DOI: 10.13382/j.jemi.B2407182

改进 U 型网络的路面缺陷图像分割算法*

邹凯鑫[1] 张自嘉[1,2] 孙 伟[1,2] 付锦燚[1]

(1. 南京信息工程大学自动化学院 南京 210044;2. 南京信息工程大学江苏省大气环境与装备技术协同创新中心 南京 210044)

摘 要:针对路面缺陷图像存在对比低以及其复杂的拓扑结构,现提出的大部分分割算法,在捕获感受野和提取路面缺陷特征时还存在很大的不足。因此本文提出改进 U-Net 的路面缺陷图像分割算法。首先,在经典的 U-Net 的卷积块中提出 SN-Disout 残差块,增强模型对过拟合的鲁棒性;其次,在编码器与解码器之间引入循环十字交叉模块,增强模型在特征图中不同位置之间捕获特征的能力,对缺陷的边界进行更准确的建模;最后,在解码器中引入空间通道挤压与激励模块,这使得网络能够更加专注于重要的特征,同时减少不相关或噪声特征的依赖;并将位置感知多头注意力加入模型的颈处,进一步有助于模型更好地理解和利用输入数据的内在关系,从而提高模型的性能和表现能力,并且使用混合损失函数 Dice+BCE 取代单一的损失函数。该算法在 Crack500 图像数据集上的交并比与 $F1$ 分别达到了 60.13% 和 75.22%,均超过 U-Net、DeepLabV3+、PSPNet、TransU-Net、UNet++等主流语义分割网络。实验结果表明,该算法能有效改善网络的预测精度与细小目标的分割结果,在保证分割精度的情况下也满足了实时性的要求。

关键词: 深度学习;语义分割;路面缺陷;U-Net;注意力机制

中图分类号: TP391.4;U418.6 **文献标识码**: A **国家标准学科分类代码**: 520.20

Improving the road defect image segmentation algorithm of U-Net

Zou Kaixin[1] Zhang Zijia[1,2] Sun Wei[1,2] Fu Jinyi[1]

(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: In view of the low contrast and complex topological structure of pavement defect images, most of the currently proposed segmentation algorithms still have great shortcomings in capturing the receptive fields and extracting pavement defect features. Therefore, this article proposes an improved U-Net road defect image segmentation algorithm. First, the SN-Disout residual block is proposed in the classic U-Net convolution block to enhance the model's robustness against overfitting. Secondly, a criss-cross module is introduced between the encoder and the decoder to enhance the model's ability to capture features between different positions in the feature map and more accurately model the boundaries of defects. Finally, the spatial channel squeeze and excitation module is introduced in the decoder, which enables the network to focus more on important features while reducing the dependence on irrelevant or noisy features; position-aware multi-head attention is added to the neck of the model to further helps the model better understanding and utilizing the internal relationship of the input data, thereby improving the performance and performance capabilities of the model, and using the hybrid loss function Dice+BCE to replace a single loss function. The intersection ratio and $F1$ of this algorithm on the Crack500 image data set reached 60.13% and 75.22% respectively, both exceeding mainstream semantic segmentation networks such as U-Net, DeepLabV3+, PSPNet, TransU-Net, and UNet++. Experimental results show that this algorithm can effectively improve the prediction accuracy of the network and the segmentation results of small targets, and it also meets the real-time requirements while ensuring

收稿日期: 2024-01-08 Received Date: 2024-01-08

*基金项目:国家自然科学基金(62376128)项目资助

segmentation accuracy.

Keywords: deep learning; semantic segmentation; pavement cracks; U_Net; attention mechanism

0 引 言

路面缺陷检测是道路维护和安全保障的重要环节。随着交通运输的日益发展,道路的使用频率也越来越高,因此路面的损坏和缺陷问题日益突出。路面缺陷不仅影响驾驶体验,还会给车辆和行人带来安全隐患。因此准确快速获取路面信息状况已成为道路养护的首要任务。

传统的路面缺陷检测方法主要依赖于人工巡查,但这种方式效率低下并且受制于人力资源。因此,借助于技术手段对路面缺陷实现快速准确的检测具有较大的实际需求。

在路面缺陷检测的早期阶段,常常使用图像处理的方法对路面进行缺陷检测,常用的图像处理方法包括边缘检测算法[1]、如 Canny 算子,Sobel 算子等,来识别图像的路面缺陷边界,或者通过阈值分割[2](根据灰度值或颜色信息来设置阈值),设置合理的阈值将路面背景和缺陷分离,进而应用形态学运算去除多余噪声,连接相邻的缺陷区域。但此类方法有一定的局限性,受限于图像质量、依赖手动提取特征、对复杂场景有局限性,并且难以满足处理实时性的要求。

近年来,深度学习的研究人员在图像分类、目标检测、目标分割等计算机视觉任务取得了很大的突破。深度学习是一种机器学习方法,其基本思想是通过多层非线性变换学习数据的高级抽象表示,从而实现对复杂模式和规律的自动发现和理解。这种方法的核心是使用神经网络构建多层非线性映射函数,将输入数据逐层变换为更高维、更抽象的表示。通过反向传播算法,神经网络可以自适应地调整网络参数,最小化输出误差并提高模型的泛化能力。

语义分割是计算机视觉领域中的一种重要技术,旨在将图像的每个像素分配给一个语义类别。与对象检测任务不同,语义分割不仅需要识别出图像中的对象,还需要对每个像素进行分类。通过将缺陷和背景分别标记为正负类别,可以生成一张二值分割结果图,从而实现对路面缺陷的检测[3]。

U-Net 是一种经典的模型,广泛用于语义分割领域[4]。它最初由 Ronneberger 等[5]在 2015 年提出,并因其独特的对称结构而得名,形状类似字母“U”。最初,这个模型主要用于医学图像分割,但由于其卓越的性能,后来被广泛应用于各种不同的分割任务[6]。Liu 等[7]研究人员在缺陷检测方面的工作中,首次采用了 U-Net 网络模型,并通过对损失函数进行改进以及使用优化器进行训练,他们在规模较小的数据集上实现了较高的精度。Cha 等[8]采用图像分块原理,并通过一种图像扫描块算法,对每一幅图像进行扫描以生成图像块。随后,使用二分类方法对这些图像块进行分类,保留了包含缺陷的部分。虽然该方法在缺陷检测方面取得了显著的成果,但无法确定缺陷的具体位置和形状。惠冰等[9]采用 VGG16 作为主要的特征提取网络,以增强其特征提取能力。在网络的解码部分引入了缺陷注意力单元,使网络能够更加关注缺陷信息,从而提高分割的准确度。然而,对于细长的缺陷,该方法的分割效果并不理想。常惠等[10]在 U-Net 网络中引入了残差模块,以增强特征传播并减少信息损失。随后,引入了注意力机制 SE 模块,用于增强对关键区域信息的提取。这一改进在铁路隧道缺陷数据集上实现了高精度的分割,并且运行速度更快。曹锦纲等[11]在 U-Net 中引入了一种名为 AFM 模块的注意力机制,旨在更好地利用全局信息并提高对各种尺度缺陷识别的准确性。他们创建的 ACNet 在公共缺陷数据集 CFD[12]和 CRACK500[13]上,相较于典型的 U-Net 和传统图像处理方法,表现出更精确的缺陷定位和更丰富的细节呈现,另一方面虽然该网络在编码和解码部分引入了基于注意力机制的特征提取模块和解码模块,实现了缺陷的精确定位。然而,该方法仍有提升准确度和精度的空间。

综上所述,当前在路面缺陷图像分割任务中,缺陷图像具有不同于常见语义分割数据集的特点,这也是分割任务中的难点所在。首先,缺陷形状曲折多变,拓扑结构和宽度各异,传统卷积无法有效提取多尺度特征,导致信息损失;其次,缺陷通常贯穿整张图片,传统 U-Net 感受野较小,全局信息提取能力不足;最后,前景与背景像素严重不平衡,传统损失函数难以有效处理背景渐变主导的问题。针对上述问题,提出一种改进 U 型网络算法来应对路面缺陷分割带来的挑战,该算法的主要工作有:

1)设计了 SD-Inception 模块,利用该模块,重新设计了 U-Net 的卷积层结构,以提高特征重用效率、确保梯度稳定性,并增强模型的泛化能力。

2)在编码器与解码器之间添置一种可以促进特征融合与信息传递的模块 RCCA,增强语义信息传输能力,有助于提升模型对路面缺陷的表示能力。

3)在模型的脖颈处引入一种带有位置感知的多头注意力机制 PA_MHSA,帮助模型更好地理解整个路面缺陷图像的上下文信息。

4)添加 BCE+Dice 混合损失函数,在路面缺陷分割中,往往存在正负样本不均衡的问题,利用 Dice 损失函

数处理不平衡数据集的问题,提高对路面缺陷的关注程度。并将其与 BCE 结合起来,从而增强模型鲁棒性并优化训练过程。

1　本文的网络结构

1.1　改进的 U-Net 算法结构

U-Net 常用于语义分割任务的卷积神经网络结构。其主要结构为:编码器和解码器。编码器由 4 个块(主要由卷积和池化组成)组成,用于逐步提取图像的特征信息并且同时减少空间维度,每经过一次下采样特征图的通道数都会增加,以此可以捕捉到不同尺度的信息。在解码器每一层中,会连接来自编码器相应层的特征图,以此来获取高层次和低层次的特征信息。

针对路面缺陷的识别特点,本文对 U-Net 算法进行了如下几方面的改进:如图 1 所示,右半边 Decoder 和左半边 Encoder 中的下采样块和上采样块引入了可变换的归一化(switchable normalization,SN)层和防过拟合层(Disout),同时加入残差单元,残差单元由两个 3 × 3 卷积块、单位映射组成。每个卷积块包括 1 个卷积层、1 个 Disout 层、1 个 SN、1 个 ReLU 激活层。其次,在瓶颈处加入 PA-MHSA 多头注意力可以增加对全局和局部特征之间关系的建模能力。PA-MHSA 利用自注意力机制,能够根据输入特征图的各个位置之间的相关性来动态调整特征的权重,使得每个位置能够更好地聚焦于与其相关的区域。这样可以提高网络对整体结构和细节特征的感知能力。在解码器上采样之前加入 scSE 模块,有助于改善网络在上采样过程中对局部细节的感知能力,使得网络能够更好地恢复图像中小尺度的缺陷信息,从而提高检测的准确性。引入轻量化的循环十字交叉模块,并将其添加到网络的解码器-编码器部分之间,以进一步提高路面缺陷分割性能。

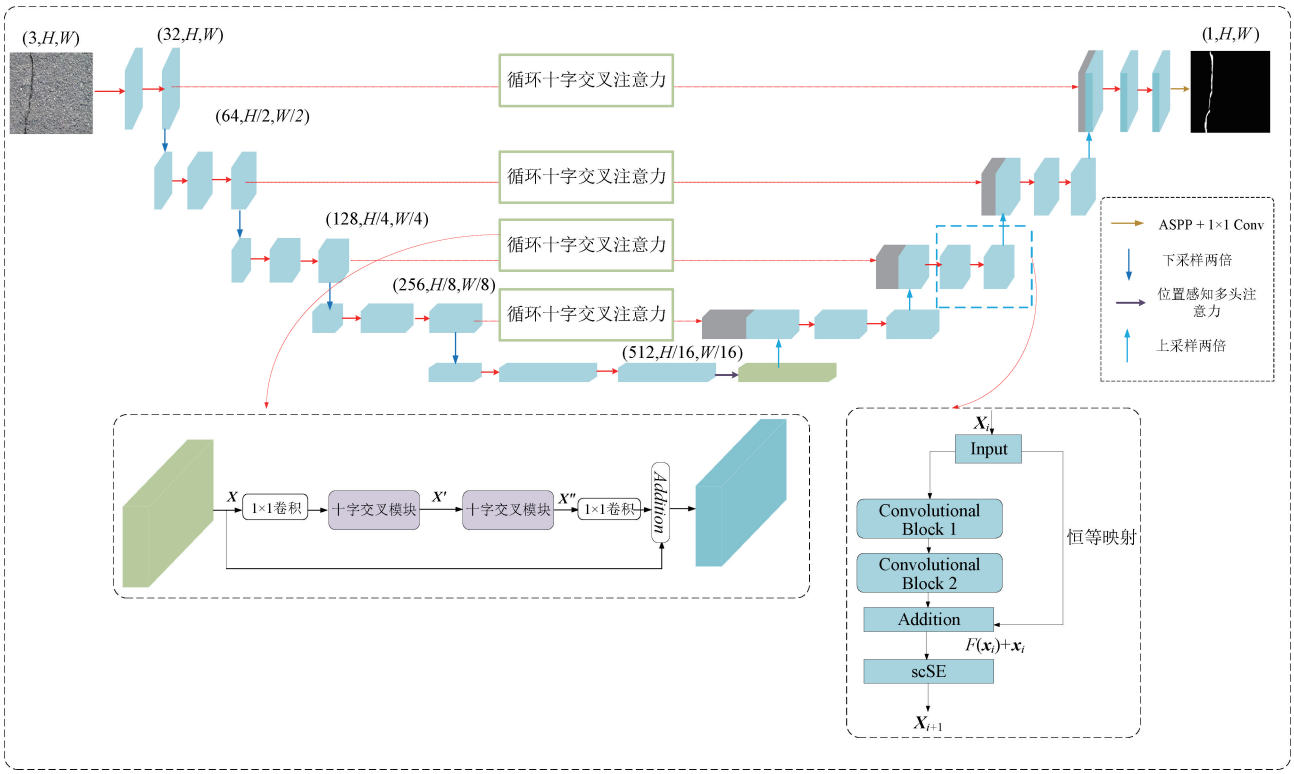


图 1　改进的 U-Net 网络模型

Fig. 1　Improved U-Net network model

1.2　更改网络结构

U-Net 算法的编码器由多个 Block 模块组成,每个 Block 仅包括 3×3 卷积层和 2×2 最大池化层(Max-Pooling)。这种设计虽然减少了参数数量和计算复杂度,但是由于缺陷图像的形状变化多样,有不同的拓扑结构。传统的标准卷积只能获取简单上下文信息,无法同时有效地提取缺陷的细纹和宽缝特征,缺乏多尺度特征提取的能力。因此本文将采用图 1 所示的编码器结构。

归一化是深度学习中非常有效的元素,传统的归一化方法,如批归一化(batch normalization,BN)和层归一化(layer normalization,LN),通常采用固定的归一化策略,无法灵活地适应不同层和不同训练阶段的特征分布变

化。而 SN[14] 则引入了一个可切换的参数来控制使用不同的归一化策略。具体而言,它通过引入一个辅助输入,由网络学习一个权重向量,来控制选择批归一化、层归一化或组归一化(group normalization,GN)中的哪一种作为当前层的归一化方式。这样做的好处是,不同类型的归一化方法适用于不同类型的数据分布和场景。比如,批归一化适用于大批量数据的训练,层归一化适用于逐层处理的网络,而组归一化适用于小批量或具有分组结构的数据。SN 的灵活性使得模型可以根据具体情况选择最适合的归一化策略,从而提升网络的表达能力和性能。SN 在一些计算机视觉任务中取得了很好的效果,特别是在处理具有不同分布特征的数据时,它的可切换性可以很好地适应不同数据的变化。

SN 定义式为:

$$\hat{\boldsymbol{h}}_{ncij}=\gamma\frac{\boldsymbol{h}_{ncij}-\sum_{k\in\Omega}w_k\mu_k}{\sqrt{\sum_{k\in\Omega}w'_k\sigma_k^2+\epsilon}}+\beta \tag{1}$$

其中,$\boldsymbol{h}_{ncij}$ 和 $\hat{\boldsymbol{h}}_{ncij}$ 分别为归一化前后的像素值并且是四维像素值,其中四维包括样本数、通道数、高度和宽度(N,C,H,W)。n、c、i、j 取值分别为 $1\sim N$、$1\sim C$、$1\sim H$、$1\sim W$。β 和 γ 分别是标度和位移参数。ϵ 是一个保持数值稳定性的小常数。Ω 是一组统计量。w_k 和 w'_k 分别是对均值和方差进行加权平均的重要比值。这些参数取 0~1 的动态值,并通过反向传播学习。

Disout 是一种用于神经网络的正则化方法,用于防止过拟合[15]。它是 Dropout 的扩展,可以在训练过程中随机丢弃部分输入特征,以减少神经网络的过度依赖于个别特征的情况。与传统的 Dropout 不同,Disout 对于每个输入特征都有一个概率 p 来决定是否应该被丢弃。Dropout 只是对每个神经元(或特征)有一个概率 p 来丢弃。

Disout 定义式为:

$$\hat{f}^{\,l}(\boldsymbol{x}_i)=f^{\,l}(\boldsymbol{x}_i)-m_i^l\circ\varepsilon_i^l \tag{2}$$

其中,$\boldsymbol{x}_i$ 是输入数据,$f^{\,l}(\boldsymbol{x}_i)$ 是 l 层的特征映射,$\hat{f}^{\,l}(\boldsymbol{x}_i)$ 是失真特征映射,m 为二值掩码,ε 为特征映射上利用的失真度。圆圈表示与元素相关的乘积。正如在全连接层一样,(m + 1) 层的 ERC 来指导 m 层 ε 优化。不是固定值,而是在 ERC 的指导下自动学习。i 的取值范围从 1 到输入样本总数。

考虑到缺陷特征在图像中占比较小,需要足够细腻的特征信息。本文在此基础上加了 scSE 模块,旨在捕获多尺度空间与通道信息来丰富特征空间[16]。如图 2 所示,scSE 模块是对 SE 模块[17-18]的改进,在此基础上提出了通道激励与空间挤压(cSE)、空间激励和通道挤压(sSE)、空间和通道挤压和激励(scSE)。scSE 模块是 cSE 和 sSE 模块的综合体,即同时对空间维度和通道维度进行信息整合,将两者的特征结果沿着通道维度使用 max-out 聚合[19]。scSE 模块是由 cSE 和 sSE 两个模块组成。

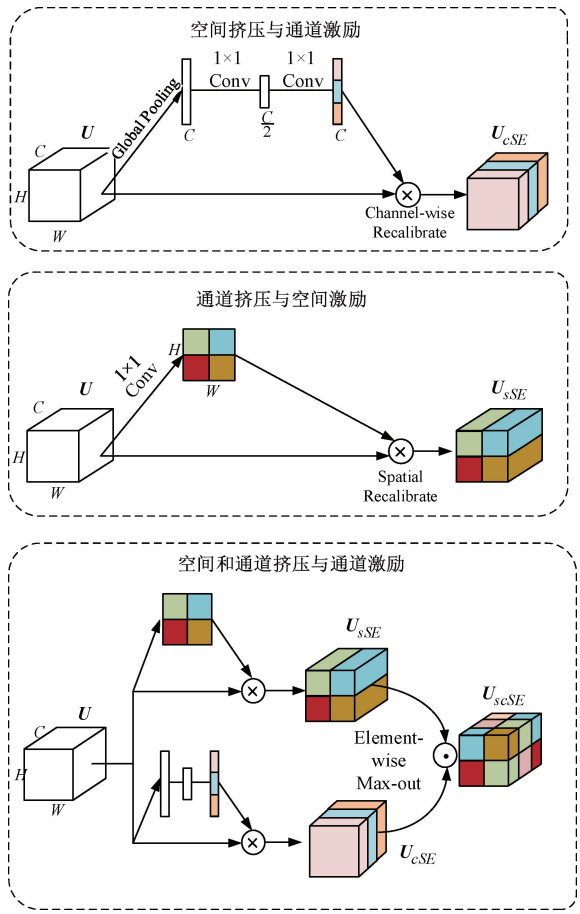


图 2 空间通道挤压与激励

Fig. 2 Space channel squeeze and excitation

其定义式如下:

$$\hat{\boldsymbol{U}}_{cSE}=[\sigma(\hat{z}_1)\boldsymbol{u}_1,\sigma(\hat{z}_2)\boldsymbol{u}_2,\cdots,\sigma(\hat{z}_c)\boldsymbol{u}_c] \tag{3}$$

$$\hat{\boldsymbol{U}}_{sSE}=\begin{bmatrix}\sigma(\boldsymbol{q}_{1,1})\boldsymbol{u}^{1,1},\cdots,\sigma(\boldsymbol{q}_{i,j})\boldsymbol{u}^{i,j},\cdots,\\ \sigma(\boldsymbol{q}_{H,W})\boldsymbol{u}^{H,W}\end{bmatrix} \tag{4}$$

$$\hat{\boldsymbol{U}}_{scSE}(i,j,c)=\max(\hat{\boldsymbol{U}}_{cSE}(i,j,c),\hat{\boldsymbol{U}}_{sSE}(i,j,c)) \tag{5}$$

其中,$\sigma(\hat{z}_1)$、$\sigma(\hat{z}_2)$、…、$\sigma(\hat{z}_c)$ 表示通道 $1\sim c$ 的重要性,它可以放大或者缩小。随着网络的学习,这些参数会自动地调整,忽略不太重要的通道,并强调重要的通道。原始特征图先通过全局平均池化(global average pooling, GAP)将其维度从 [C,H,W] 变为 [C,1,1](进行空间挤压),使用两个 1×1 卷积进行处理(先降维后升维),最终得到 C 维向量,紧接着使用 sigmod 函数进行归一化,得到 $\sigma(\hat{z}_1)$、$\sigma(\hat{z}_2)$、…、$\sigma(\hat{z}_c)$,最后通过 channel-

wise 与原始特征图相乘(通道激励),最终得到 $\hat{\boldsymbol{U}}_{cSE}$。

$\sigma(\boldsymbol{q}_{1,1})$、…、$\sigma(\boldsymbol{q}_{i,j})$、…、$\sigma(\boldsymbol{q}_{H,W})$ 表示对于给定特征映射的空间信息 (1,1)、…、(i,j)、…、(H,W) 相对重要性。随着反向传播,参数更新,这些参数会更加注重相关的空间位置,忽略不相关的空间位置。原始特征图先通过输出通道为 1,卷积核大小为 1×1 的卷积层,得到一个维度为 $[1,H,W]$ 参数矩阵(进行通道挤压),之后将参数矩阵进行 sigmod 归一化处理,得到最终的参数矩阵 $\sigma(\boldsymbol{q}_{1,1})$、…、$\sigma(\boldsymbol{q}_{i,j})$、…、$\sigma(\boldsymbol{q}_{H,W})$。将参数矩阵同原始特征图在空间维度相乘(空间激励),最终得到 $\hat{\boldsymbol{U}}_{sSE}$。

$\hat{\boldsymbol{U}}_{scSE}(i,j,c)$ 引入 cSE 和 sSE 模块能够使网络更加关注输入特征图中的重要空间信息。这有助于捕捉到物体的轮廓、边界以及其他局部信息,从而提高感知能力。还有助于放大具有更强判别性的通道,抑制噪声和不重要的通道,从而增强特征的表达能力。结合空间和通道两个模块,能够更好捕捉不同空间位置和通道之间相关性,提高模型的性能。

受深度残差学习[20-21]和 U-Net 的启发,本文提出了深度残差 U-Net,如图 3 所示引入的残差模块,这是一种利用深度残差学习和 U-Net 架构优势的架构[22]。本文提出的深度残差 U 网是在 U 网架构的基础上构建的。这种组合带来了两个好处:1)残差单元将简化网络的训练;2)残差单元内部和网络的低级和高级之间的跳过连接将促进信息的传播而不会退化,一方面可以减少模型的参数量,一方面也可提高模型的分割性能。

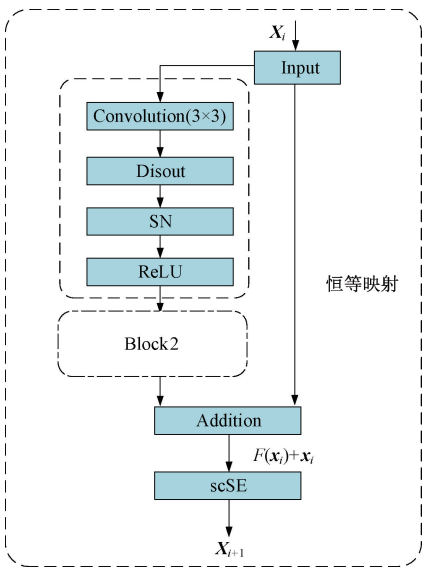


图 3　残差块结构

Fig. 3　Residual block structure

损失函数是用来衡量模型预测结果和真实结果之间的差异或错误程度的函数。在训练神经网络时,需要通过优化损失函数来使模型的预测结果尽可能地接近真实结果。由于路面缺陷分割属于二分类问题并且由于其正负样本及其不平衡等特点,本文采用了一种结合二分类交叉熵损失(binary cross entropy,BCE)和 Dice 损失的混合损失函数来克服该问题。混合损失函数公式如式(6)所示。

$$L_{BCED} = L_{BCE} + L_{Dice} \tag{6}$$

其中,L_{BCE} 与 L_{Dice} 公式如下:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \lg p_i + (1-y_i)\lg(1-p_i)] \tag{7}$$

式中:N 为像素总数,y_i 为第 i 个像素的标签像素值,p_i 为第 i 个像素的预测值。

$$L_{Dice} = 1 - \frac{|X \cap Y|}{|X| + |Y|} \tag{8}$$

式中:$|X|$ 和 $|Y|$ 表示 X 和 Y 的像素个数,而 $|X \cap Y|$ 表示 X 和 Y 具有相同像素点的个数。Dice Loss 通过 mask 掩码操作,固定大小的正样本区域计算的损失不因图片大小变化,有利于挖掘前景信息。然而,单一的 Dice 损失函数有时会导致梯度波动较大。因此,采用“BCE+Dice”混合损失函数可以在保证训练稳定的条件下,很好地处理正负样本不平衡的问题。

1.3　位置感知多头注意力(PA-MHSA)

由于 U-Net 的结构限制,模型的感知范围可能受限,难以有效地处理远距离像素之间的关系。通过引入位置感知多头注意力(PA-MHSA),可以扩大模型的感知范围,有助于更好地捕获像素之间的长距离依赖关系[23-24]。它位于 U-Net 瓶颈处,MHSA 的结构如图 4 所示。

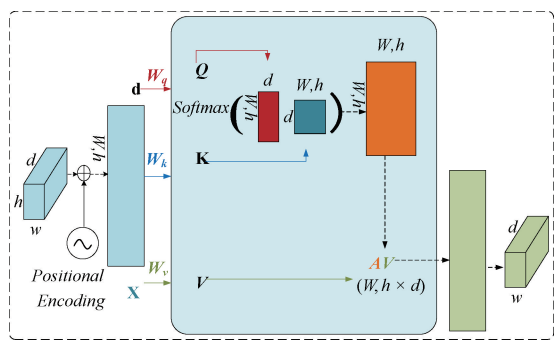


图 4　PA-MHSA 位置感知多头注意力

Fig. 4　PA-MHSA position-aware multihead self-attention

MHSA 公式如式(9)所示。

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} = \boldsymbol{A}\boldsymbol{V} \tag{9}$$

其中,自关注模块有 3 个输入,1 个查询矩阵 $\boldsymbol{Q}$,1 个键矩阵 $\boldsymbol{K}$ 和 1 个值矩阵 $\boldsymbol{V}$,$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 具有相同的大小。此外,为了考虑绝对上下文信息,在输入特征中添加了位置编码。可以利用位置编码来捕获 MHSA 中相关特征的绝

对和相对位置。

1.4 十字交叉模块(Criss-Cross Module)

路面缺陷具有不同尺度的特征,有些缺陷非常细小,有些则比较大,并且其拓扑结构极不规则,分布在路面上各个位置,十字交叉模块(criss-cross module)能够在全局和局部层面上对信息进行整合,有助于更全面地理解缺陷的分布情况和特征。因此将轻量化十字交叉模块添加到网络的编码器-解码器部分之间。CCA 的结构如图 5 所示。

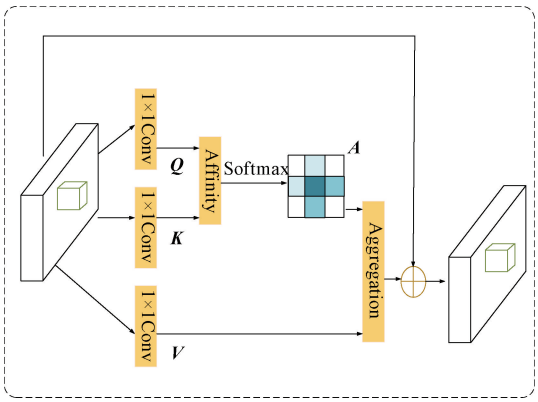


图 5 十字交叉模块

Fig. 5 Criss-Cross Module

引入 CCA 可以帮助模型更好地理解图像中缺陷的边界信息[25]。CCA 能够捕捉到特征图中不同位置之间的依赖关系,对缺陷的边界进行更准确的识别。这有助于提高缺陷边界分割的准确性和清晰度。CCA 公式如式(10)所示。

$$\boldsymbol{X}'_u = \sum_{i=0}^{H+W-1} \boldsymbol{A}_{i,u} \boldsymbol{\Phi}_{i,u} + \boldsymbol{X}_u \tag{10}$$

其中,输入特征 $\boldsymbol{X} \in \mathbf{R}^{C\times W\times H}$,$\boldsymbol{X}$ 经过 3 个卷积块分别生成 $\boldsymbol{Q}$、$\boldsymbol{K}$、和 $\boldsymbol{V}$,其中 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的通道大小小于输入特征 $\boldsymbol{X}$ 的通道 C。$\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 经过 Affinity 和 Softmax 操作计算生成 $\boldsymbol{A} \in \mathbf{R}^{(H+W-1)\times W\times H}$,对于生成的 $\boldsymbol{V} \in \mathbf{R}^{C\times W\times H}$,在 $\boldsymbol{V}$ 的空间维度的每个位置 u,可以得到 1 个向量 $\boldsymbol{V}_u \in \mathbf{R}^C$ 和 1 个集合 $\boldsymbol{\Phi}_u \in \boldsymbol{R}^{(H+W-1)\times C}$,集合 $\boldsymbol{\Phi}_u$ 是 $\boldsymbol{V}$ 中与位置 u 在同一行或同列的特征向量的集合。将这个向量集与生成的 A 相乘,完成 Aggregation 操作,加上原始输入特征 $\boldsymbol{X}$,输出 $\boldsymbol{X}'$。虽然十字交叉模块可以获取水平和垂直方向的上下文信息,减少模型参数量,但是在非水平和垂直方向的联系缺失了,所以为了捕获密集的全局上下文信息,这里采用循环十字交叉模块。

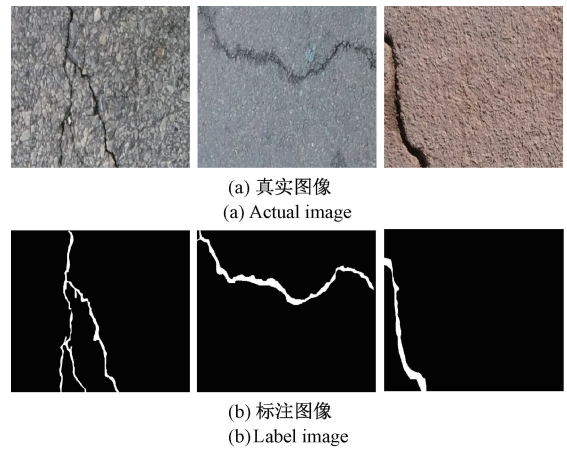
2 数据集

2.1 数据集介绍

本实验采用公开路面缺陷数据集 CRACK500,其包含 500 张高分辨率(2 560×1 440 pixels)的数字图像,该数据集主要是利用手机在美国天普大学(Temple University)校园道路上所拍摄的路面缺陷图像。每张图像都经过了像素级别的标注,包含了缺陷和非缺陷两类区域。CRACK500 包含了各种类型和尺寸的缺陷,以及不同的光照和背景条件。因此,使用 Crack500 数据集进行路面缺陷检测和分割的研究,可以帮助改进算法的鲁棒性和适应性。所以本文采用 CRACK500 数据集。

2.2 数据预处理

CRACK500 中的路面缺陷图像分辨率较大,因此将其裁剪成 3 368 张 640×360 大小的图像。其中训练集大小为 1 896 张,验证集为 348 张,测试集为 1 124 张。在模型训练阶段使用训练集训练模型参数,使模型能够学习数据的特征和模式,用验证集调整模型的超参数、提高模型的泛化能力,最后使用测试集评估模型的性能。图 6 展示了一些路面缺陷的实际图片及其相应标签。



(a) 真实图像

(a) Actual image

(b) 标注图像

(b) Label image

图 6 裁剪之后的真实图像和标签图像

Fig. 6 Cropped real image and label image

3 实 验

3.1 实验环境

实验模型搭载 Pytorch 深度学习框架,基于 64 位 Ubuntu 18.04 操作系统,硬件配置为 Intel Core i7-4790k CPU、Nvidia RTX4060 显卡、16 GB 内存。Batch Size(批量大小,每次输入图片的个数)为 4,初始学习率为 0.001,训练过程使用 Adam 优化器对权重进行更新,代码均使用 Python3.6 实现。

3.2 模型评价指标

本文的路面缺陷分割属于二分类问题,其中正样本为路面裂缝特征,负样本为路面背景。TP(true positive)为路面裂缝预测为路面裂缝,FP(false positive)为路面背

景预测为路面裂缝,FN(false negative)为路面裂缝预测为路面背景,TN(true negative)为路面背景预测为路面背景。如表 1 所示。

表 1　二分类预测结果

Table 1　Prediction results of binary classification

真实值	预测值	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

根据上述表格,本文提出以下几种评价模型的指标:

准确率(precious,P)衡量模型预测为正类的实例中有多少是真正的正类。主要关注模型的误检测情况,实际上是将路面背景特征错误地分类为正例:

$$P = \frac{TP}{TP + FP} \tag{11}$$

召回率(recall,R)衡量模型在所有实际为正类的样本中,正确识别为正类的比例。主要关注模型漏检的情况,即未能正确地将实际上路面裂缝特征分类为正例的比例:

$$R = \frac{TP}{TP + FN} \tag{12}$$

交并比(intersection-over-union,IoU)用于图像分割任务中的指标,衡量模型预测的区域与真实区域之间的重叠程度:

$$IoU = \frac{TP}{TP + FN + FP} \tag{13}$$

$F1$ 分数综合考虑了准确率和召回率的指标,能够平衡模型的精确性和完整性:

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{14}$$

3.3　验证网络结构性能

为验证本文模型中每个改进部分的有效性,接下来对改进模块进行消融实验,以验证提出模型的有效性。

如图 7 与表 2 所示,在实验 b、c、d、e 中每个改进方法都对模型性能有一定的提升效果并且各项精度指标都高于 U-Net。在实验 f 中,将 SD-Inception 网络结构代替原有卷积模块并将 scSE 模块添加到解码器中,IoU 与 $F1$ 两个指标对比实验 b 又提升了 1.1 与 1.31。在实验 f 的基础上,实验 g 将 RCCA 模块嫁接在编码器与解码器之间,IoU 与 $F1$ 两个指标分别提升了 1.9 与 1.57,为了进一步提升路面细小缺陷检测能力,在实验 g 基础上,将位置感知多头注意力机制添加到模型的脖颈处,IoU 与 $F1$ 两个指标分别又提升了 0.38 与 0.57。最后使用 Dice+BCE Loss,IoU 与 $F1$ 提升了 0.7 与 0.59。

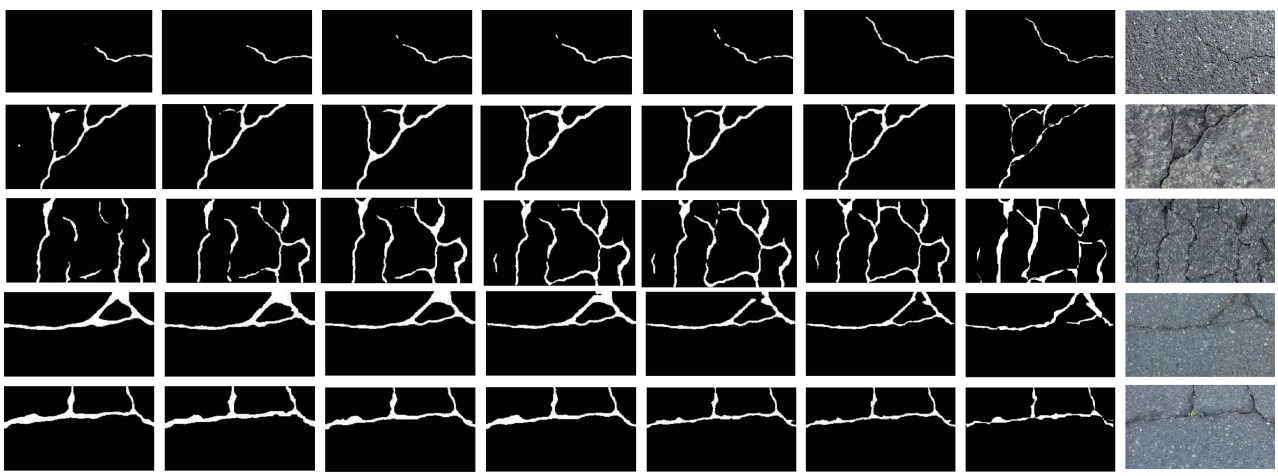


图 7　消融实验的可视化结果

Fig. 7　Visualized results of ablation experiments

3.4　对比实验

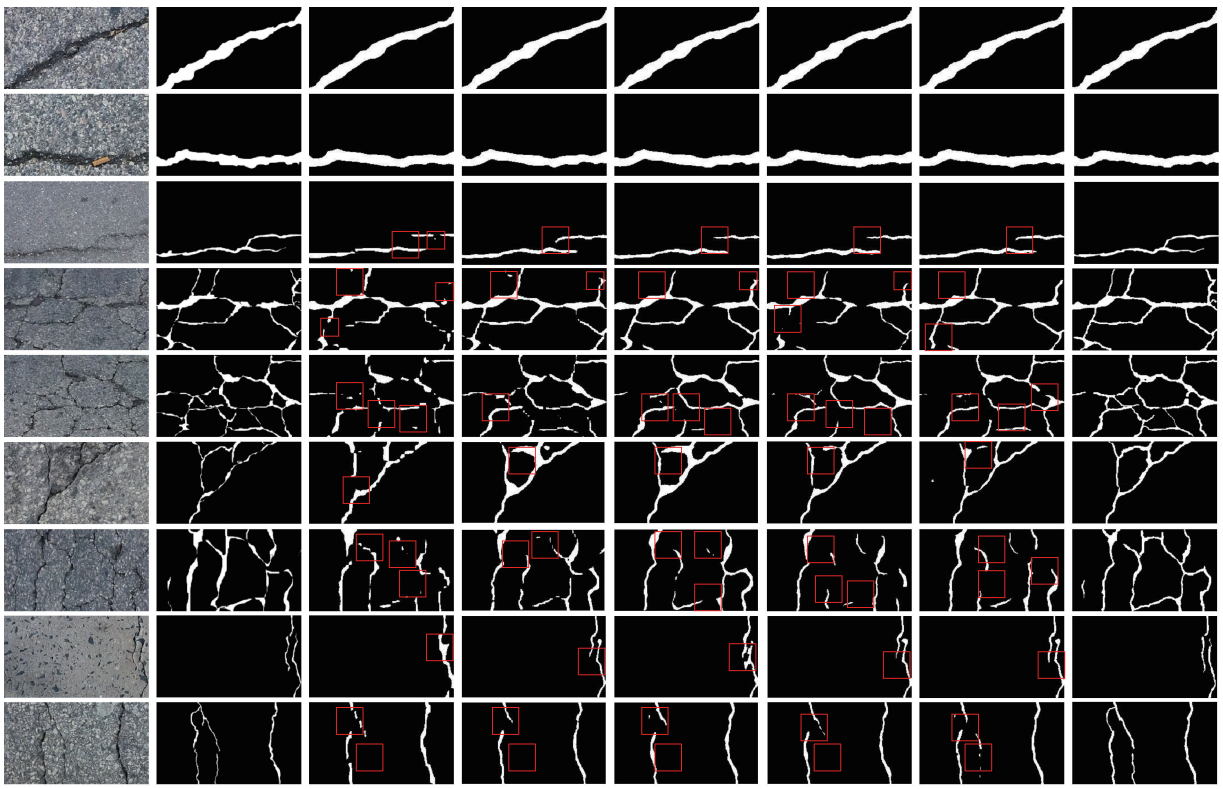
为了验证本文所提出的网络结构的优越性,本实验选取了 5 个较为主流的语义分割方法与本文方法进行对比实验,分别为编码器和解码器的 U_Net 模型、通过引入改进的金字塔池化模块来捕捉不同尺度下的上下文信息的 PSPNet[26]、融合空洞卷积、ASPP 模块和编码器-解码器结构来实现高精度、高分辨率的图像语义分割 DeepLabv3+[27]、基于 Transformer 架构的模型,它利用自注意力机制来捕捉图像中的长距离依赖关系 TransUnet[28],该模型能够自适应地学习图像中的重要区域,并在分割过程中更加关注这些重要区域,以提高分割准确性、引入额外的上采样路径和多级聚合模块的 UNet++[29]。各个模型使用效果最佳的网络模型进行识别。

表 2 消融实验有效性

Table 2 Effectiveness of ablation experiments

实验	方法	*P*/%	*R*/%	*IoU*/%	*F*1/%
a	U-Net	70.45	68.89	54.59	69.65
b	+SD-Inception	71.20	71.25	56.06	71.18
c	+scSE	70.89	70.37	55.79	70.63
d	+RCCA	71.17	71.97	56.63	71.56
e	+PA-MHSA	70.67	71.56	55.93	71.12
f	+SD-Inception+scSE	72.42	72.97	57.16	72.49
g	+SD-Inception+scSE+RCCA	71.73	76.53	59.06	74.06
h	+SD-Inception+scSE+RCCA+PA-MHSA	71.77	77.83	59.44	74.63
i	+SDInception+scSE+RCCA+PA-MHSA+Loss function	72.04	78.70	60.14	75.22

如表 3 所示,本文所提算法的各项精度指标均高于其他模型。其中 IoU 指标对比表 3 实验模型分别上升了 2.61%~6.35%,*F*1 指标也上升了 3.5%~6.45%。与 PSPNet 模型相比,本文方法模型体积减少 144 MB,检测效率下降了 22.89 fps。虽然 PSPNet 检测效率高,但是却以牺牲精度为代价,其各项精度指标都非常低。相较于传统的 U-Net 模型,本研究中改进后的模型在准确率上提升了 1.59%,召回率提高了 9.81%。这主要归因于召回率和准确率之间相互影响的关系,提高其中一个指标往往会导致另一个指标下降。召回率的增加有助于更好地检测路面缺陷,减少漏检情况,从而最大程度地降低潜在的安全隐患,并有利于制定精准的路面养护决策。由图 8 的可视化结果可知,当路面缺陷特征较为明显时,这 6 种模型都能很好的识别出路面缺陷特征。而当路面缺陷特征变得细长并且路面缺陷特征与路面背景对比度低的情况下,其他实验模型均出现了漏检的情况,而本文模型模漏检少,能够更好的识别出这种细小的路面缺陷特征,分割结果更加贴近真实标签。因此,本文改进的模型在路面缺陷图像分割方面实现了更加精细的效果。



(a) 原图像 (a) Original image (b) 标签 (b) Tags (c) PSPNet (d) 改进的DeepLabV3+ (d) Improved DeepLabV3+ (e) TransUet (f) UNet++ (g) U-Net (h) 本文方法 (h) Method of this article

图 8 6 种模型的可视化结果

Fig. 8 Visualization results of 6 models

表 3　各模型在测试集上的分割结果

Table 3　Segmentation results of each model on the test set

模型名称	*P*/%	*R*/%	*IoU*/%	*F*1/%	Parameter/M	FPS
PSPNet	63.56	74.88	53.79	68.77	187	49.06
改进的 DeepLabV3+	67.69	76.25	56.53	71.72	227	19.56
TransUet	67.66	75.38	56.31	71.32	133	21.45
UNet++	68.94	74.08	57.53	71.42	55	26.53
U-Net	70.45	68.89	54.59	69.65	37	29.51
本文方法	72.04	78.70	60.14	75.22	43	27.17

4　结　论

针对路面缺陷图像存在对比度低、拓扑结构复杂等问题,本文设计了一种改进 U 型网络路面缺陷图像分割算法。通过使用 SD-Inception 网络结构、RCCA 模块、引入带有位置感知的 PA_MHSA 多头注意力机制、添加混合损失函数 BCE+Dice,提升了模型对路面缺陷的表示能力,并且增强了模型的泛化能力。实验结果显示,提出的方法对路面缺陷分割的 IoU 为 60.14%,*F*1 分数为 75.22%,与其他模型对比,也存在着显著优势。在以后的工作中,将研究如何降模型的体积并且同时提高模型对路面缺陷的分割能力,保证分割精度同时加快运行速率,适应资源有限场景下的应用。

参考文献

[1] 王博, 王霞, 陈飞, 等. 航拍图像的路面缺陷识别[J]. 光学学报, 2017, 37(8): 0810004.
WANG B, WANG X, CHEN F, et al. Pavement crack recognition from aerial images [J]. Acta Optica Sinica, 2017, 37(8): 0810004.

[2] PENG L, CHAO W, LI S M, et al. Research on crack detection method of airport runway based ontwice-threshold segmentation [C]. 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), September 18-20, 2015, Qinhuangdao, China. New York: IEEE, 2015: 1716-1720.

[3] KNIG J, JENKINS M D, BARRIE P, et al. A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating[C]. 2019 IEEE International Conference on Image Processing(ICIP). IEEE, 2019:1460-1464.

[4] 刘国华,闫克丁,邢静,等. 基于改进 U 型网络的细胞核分割方法[J]. 国外电子测量技术,2022,41(11): 1-6.
LIU G H, YAN K D, XING J, et al. Cell nucleus segmentation method based on improved U-shaped network [J]. Foreign Electronic Measurement Technology, 2022, 41(11): 1-6.

[5] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. MICCAI. Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Berlin: Springer, 2015: 234-241.

[6] 何晓云,许江淳,陈文绪. 基于改进 U-Net 网络的眼底血管图像分割研究[J]. 电子测量与仪器学报,2021, 35(10):202-208.
HE X Y, XU J CH, CHEN W X. Research on fundus blood vessel image segmentation based on improved U-Net network [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 202-208.

[7] LIU Z Q, CAO Y W, WANG Y Z, et al. Computer vision-based concrete crack detection using U-Net fully convolutional networks[J]. Automation in Construction, 2019, 104: 129-139.

[8] CHA Y J, CHOI W, BUYUKOZTURK O. Deep learning-based crack damage detection using convolution neural networks [J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.

[9] 惠冰,李远见. 基于改进 U 型神经网络的路面缺陷检测方法[J]. 交通信息与安全, 2023, 41(1): 105-114,131.
HUI B, LI Y J. Pavement crack detection method based on improved U-shaped neural network [J]. Traffic Information and Safety, 2023, 41(1): 105-114,131.

[10] 常惠, 饶志强, 赵玉林, 等. 基于改进 U-Net 网络的隧道缺陷分割算法研究[J]. 计算机工程与应用, 2021, 57(22): 215-222.
CHANG H, RAO ZH Q, ZHAO Y L, et al. Research on tunnel crack segmentation algorithm based on improved U-Net network [J]. Computer Engineering and Applications, 2021, 57(22): 215-222.

[11] 曹锦纲, 杨国田, 杨锡运. 基于注意力机制的深度学

习路面缺陷检测[J]. 计算机辅助设计与图形学学报, 2020, 32(8): 1324-1333.
CAO J G, YANG G T, YANG X Y. Deep learning pavement crack detection based on attention mechanism [J]. Journal of Computer-Aided Design and Graphics, 2020, 32(8): 1324-1333.

[12] SHI Y, CUI L M, QI Z Q, et al. Automatic road crack detection using random structured forests [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(12): 3434-3445.

[13] YANG F, ZHANG L, YU S J, et al. Feature pyramid and hierarchical boosting network for pavement crack detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(4): 1525-1535.

[14] LUO P, ZHANG R, REN J, et al. Switchable normalization for learning-to-normalize deep representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 712-728.

[15] TANG Y, WANG Y, XU Y, et al. Beyond dropout: Feature map distortion to regularize deep neural networks [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5964-5971.

[16] 仝卫国,仪小龙,李冰,等. 融合多尺度特征与注意力机制的风机桨叶缺陷检测方法[J]. 电子测量技术, 2022,45(24):166-172.
TONG W G, YI X L, LI B, et al. Wind turbine blade defect detection method integrating multi-scale features and attention mechanism [J]. Electronic Measurement Technology, 2022, 45(24): 166-172.

[17] 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.

[18] 张平, 佟昆宏, 王学珍. 基于改进 U-Net 网络的液压管路分割方法[J]. 电子测量与仪器学报, 2023, 37(1): 123-129.
ZHANG P, TONG K H, WANG X ZH. Hydraulic pipeline segmentation method based on improved U-Net network [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 123-129.

[19] ROY A G, NAVAB N, WACHINGER C. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks[J]. IEEE Transactions on Medical Imaging, 2018, 38(2): 540-549.

[20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[21] 张栋,姜媛媛. 融合注意力机制与逆残差结构的轻量级钻机目标检测方法[J]. 电子测量与仪器学报, 2022,36(11):201-210.
ZHANG D, JIANG Y Y. Lightweight drilling rig target detection method integrating attention mechanism and inverse residual structure [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(11): 201-210.

[22] 邸拴虎, 杨文瀚, 廖苗, 等. 基于 RA-Unet 的 CT 图像肝脏肿瘤分割[J]. 仪器仪表学报, 2022, 43(8): 65-72.
DI SH H, YANG W H, LIAO M, et al. Liver tumor segmentation in CT images based on RA-Unet [J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 65-72.

[23] 陈仁祥,詹赞,胡小林,等. 基于多注意力 Faster RCNN 的噪声干扰下印刷电路板缺陷检测[J]. 仪器仪表学报,2021,42(12):167-174.
CHEN R X, ZHAN Z, HU X L, et al. Printed circuit board defect detection under noise interference based on multi-attention Faster RCNN [J]. Chinese Journal of Scientific Instrument, 2021, 42(12): 167-174.

[24] TAN H, LIU X, YIN B, et al. MHSA-Net: Multihead self-attention network for occluded person re-identification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022,33(11):8210-8224.

[25] ZHANG L, CHENG Y. A densely connected criss-cross attention network for document-level relation extraction[J]. ArXiv preprint arXiv:2203.13953, 2022.

[26] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.

[27] WANG Z, ZHANG H, HUANG Z, et al. Multi-scale dense and attention mechanism for image semantic segmentation based on improved DeepLabV3 + [J]. Journal of Electronic Imaging, 2022, 31(5): 053006.

[28] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. ArXiv preprint arXiv:2102.04306, 2021.

[29] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHSH N, et al. Unet + +: A nested U-Net architecture for medical image segmentation [C]. Deep Learning in

Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing, 2018: 3-11.

作者简介

邹凯鑫(通信作者),2020 年于南京信息工程大学获得学士学位,现为南京信息工程大学硕士研究生,主要研究方向计算机视觉、机器人感知与控制。

E-mail: 2794827428@ qq. com

Zou Kaixin (Corresponding author) received his B. Sc. degree in 2020 from Nanjing University of Information Science & Technology. Now he is a M. Sc. candidate of Nanjing University of Information Science & Technology. His main research interests include intelligent information processing and application、robot perception and control.

张自嘉,2007 年于上海交通大学获得博士学位,现为南京信息工程大学教授,主要研究方向为智能仪器、嵌入式系统应用、机器人感知与控制。

E-mail: zhzijia@ 126. com

Zhang Zijia received his Ph. D. from Shanghai Jiao Tong University in 2007. He is now a professor at Nanjing University of Information Science and Technology. His main research interests include intelligent instruments, embedded system applications, and robot perception and control.

孙伟,2010 年于东南大学获得博士学位,主要研究方向为交通目标视频检测与环境感知、异常事件视频检测和辨识。

E-mail: sunw0125@ 163. com

Sun Wei received his Ph. D degree from Southeast University in 2010. His research interests include traffic target video detection and environmental perception, and abnormal event video detection and identification.