

文章编号:1673-9469(2025)05-0010-09

DOI:10.3969/j.issn.1673-9469.2025.05.002

## 基于改进 CNN 和图像处理的钢梁裂缝识别方法

赵丽洁<sup>1,2</sup>, 和子硕<sup>1</sup>

(1. 河北工程大学 土木工程学院, 河北 邯郸 056038; 2. 天津农学院 水利工程学院, 天津 300384)

**摘要:** 裂缝检测是结构健康监测的重要内容之一,为实现复杂背景下钢梁裂缝的定性分析,提出一种基于改进卷积神经网络和数字图像处理的裂缝损伤定位和裂缝图像分割两阶段检测方法。第一阶段通过构建多尺度卷积神经网络来识别复杂背景下的裂缝图像,该网络由多尺度卷积模块 Inception 与残差模块组成,其中,多尺度卷积模块 Inception 包含  $1\times 1$ 、 $3\times 3$ 、 $5\times 5$  三种不同尺寸的卷积核,用于图像的多尺度特征提取,在残差模块中引入卷积层和非线性激活函数以增强跨层融合能力进而提取更深层次的特征。采用 Grad-CAM 可视化分析突出多尺度卷积神经网络的预测依据,证明其分类性能和判别依据。第二阶段,针对识别的裂缝图像,提出图像滤波去噪、阈值分割分离裂缝像素和形态学处理优化分割结果的组合流程对裂缝进行像素级别的分割与提取,以人工标注的像素标定结果作为真实标签评估图像分割的识别结果。在数据集上的训练结果表明:多尺度卷积神经网络对钢梁裂缝图像的识别准确率可达 98.8%,提出的图像处理组合流程最大交并比为 0.819,可较好地对裂缝进行分类和提取。

**关键词:** 卷积神经网络;钢梁裂缝识别;多尺度卷积模块;残差模块;图像处理

**中图分类号:** TP391.4; TU3

**文献标识码:** A

## Crack Identification Method of Steel Girder Based on Improved CNN and Image Processing

ZHAO Lijie<sup>1,2</sup>, HE Zishuo<sup>1</sup>

(1. School of Civil Engineering, Hebei University of Engineering, Handan, Hebei 056038, China;

2. School of Water Conservancy Engineering, Tianjin Agricultural University, Tianjin 300384, China)

**Abstract:** Crack detection is one of the important aspects of structural health monitoring. To achieve qualitative analysis of cracks in steel beams in complex backgrounds, a two-stage detection method based on improved Convolutional Neural Network (CNN) and digital image processing is proposed for crack damage location and crack image segmentation. The first stage uses a multi-scale convolutional neural network to identify crack images in complex backgrounds. This network consists of a multi-scale convolution module Inception and a residual module. The multi-scale convolution module Inception contains three different-sized convolution kernels ( $1\times 1$ ,  $3\times 3$ ,  $5\times 5$ ) for multi-scale feature extraction of the image. In the residual module, convolution layers and nonlinear activation functions are introduced to enhance cross-layer fusion ability and extract deeper features. The Grad-CAM visualization analysis highlights the prediction basis of the multi-scale convolutional neural network, proving its classification performance and discrimination basis. In the second stage, for the identified crack images, a combined process of image filtering denoising, threshold segmentation to separate crack pixels, and morphological processing to optimize the segmentation result is proposed for pixel-level segmentation and extraction of cracks. The pixel marking results manually annotated are used as the true labels to evaluate the recognition effect of image segmentation. The training results on the dataset show that the multi-scale convolu-

收稿日期:2023-12-26

基金项目:国家自然科学基金资助项目(52208193)

第一作者:赵丽洁(1988—),女,河北邯郸人,博士,副教授,从事结构健康监测方面的研究。

tional neural network has an identification accuracy of 98.8% for steel beam crack images. The proposed image processing combination process has a maximum intersection-over-union (IOU) of 0.819, which can better classify and extract cracks.

**Key words:** convolutional neural network; crack recognition of steel girder; multi-scale convolution module; residual module; image processing

近年来,深度学习技术快速发展,为结构损伤识别任务提供了新的解决方案。作为一种自动特征提取方法,深度学习模型通过在大规模损伤数据集上进行训练,能够自动提取裂缝特征,并广泛应用在实际工程场景中。

基于图像的裂缝识别任务主要包括3种:裂缝分类、裂缝目标检测和裂缝分割<sup>[1]</sup>。裂缝分类用于判断图像或图像块是否含有裂缝,Cha等<sup>[2]</sup>采用滑动窗口技术训练卷积神经网络(Convolutional Neural Network, CNN)分类器检测小图像块是否存在裂缝,其准确率为98.22%,优于传统的边缘检测方法。Xu等<sup>[3]</sup>提出一种端到端的CNN裂缝识别算法,在不进行预训练的情况下,对裂缝和背景分类的准确率达到96.4%。Xu等<sup>[4]</sup>结合输入子图像的多级卷积特征,采用改进的融合CNN架构对裂缝进行分类。裂缝目标检测旨在全景图像中找到裂缝的具体位置,并生成包含裂缝的矩形框。Cha等<sup>[5]</sup>提出了一种基于更快速区域卷积神经网络的结构视觉检测方法,用于检测5类损伤,实现了对多种损伤的同时检测。与目标检测相比,裂缝分割实现了像素级的裂缝分类,可得到更详细的裂缝结果。目前主流的裂缝分割是基于编码器-解码器模型的使用与优化,如全卷积神经网络(Fully Convolutional Network, FCN)和U型网络(U-Net)<sup>[6]</sup>。Li等<sup>[7]</sup>提出一种基于特征金字塔网络改进的FCN裂缝自动分割方法,交并比(Intersection over Union, IoU)达到0.78。此外,两阶段的裂缝检测方法也取得了良好效果,Xiong等<sup>[8]</sup>首先对栈桥图像进行边缘检测,将图像划分为大小合适的区域,再利用改进的DeepLabV3+模型对划分的区域进行像素级语义分割,实现裂缝的定位和识别。Han等<sup>[9]</sup>提出了YOLOv3和DeepLabV3+模型,结合无人机影像识别和定位钢结构表面裂缝,利用YOLOv3定位存在裂缝的区域,之后输入DeepLabV3+对裂缝图像进行分割。夏子祺等<sup>[10]</sup>提出了一种结合深度学习与图像处理的建筑外墙裂缝检测方法,以RetinaNet作为裂缝区域定位的检测网络,对定位的单幅图像进行图像处理,进而实现裂缝的提取。Nguyen等<sup>[11]</sup>提出两阶段卷积神经网络,

用于道路裂缝的检测和分割,第一阶段使用CNN定位含裂缝的区域,第二阶段对含裂缝的图像进行分割。

虽然基于深度学习的裂缝识别方法已经取得显著进展,但在分析真实结构图像时,仍然面临诸多挑战。在人工检查时,工程师通常会用线标记裂缝,并附上标定尺进行拍照,现场获取的裂缝图像存在大量的人工干扰。此外,传统的CNN卷积核尺寸单一,在复杂背景下缺乏多尺度特征提取的能力<sup>[12]</sup>,利用单阶段的CNN只能定位含裂缝区域,不能进行提取和分割裂缝;使用裂缝分割网络则需要大量的标注和数据集,现实中很难收集到有标注的裂缝图像。因此,难以从复杂背景中提取裂缝的二值图像。针对以上问题,本文在CNN和裂缝识别研究的基础上,以钢结构裂缝为研究对象,提出一种基于改进CNN和图像处理的两阶段检测方法:第一阶段构建多尺度卷积神经网络,定位裂缝损伤;第二阶段结合数字图像处理技术对裂缝图像进行像素级分割。考虑到现实图像往往具有复杂背景干扰,本文选取具有复杂背景的钢梁裂缝图像构建数据集,用于方法的训练与验证。

## 1 基本理论

### 1.1 模型构建

在某种程度上,卷积神经网络越深,其学习能力越强,增加卷积核(Conv)的数量可以提取图像中更多尺度的特征信息。以VGG16网络模型为例(图1),该模型共有5个卷积模块,13个卷积层,虽然提取特征的能力很强,但是每个卷积核尺寸仅为 $3 \times 3$ ,若想提取更多尺度的特征,只能增加卷积核的数量。然而,卷积核数量的增加很可能导致模型出现过拟合现象。多尺度卷积神经网络的核心思想是将多尺度特征学习能力融入传统的卷积神经网络<sup>[13]</sup>,利用不同尺寸的卷积核提取不同的特征,再进行特征融合。因此,构建如图2所示的多尺度卷积模块,卷积核尺寸分别为 $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$ ,每个卷积模块激活函数均采用修正线性单元(Rectified Linear Unit, ReLU)。在多尺度卷积

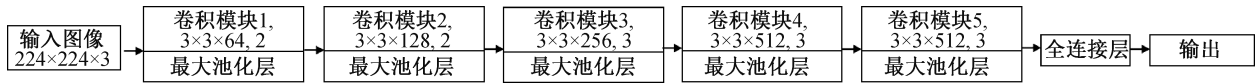


图1 VGG16网络模型

Fig. 1 VGG16 network model

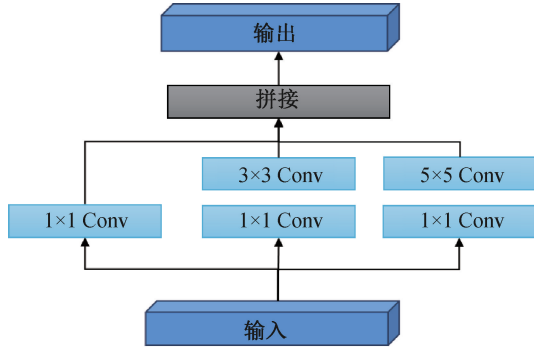


图2 多尺度 Inception 模块

Fig. 2 Multi-scale Inception module

Inception 模块<sup>[14]</sup>中引入  $1 \times 1$  的卷积核可实现特征图的降维,这既有助于实现特征融合,又能降低模型参数量,从而提升网络训练速度。

在残差模块中引入卷积层、批量归一化与非线性激活函数,如图3所示。卷积层用于进一步提取图像的局部特征,有助于在网络深层进行信息融合时获得更丰富的高级语义信息;批量归一化层则用于缓解因卷积层增加而可能出现的梯度消失或梯度爆炸问题<sup>[15]</sup>。改进后的残差模块对输入信息的计算如式(1)所示:

$$H(x) = F(x) + y \quad (1)$$

$$y = \text{ReLU}\{\text{BN}[\text{Conv}(x)]\} \quad (2)$$

式中,  $H(x)$  为残差模块输出;  $x$  为上一层的输入;

$F(x)$  为对  $x$  的特征提取和非线性变换;  $\text{Conv}(\cdot)$  为卷积操作;  $\text{BN}(\cdot)$  为批量归一化操作;  $\text{ReLU}(\cdot)$  为非线性激活函数。

为使残差模块与相应的卷积模块通道数相同,残差模块1、模块2、模块3分别选择128、192、256个卷积核,每个卷积核尺寸均为  $1 \times 1$ ,步长均为1,激活函数采用 ReLU。

本研究基于多尺度 Inception 模块和残差模块构建多尺度卷积神经网络模型(Inception-Residual net, IRnet),IRnet 模型由3个 Inception 模块和3个残差模块构成,如图3所示。首先,将图像输入一个卷积层,之后接最大池化层抑制噪声,降低冗余特征,利用 Inception 模块对图像进行多尺度特征提取;其次,通过残差模块对图像的局部特征进行组合,进而提取更深层次的特征,在 Inception 模块与残差模块输出值相加后引入最大池化层,防止图像边缘和纹理特征丢失;最后,基于 Inception 模块和残差模块的特征融合,使用全局平均池化层对输出特征进行处理,通过 softmax 分类层输出识别的结果。模型整体结构如表1所示。

## 1.2 图像分割

最大类间方差法(OTSU)是一种自适应阈值分割算法<sup>[16]</sup>,在目标和背景灰度差异较小时具有

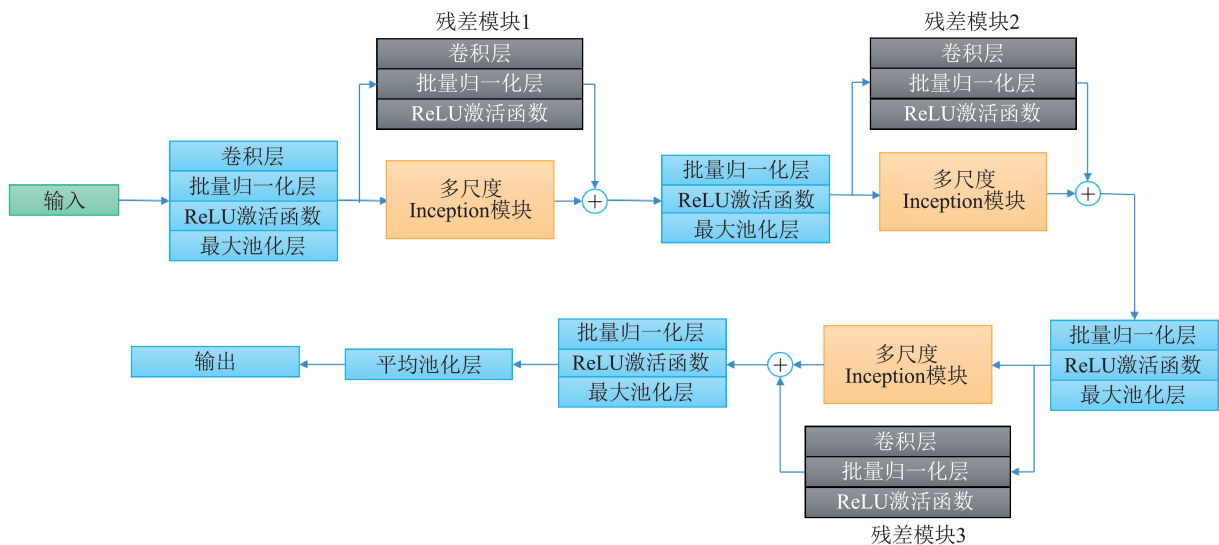


图3 多尺度卷积神经网络

Fig. 3 Multi-scale convolutional neural network

表 1 网络模型具体参数

Tab. 1 Specific parameters of the network model

层数 (Layer)	名称	输入 (Input)	卷积核 (Conv)	步长 (Stride)	输出 (Output)
第一层	Conv	224×224×3	7×7×64	2	112×112×64
最大池化层	Maxpooling	112×112×64	3×3	2	56×56×64
残差模块	Conv1	56×56×64	1×1×128	1	56×56×128
Inception 模块	Conv2	56×56×64	1×1×48	1	56×56×48
	Conv3-1	56×56×64	1×1×16	1	56×56×16
	Conv3-2	56×56×48	5×5×32	1	56×56×32
	Conv4-1	56×56×64	1×1×24	1	56×56×24
	Conv4-2	56×56×24	3×3×48	1	56×56×48
最大池化层	Maxpooling	56×56×128	3×3	2	28×28×128
残差模块	Conv5	28×28×128	1×1×192	1	28×28×192
Inception 模块	Conv6	28×28×128	1×1×48	1	28×28×48
	Conv7-1	28×28×128	1×1×32	1	28×28×32
	Conv7-2	28×28×32	5×5×48	1	28×28×48
	Conv8-1	28×28×128	1×1×64	1	28×28×64
	Conv8-2	28×28×64	3×3×96	1	28×28×96
最大池化层	Maxpooling	28×28×192	3×3	2	14×14×192
残差模块	Conv9	14×14×192	1×1×256	1	14×14×256
Inception 模块	Conv10	14×14×192	1×1×64	1	14×14×64
	Conv11-1	14×14×192	1×1×32	1	14×14×32
	Conv11-2	14×14×32	5×5×64	1	14×14×64
	Conv12-1	14×14×192	1×1×96	1	14×14×96
	Conv12-2	14×14×96	3×3×128	1	14×14×128
最大池化层	Maxpooling	14×14×256	3×3	2	7×7×256
全局平均池化层	Avgpooling	7×7×256	—	—	1×1×256
全连接层	Fully Connected Layer	—	—	—	3

注: Conv1、Conv2、Conv3、Conv4 等代表不同的卷积层数,层数依次加深。

更高的分割效果。按照某一阈值  $T$  可将图像的像素分为前景和背景两类,其方差计算方法见公式(3)(4),图像的内类方差  $\delta^2$ 、类间方差  $\sigma^2$  计算见公式(5)(6)<sup>[17]</sup>,图像类内方差  $\delta^2$  最小或者类间方差  $\sigma^2$  最大对应的阈值  $T$  为最佳阈值。

$$\delta_0^2 = \frac{1}{P_0} \sum_{i=0}^T P_i \times (i - u_0)^2 \quad (3)$$

$$\delta_1^2 = \frac{1}{P_1} \sum_{i=T+1}^{L-1} P_i \times (i - u_1)^2 \quad (4)$$

$$\delta^2 = P_0 \times \delta_0^2 + P_1 \times \delta_1^2 \quad (5)$$

$$\sigma^2 = P_1 \times (u_0 - u)^2 + P_1 \times (u_1 - u)^2 \quad (6)$$

式中,  $\delta_0^2$ 、 $\delta_1^2$  分别为前景、背景像素的方差;  $P_0$ 、 $P_1$  分别为前景和背景像素在图像中的分布概率;  $P_i$  为第  $i$  个像素的分布概率;  $u_0$ 、 $u_1$  分别为前景、背景像素值均值;  $u$  为总体灰度均值;  $L$  为像素的灰度值。

## 2 裂缝损伤定位

### 2.1 数据来源

为了识别复杂背景下的裂缝,本研究选取 APSS2018 Steel Girder Crack ID Dataset<sup>[18]</sup> 作为数据集,该数据集反映了我国大跨度桥梁人工裂缝检测的实际情况,数据集图像如图 4 所示。可以看出,图像包含大量手写体、卷尺等干扰因素。考虑到数据集规模较小,不满足卷积神经网络需要大量数据样本的前提条件,且易造成模型过拟合,因此,采用数据增强技术<sup>[19]</sup> 来提高模型的泛化能力,同时可以降低数据收集成本。

采用滑动窗口策略把原始图像裁剪为 224×224 像素的子图像,将子图像分为背景、手写体(含卷尺)和裂缝 3 类。由于每幅原始图像中裁剪的裂缝子图像数量较少,且很难收集到大量带有标

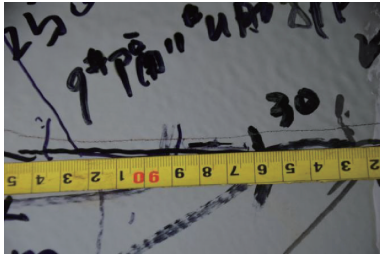


图4 原始图像

Fig. 4 The raw image

签的裂缝子图像,因此,采用水平翻转,垂直翻转,逆时针旋转 $90^\circ$ 、 $180^\circ$ 和 $270^\circ$ 等5种增强算法对裂缝子图像进行扩充,将每幅子图像扩充为5幅图像。经处理后,最终获得7 331幅背景图像数据集、4 650幅手写体图像数据集和2 406幅裂缝图像数据集,扩充后的3类子图像如图5所示。将所有图像按8:2的比例分为训练集和测试集,同时在训练集中按7:3的比例分为训练集和验证集。训练集、验证集、测试集图像数量如表2所示。

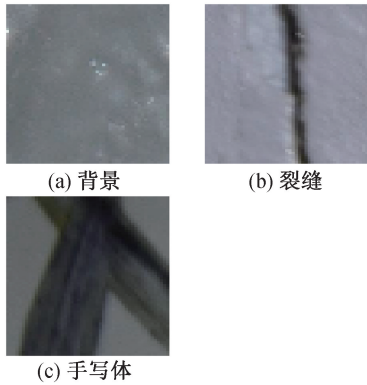


图5 子图像

Fig. 5 The sub image

表2 各图像集的数量 (单位:幅)

Tab. 2 Number of each dataset (unit: piece)

数据集	训练集	验证集	测试集	总和
背景	4 105	1 760	1 466	7 331
裂缝	1 347	577	482	2 406
手写体	2 604	1 116	930	4 650

## 2.2 网络训练

为评估基于多尺度 Inception 模块和残差模块构成的 IRnet 对钢梁裂缝识别的性能,本研究在经数据增强处理的数据集上进行试验,并与以下4种基于 CNN 的模型进行对比:第1种是只含有 Inception 模块的网络(Inet),第2种是只含有残差模块网络(Rnet),第3种是微调的 AlexNet<sup>[20]</sup>,第4种是微调的 VGG16<sup>[21]</sup>。网络训练学习率采用学习率衰减策略,初始学习率为0.001,每经过1轮

训练,学习率减少20%。训练时批尺寸为64,轮数为20,模型优化器选择 Adam 优化。在对改进的 AlexNet 与 VGG16 网络进行训练时,由于参数量较大,将训练批尺寸和初始学习率进行微调。试验的硬件配置为 Intel(R) Core(TM) i7-12700 H CPU,16.0 GB 内存,NVIDIA GeForce RTX 3060,Windows 11 操作系统。

为更客观地评估裂缝的识别效果,引入准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 $F_1$ 分数作为评估指标。4个指标的精度范围均为0~1,越接近1说明效果越好,其计算公式如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

式中,Accuracy 表示正确预测的样本占总样本的比例,%;Precision 表示所有预测为裂缝的样本其真实也为裂缝的比例,%;Recall 表示所有真实裂缝样本的比例,%; $F_1$ 为 Precision 和 Recall 的加权平均值,%;TP 为实际和预测均为裂缝的样本数量,幅;TN 为实际和预测均为非裂缝的样本数量,幅;FN 表示实际为裂缝,预测为非裂缝的样本数量,幅;FP 表示实际为非裂缝,预测为裂缝的样本数量,幅。

## 2.3 训练结果

IRnet 与其他4种网络模型的训练结果如图6所示,从图6(a)可以看出,各个模型的准确率在训练过程中总体呈现稳步上升趋势,在第5轮训练后基本上趋于稳定,准确率达到90%以上;由11~16轮训练过程可以看出,IRnet 的准确率最高,Inet 其次,VGG16 的准确率出现波动,AlexNet 准确率最低,表明 IRnet 在训练集中具有良好的性能。由图6(b)可知,各个模型的损失值在训练过程中总体呈现下降趋势,其中 AlexNet 损失值最高;由11~16轮的训练过程可以看出,IRnet 的损失值低于 Inet 和 Rnet,各模型的损失值最终趋于收敛,IRnet 的损失值最小,约为0.005。

5种模型在测试集上的评估指标如表3所示。结果显示,IRnet、Inet、Rnet 的总体准确率相差不大,但在裂缝的识别准确率上,IRnet 最高,为98.8%,其 $F_1$ 分数相较于 Inet、Rnet 分别提升0.7%和0.8%,并且 IRnet 的4种评估指标均优于传统的 AlexNet、VGG16 模型,表明 IRnet 具有良好的裂缝识别性能。

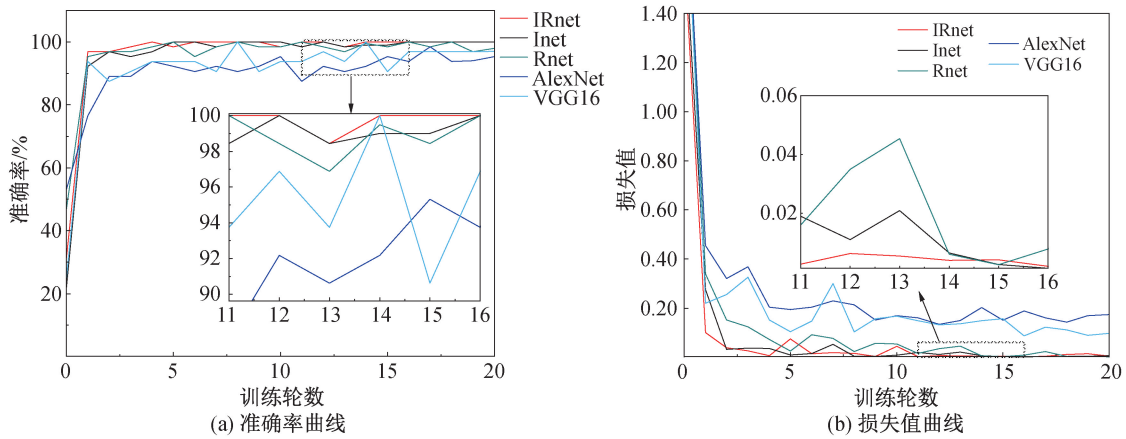


图 6 训练结果曲线

Fig. 6 Curves of training results

表 3 评估指标对比

Tab. 3 Comparison of the evaluation indicators

模型	准确率/%	精确率/%	召回率/%	$F_1$ 分数/%
IRnet	99.2	98.8	98.6	98.7
Inet	98.9	97.7	98.3	98.0
Rnet	98.9	98.8	97.1	97.9
AlexNet	93.7	75.9	87.8	81.4
VGG16	94.0	76.3	82.2	81.4

2.4 可视化分析

本文采用 Grad-CAM<sup>[22]</sup> 捕捉裂缝在 IRnet 中的梯度,生成一个粗略的定位图,突出显示图像中裂缝的关键区域。可视化 IRnet 的卷积模块的部分输出结果如图 7 所示,红色边框内为经过 Grad-CAM 分析的裂缝图像。可以看出,在较浅的卷积模块(Conv、Conv4-2)中,由于感受野较小,只关注局部特征,定位能力有限,提取的特征为裂缝的边缘、纹理等低级信息。而卷积模块 Conv7-2 提取的特征为裂缝的具体形状等信息,Conv12-2 有更大的感受野,所提取的特征更多、更抽象。由此可以得出,随着网络层数的加深,卷积模块提取的特征越来越抽象。

2.5 损伤定位

在得到经过训练的网络模型后,选取未经训练的整幅图像进行裂缝损伤区域定位。对整幅图像同样采用  $224 \times 224$  的滑动窗口进行裁剪,将裁剪后的子图像输入网络模型,若预测结果为裂缝图像则保留,若预测结果为背景或者手写体图像则将其调暗,错误的预测用红色边框来显示。最后将经过 IRnet 预测的子图像进行拼接,得到裂缝损伤定位图。不同网络模型的裂缝损伤定位图如图 8 所示,可以看出,含有多尺度卷积模块的 IRnet 裂缝损伤定位精度更高,错误率更小。但是也有检测错误的图像,如图 8 错例所示,第 1 幅图像中手写体占比较大,而裂缝只出现在图像右下角边缘,网络预测结果为手写体;第 2 幅图像中裂缝出现在上部边缘,网络预测结果为背景。

为了进一步评估 IRnet 在手写体干扰因素下的准确性,选取同时包含裂缝和手写体的图像进行 Grad-CAM 可视化分析。从图 8 可以看出,IRnet 在提取裂缝特征时准确地找到了裂缝的位置,并展示了裂缝的形状和大小,而 VGG16 关注的特征则是手写体,并未定位到裂缝的位置。

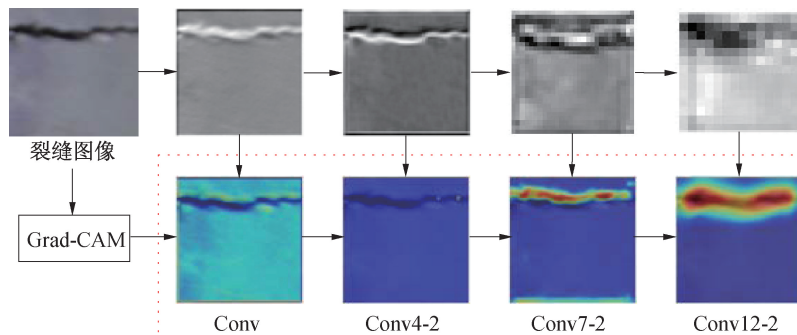


图 7 可视化特征图

Fig. 7 Visualization of feature map

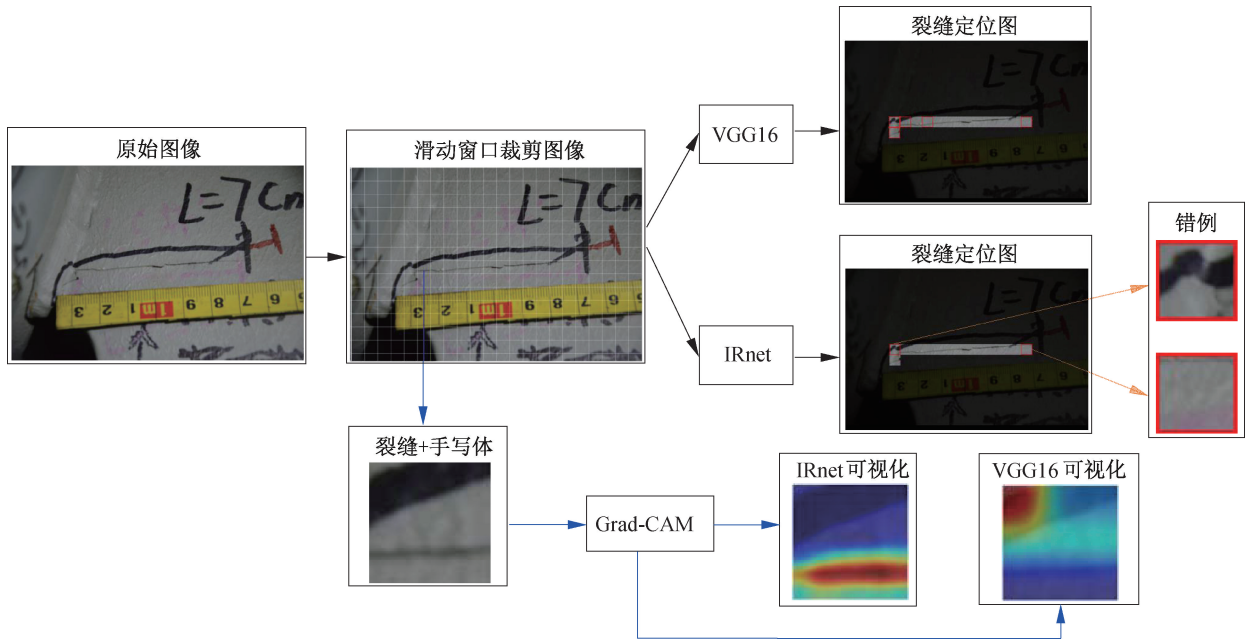


图8 裂缝损伤定位流程及可视化

Fig. 8 Crack damage location process and visualization

### 3 裂缝图像分割

#### 3.1 数据集标定

图像分割的评价指标选用 IoU, 其计算见式 (8), 含义为二值图像、真实标注的交集除以二值图像、真实标注的并集。

$$IoU = (A \cap B) / (A \cup B) \quad (8)$$

式中,  $A, B$  分别为二值图像、真实标注。

针对 IRnet 分类的测试集裂缝图像, 本文选取不同类型的裂缝并利用标注软件 Labelme 进行标注, 标注效果如图 9 所示。

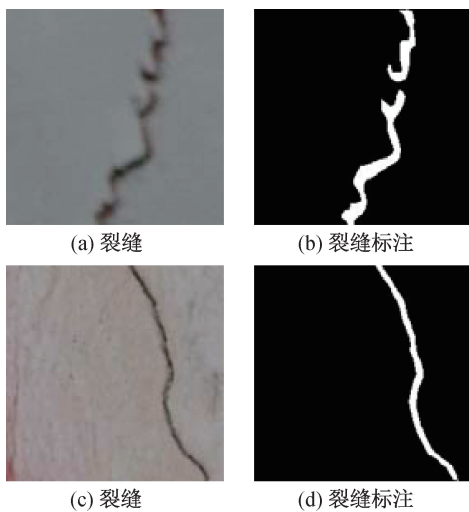


图9 裂缝图像标注示意图

Fig. 9 Schematic diagram of the crack image annotation

#### 3.2 裂缝分割

本次研究选取的测试集裂缝图像存在以下问题: 在二值化图像过程中, 裂缝背景的噪声会导致二值图像出现多余的零星散点; 裂缝呈细长状, 在图像处理过程中极易发生像素点的缺失, 进而导致裂缝图像断裂。针对第 1 个问题, 本文对比了 3 种常用的图像平滑去噪的方法, 即均值滤波、高斯滤波和中值滤波, 根据对比结果, 选用滤波器尺寸为  $5 \times 5$  的中值滤波对裂缝图像进行预处理。此外, 引入基于连通区域的图像特征分析算法, 通过计算连通区域的面积, 删除二值图像中的零星散点。针对第 2 个问题, 使用基于最大类间方差法的 OTSU 阈值分割, 叠加预处理, 采用形态学处理方法强化裂缝特征, 减少裂缝的断裂现象。整体处理流程及识别效果如表 4 所示。

在对子图像进行准确分类后, 采用图像处理的方法提取裂缝掩膜。将所有子图像转换为一个通道的灰度图像, 如果子图像被分类为裂缝, 则按照表 4 的分割流程将子图像的裂缝与背景分离, 否则, 将整幅图像的灰度值设置为 0。将所有子图像的结果进行拼接, 得到整体的裂缝掩膜, 裂缝掩膜的提取结果见图 10。可以看出, 在复杂背景的干扰下, 本文提出的两阶段检测方法优于传统的图像分割方法。

表 4 裂缝各典型图像处理流程及识别效果

Tab. 4 Typical image processing process and identification effect of cracks

原图	灰度图	中值滤波	阈值分割	闭合	小连通区域去噪	叠加原图	IoU
							0.733
							0.735
							0.73
							0.819
							0.716



图 10 裂缝掩膜的提取

Fig. 10 Extraction of the crack mask

### 4 结论

本文提出一种基于多尺度模块和残差模块的网络架构,对具有复杂背景的钢梁裂缝进行识别,针对数据不足的问题,采用滑动窗口和数据增强来扩充数据集,训练结果表明:

1) 与传统的 AlexNet、VGG16 网络模型相比,IRnet 模型收敛速度更快,训练过程更加平稳。此外,使用多尺度模块与残差模块融合对裂缝的识别准确率达到 98.8%,高于其他网络模型。

2) 对提取的裂缝图像进行阈值分割,使用 OTSU 法和中值滤波平滑裂缝图像,结合形态学处理和小连通区域去噪实现了较好的裂缝提取。

3) 改变传统的阈值分割方法,利用 CNN+图像处理的两阶段检测方法,可以在复杂背景下较好地提取裂缝掩膜,从而使其更容易测量裂缝。

### 参考文献:

[1] HSIEH Y A, TSAI Y J. Machine learning for crack detection: review and model performance comparison [J]. Journal of Computing in Civil Engineering, 2020, 34 (5): 04020038.

[2] CHA Y J, CHOI W, ORAL B. Deep learning-based crack damage detection using convolutional neural networks[J]. Computer-aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.

[3] XU H, SU X, WANG Y, et al. Automatic bridge crack detection using a convolutional neural network [J]. Applied Sciences, 2019, 9(14): 2867.

[4] XU Y, BAO Y, CHEN J, et al. Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images [J]. Structural Health Monitoring, 2019, 18(3): 653-674.

[5] CHA Y J, CHOI W, SUH G, et al. Autonomous struc-

- tural visual inspection using region-based deep learning for detecting multiple damage types[J]. *Computer-aided Civil and Infrastructure Engineering*, 2018, 33(9): 731-747.
- [6] 杨铄, 许清风, 王卓琳. 基于卷积神经网络的结构损伤识别研究进展[J]. *建筑科学与工程学报*, 2022, 39(4): 38-57.
- YANG S, XU Q F, WANG Z L. Research progress on structural damage detection based on convolutional neural networks[J]. *Journal of Architecture and Civil Engineering*, 2022, 39(4): 38-57.
- [7] ZHANG J, LI J, LY R, et al. Deep learning-based fatigue cracks detection in bridge girders using feature pyramid networks [DB/OL]. (2024-10-28) [2025-10-02]. <https://arxiv.org/abs/2410.21175>.
- [8] XIONG C B, LIAN S, CHEN W. Detection and location of steel structure trestle surface cracks based on consumer-grade camera system[J]. *KSCE Journal of Civil Engineering*, 2023, 27(3): 1150-1165.
- [9] HAN Q, LIU X, XU J. Detection and location of steel structure surface cracks based on unmanned aerial vehicle images[J]. *Journal of Building Engineering*, 2022, 50: 104098.
- [10] 夏子祺, 马临原, 单伽程, 等. 基于计算机视觉的建筑外墙剥落和裂缝两阶段检测方法[J]. *建筑结构学报*, 2023, 44(2): 207-216.
- XIA Z Q, MA L Y, SHAN J Z, et al. Computer vision-based two-step inspection method for spalling and cracks of building facades[J]. *Journal of Building Structures*, 2023, 44(2): 207-216.
- [11] NGUYEN N H T, PERRY S, BONE D, et al. Two-stage convolutional neural network for road crack detection and segmentation[J]. *Expert Systems with Applications*, 2021, 186: 115718.
- [12] WEN X, LI S, YU H, et al. Multi-scale context feature and cross-attention network-enabled system and software-based for pavement crack detection [J]. *Engineering Applications of Artificial Intelligence*, 2024, 127: 107328.
- [13] 张健飞, 蔡东成. 基于多尺度卷积神经网络的结构损伤识别研究[J]. *地震工程与工程振动*, 2022, 42(1): 132-142.
- ZHANG J F, CAI D C. Research on structural damage identification based on multi-scale convolutional neural networks[J]. *Earthquake Engineering and Engineering Dynamics*, 2022, 42(1): 132-142.
- [14] 谢东阳, 李丽宏, 苗长胜. 基于改进 AlexNet 卷积神经网络的手写体数字识别[J]. *河北工程大学学报(自然科学版)*, 2021, 38(4): 102-106.
- XIE D Y, LI L H, MIAO C S. Handwritten number recognition based on improved AlexNet convolutional neural network[J]. *Journal of Hebei University of Engineering (Natural Science Edition)*, 2021, 38(4): 102-106.
- [15] 张会敏, 谢泽奇. 基于多尺度注意力卷积神经网络的苹果叶部病害识别方法[J]. *江苏农业科学*, 2023, 51(16): 154-161.
- ZHANG H M, XIE Z Q. An apple leaf disease identification method based on multi-scale attention convolution neural network [J]. *Jiangsu Agricultural Sciences*, 2023, 51(16): 154-161.
- [16] OTSU N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1):62-66.
- [17] 梁肇峻, 钟俊. 基于 Otsu 算法与直方图分析的自适应 Canny 算法的改进[J]. *现代电子技术*, 2019, 42(11): 54-58.
- LIANG Z J, ZHONG J. Adaptive canny algorithm improvement based on Otsu algorithm and histogram analysis [J]. *Modern Electronics Technique*, 2019, 42(11): 54-58.
- [18] TANG Z, HU F Q, QUAN Y, et al. APES2018 steel girder crack ID data set [EB/OL]. (2018) [2025-10-25]. [https://github.com/dawnao/APES2018\\_Steel\\_Girder\\_Crack\\_ID\\_dataset](https://github.com/dawnao/APES2018_Steel_Girder_Crack_ID_dataset).
- [19] 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.
- MENG Q C, WAN D, WU H J, et al. Image recognition method of concrete cracks based on convolutional neural network[J]. *Journal of Shenyang Jianzhu University (Natural Science)*, 2021, 37(5): 832-840.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [21] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. (2015-04-10) [2025-10-02]. <https://arxiv.org/abs/1409.1556>.
- [22] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization [J]. *International Journal of Computer Vision*, 2020, 128(2):336-359.

(责任编辑 张爱丽)