sup>, SU Chao<sup>2</sup>, WANG Wen-jun<sup>2</sup>

(1. Guangzhou Liuxihe Irrigation District Management Center, Guangzhou 510920, China;

2. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210024, China)

**Abstract:** Early and timely defect detection is essential to ensure the safe operation of hydraulic concrete structures. The deep learning-based computer vision method does not require complex manual feature engineering, and can automatically determine the category of structural defects in remote images, overcoming the shortcomings of traditional manual vision that are labor-intensive, subjective and prone to errors. Inspired by this, this paper proposes a deep learning-based defect detection method, which introduces attention mechanism into the ResNeXt50 network to adaptively recalibrate the channel-level feature responses, so that the model can pay more attention to the defect information in the image and enhance the feature extraction ability. Test results show that the proposed method can achieve an F1 score of 88.0%, and realize a good classification effect for common concrete defects.

**Key words:** hydraulic concrete structure; defect classification; deep learning; convolutional neural network; attention mechanism