



计算机科学

COMPUTER SCIENCE

基于Sc-DeepLabV3+模型的铁轨扣件分割方法

黄坤, 何朗, 王展青

引用本文

黄坤, 何朗, 王展青. 基于Sc-DeepLabV3+模型的铁轨扣件分割方法[J]. 计算机科学, 2025, 52(12): 166-174.

HUANG Kun, HE Lang, WANG Zhanqing. [Railway Fastener Segmentation Method Based on Sc-DeepLabV3+ Model](#) [J]. Computer Science, 2025, 52(12): 166-174.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于改进ModernTCN的光伏发电中短期预测](#)

Prediction of Short-and-Medium Term Photovoltaic Power Generation Based on Improved ModernTCN
计算机科学, 2025, 52(11A): 241000164-7. <https://doi.org/10.11896/jsjcx.241000164>

[基于消除语义特征的图像篡改定位模型对抗攻击](#)

Attacking Image Manipulation Localization Model by Eliminating Semantic Features
计算机科学, 2025, 52(11A): 241100104-7. <https://doi.org/10.11896/jsjcx.241100104>

[自动化软件缺陷定位技术研究](#)

Advances in Automatic Software Defect Location Techniques
计算机科学, 2025, 52(11A): 250200024-14. <https://doi.org/10.11896/jsjcx.250200024>

[层次时间序列预测方法与应用综述](#)

Comprehensive Review of Hierarchical Time Series Forecasting Methods and Applications
计算机科学, 2025, 52(11A): 241000139-7. <https://doi.org/10.11896/jsjcx.241000139>

[融合自适应优化与多维聚焦的点云配准网络](#)

Point Cloud Registration Network Integrating Adaptive Optimization and Multi-dimensional Focusing
计算机科学, 2025, 52(11A): 250100019-7. <https://doi.org/10.11896/jsjcx.250100019>

基于 Sc-DeepLabV3+ 模型的铁轨扣件分割方法

黄坤 何朗 王展青

武汉理工大学理学院 武汉 430070

(417136363@qq.com)

摘要 铁轨扣件病害是影响铁路交通安全的重要因素。利用深度学习图像识别方法对铁轨扣件检测机器人所采集的图像进行分割,可以有效提高扣件病害检测的效率。针对目前缺乏公开可用的铁轨扣件数据集,以及扣件数据量大但背景环境复杂导致分割难度大、耗时长等问题,人工制作了 RFS 铁轨扣件数据集并提出基于 Sc-DeepLabV3+ 模型的铁轨扣件分割方法。在原始 DeepLabV3+ 模型的基础上,替换其主干网络为轻量 MobileNetV4 网络以加快运算速度,提出改进的 S-ASPP 模块,使网络能够获得更密集的像素采样,从而增强网络提取细节特征的能力。此外,加入 CSWin 注意力机制并行地计算横向和纵向的注意力,减少复杂背景环境的干扰。实验部分,提出了 RailAugment 数据增强技术,有效增加了数据集的多样性和覆盖度,最终获得的扣件数据集共有 6832 张图像,其中训练集 4782 张,验证集 1366 张,测试集 684 张。实验结果表明,mIoU 和 mPA 分别达到了 95.17% 和 97.14%,相较于原模型提高了 2.19 个百分点和 0.3 个百分点。尽管性能提升幅度较小,但在细节特征提取和背景干扰处理上有明显改善。在公共 DeepGlobe 数据集上验证了 Sc-DeepLabV3+ 模型的鲁棒性和泛化能力,其推理速度较主流 Swin-UNet 模型和 Segmenter 模型快 51.4ms 和 66.5ms,展现了良好的效率与实时性。因此,该模型在铁路维护等领域具有广泛应用潜力,能够有效降低人力和算力成本,提高检测效率。

关键词: 深度学习; 图像语义分割; DeepLabV3+; 铁轨扣件; 数据增强

中图分类号 TP319

Railway Fastener Segmentation Method Based on Sc-DeepLabV3+ Model

HUANG Kun, HE Lang and WANG Zhanqing

School of Science, Wuhan University of Technology, Wuhan 430070, China

Abstract The deterioration of track fasteners is a critical factor affecting railway traffic safety. Utilizing deep learning image recognition methods for segmenting images collected by track fastener detection robots can significantly improve the efficiency of fastener defect detection. This paper addresses the current lack of publicly available datasets for track fasteners and the challenges posed by complex backgrounds that increase segmentation difficulty and processing time. This paper manually creates the RFS (Rail Fastener Segmentation) track fastener dataset and proposes a segmentation method based on the Sc-DeepLabV3+ model. By replacing the backbone network of the original DeepLabV3+ model with the lightweight MobileNetV4, it accelerates computation speed and introduces an improved S-ASPP (Switchable Atrous Spatial Pyramid Pooling) module to enable the network to achieve denser pixel sampling, enhancing its ability to extract detailed features. Additionally, it incorporates the CSWin (Cross-Shaped Window Self-Attention) attention mechanism to compute horizontal and vertical attention in parallel, reducing interference from complex backgrounds. In the experimental section, this paper proposes the RailAugment data augmentation technique to effectively increase the diversity and coverage of the dataset, ultimately resulting in a total of 6832 images, including 4782 for training, 1366 for validation, and 684 for testing. Experimental results show that the mIoU and mPA reach 95.17% and 97.14%, respectively, which represent improvements of 2.19 percentage point and 0.36 percentage point compared to the original model. Although the performance improvement is relatively small, significant improvements are observed in detailed feature extraction and background interference handling. Furthermore, the Sc-DeepLabV3+ model is validated on the DeepGlobe dataset, demonstrating its robustness and generalization ability. Its inference speed is 51.4 ms and 66.5 ms faster than the mainstream Swin-UNet and Segmenter models, respectively, showing good efficiency and real-time performance. Therefore, this model has broad application potential in railway maintenance and other fields, effectively reducing labor and computational costs while improving detection efficiency.

到稿日期:2024-10-22 返修日期:2025-02-08

基金项目:国家自然科学基金青年科学基金(12201475)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China(12201475).

通信作者:何朗(helang@whut.edu.cn)

Keywords Deep learning, Image semantic segmentation, DeepLabV3+, Railway fasteners, Data augmentation

1 引言

扣件是铁路轨道的重要连接零件,对保证铁轨的稳定性、可靠性、安全性有重要作用^[1]。扣件腐蚀、疲劳、磨损、松动以及断裂等是引起铁路交通事故的重要原因。随着我国铁路线路营业里程持续增长,以及列车运行速度大幅度提升,及时而有效地对铁轨扣件进行检测与维护能极大地减少铁路交通事故的发生^[2]。

随着图像处理和计算机视觉技术的不断发展,利用图像处理技术能够实现自动化的轨道扣件巡检,减少人工巡道的不确定性,节省线路运维成本^[3]。因此,铁轨扣件的图像分割技术成为一种有效的伤损检测手段。通过对铁轨扣件图像进行处理和分割,可以显著提高扣件伤损检测效率。

针对铁轨扣件图像分割研究中,传统分割方法高度依赖于人工特征提取设计,Hsieh等^[4]采用小波变换获取扣件的位置区域,通过阈值分割、形态学操作获取扣件的二值图像。Li等^[5]对铁轨扣件图像使用均值滤波,起到了去噪效果,使用Canny算子对铁轨扣件进行边缘检测。以上模型虽然能够在一定程度上自动提取扣件图像,但仍存在泛化性能较差的问题,且只在特定场景下具备较高的准确性,而铁路场景复杂多变,受天气、遮挡、光影、涂装等因素影响。如图1所示,图1(a)中的雪天环境因反射光线而和背景对比减弱,传统分割难以准确识别扣件;图1(b)中隧道内昏暗且有遮挡,导致光照不足和部分可见性问题,使分割效果下降;图1(c)中涂装的扣件颜色多样,增加了复杂性,传统方法可能无法适用;图1(d)中腐蚀的扣件表面纹理变化使得边界模糊,影响分割精度。因此,传统分割方法无法处理背景环境复杂的铁轨扣件分割。



图1 复杂环境下的扣件图像

Fig. 1 Fastener images in complex environments

随着人工智能的发展,基于深度学习的方法推动了自动化扣件巡检技术的进步,弥补了传统分割方法中的不足。Long等^[6]在2015年提出全卷积神经网络(Fully Convolutional Network, FCN)架构,这是首个端对端的像素级图像语义分割模型,使用上采样方式将低分辨率的图像或特征图放大到原始分辨率以恢复图像尺寸,大大提升了图像分割任务的精度和效率。Nash等^[7]在2018年提出了以VGG-16为主

干网络的FCN模型,通过将全连接层替换为卷积层,实现了像素级别的图像分割,并且SegNet在此基础上引入了解码器结构,通过逐步上采样和特征融合,恢复空间分辨率,从而提高了分割精度。上述两种方法虽然在一定程度上提高了分割的精度,但均采用了多次上采样加卷积的操作,在实际应用中会损失细粒度信息,不利于颜色纹理细节繁多的扣件分割。Katsamenis等^[8]在2022年将U-Net模型用于分割金属结构,其独特结构可以兼容不同尺寸输入,然而U-Net缺少对上层信息的融合,易导致扣件位置及边缘信息丢失。Han等^[9]采用线性迭代聚类对图像预分割后,结合FPN在PANet模型上进行训练,在大目标分割任务中表现良好,但模型对于小目标且背景复杂的铁路扣件的细节捕捉存在不足,容易出现细节丢失从而导致误分割。Qian^[10]对比RefineNet^[11], PSPNet^[12], U-Net, DeepLab系列模型在金属表面分割任务上的表现发现:RefineNet设计侧重于全局特征的融合和细化,不适用于金属表面;PSPNet使用多尺度卷积,易模糊扣件、螺栓等细小物体的精细边缘;DeepLab表现出最佳性能,但Rahman等^[13]制作的数据集在使用DeepLabV3+模型训练后,仍面临信息相关性弱、局部细节丢失、耗时长及参数量大等挑战,不适用于铁路线路上资源受限的设备。

尽管现有文献中已经有许多关于金属表面分割的方法,但基于深度学习的铁轨扣件分割仍未得到充分研究。一方面是因为目前缺乏公开可用的铁轨扣件数据集,加之铁轨扣件样本数量少、标注成本高,导致模型的训练和评估面临挑战^[14]。另一方面,铁轨扣件所处的环境背景复杂,包括颜色纹理相近的钢轨及其他轨件、光照变化、阴影和其他杂物等,使得精确分割更加困难。因此,现有方法在实际应用中往往难以实现高精度和鲁棒性。

为了应对上述问题,从上海铁路局、武汉铁路局和浙江宝晟铁路新材料科技有限公司获得了一部分铁轨扣件原始图像,并进行了详细的人工标注,以创建一个高质量的铁轨扣件数据集。为了进一步提升模型的泛化能力,提出了RailAugment数据增强技术,模拟扣件的实际环境,从而扩展了数据集的多样性。

在此基础上,本文提出一种基于Sc-DeepLabV3+模型的铁轨扣件分割方法。采用轻量级MobileNetV4^[15]作为主干特征提取网络,降低了参数量并加快了分割速度。同时,改进ASPP模块,引入可控制函数和多尺度融合,实现更密集的像素采样,增强细节特征提取能力。此外,使用CSWin注意力机制^[16]提升对扣件信息的关注,减少背景干扰,从而提高分割精度。

2 基础 DeepLabV3+ 网络模型

DeepLabV3+主要为Encoder-Decoder结构,其中包含一个编码器模块(Encoder)和一个解码器模块(Decoder)。编码器模块逐步减少特征映射并捕获更高的语义信息,解码器逐步恢复空间信息。Encoder部分主要包括backbone(DCNN)和ASPP两部分。Encoder中连接的第一个模块是DCNN,

它代表的是用于提取图片特征的主干网络;DCNN 模块后是 ASPP 网络,它用一个 1×1 的卷积、3 个 3×3 的空洞卷积和一个全局池化来对主干网络的输出进行处理。之后将结果都连接起来,并用一个 1×1 的卷积来缩减通道数^[17]。

Decoder 部分接收来自 backbone 中间层的低级 feature maps 和来自 ASPP 模块的输出作为输入,利用 ASPP 模块融合深层和浅层信息。不同空洞率的空洞卷积可以获取更多的上下文信息,并且能够让每个卷积输出都包含较大范围的信息,有较好的提取能力。其网络结构如图 2 所示。

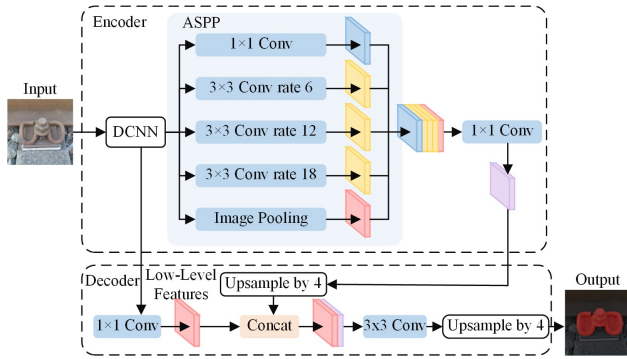


图 2 DeepLabV3+ 模型结构

Fig. 2 Structure of DeepLabV3+ model

但空洞卷积也可能导致局部细节信息丢失和细节信息表征能力不足等问题,特别是在处理模糊边界或精细结构时。

在铁轨扣件的应用场景中,扣件的边缘细节非常微小或与周围环境(如钢轨、其他轨件等)有相似的颜色和纹理特征,这使得空洞卷积可能忽略一些重要的局部细节并且无法有效保持边界信息。此外,DeepLabV3+ 采用了复杂的多尺度信息处理方式,在理论上能够提升性能,但实际运算时,由于需要处理更多的参数和计算量,因此可能训练和推理时间较长。而在资源有限的环境中需要实时检测(如铁路检测设备),长时间的计算可能并不适用,会影响系统的效率和响应速度。

3 Sc-DeepLabV3+ 网络模型

由于铁轨扣件的实际背景复杂环境,易受噪声干扰,因此基础 DeepLabV3+ 模型存在局部信息丢失、细节信息表征能力不足和运算时间长等问题。为了更好地满足实际需求,本文提出了 Sc-DeepLabV3+ 模型,其网络结构如图 3 所示。在 DeepLabV3+ 的基础上,首先采用轻量级的 MobileNetV4 网络作为主干特征提取网络,降低模型的参数量,加快分割速度;其次,对 ASPP 模块进行改进,在金字塔池化操作时加入可控制函数合成结果,再进行特征多尺度融合,使网络在原来的基础上能够获得更密集的像素采样,从而增强网络提取细节特征的能力,加强了特征复用和表达能力。此外,在改进的 ASPP 模块中的每层卷积和浅层特征后加入 CSWin 注意力机制,并行地计算横向和纵向的注意力,减少复杂背景环境的干扰,使模型更关注扣件信息,从而提升分割精度。

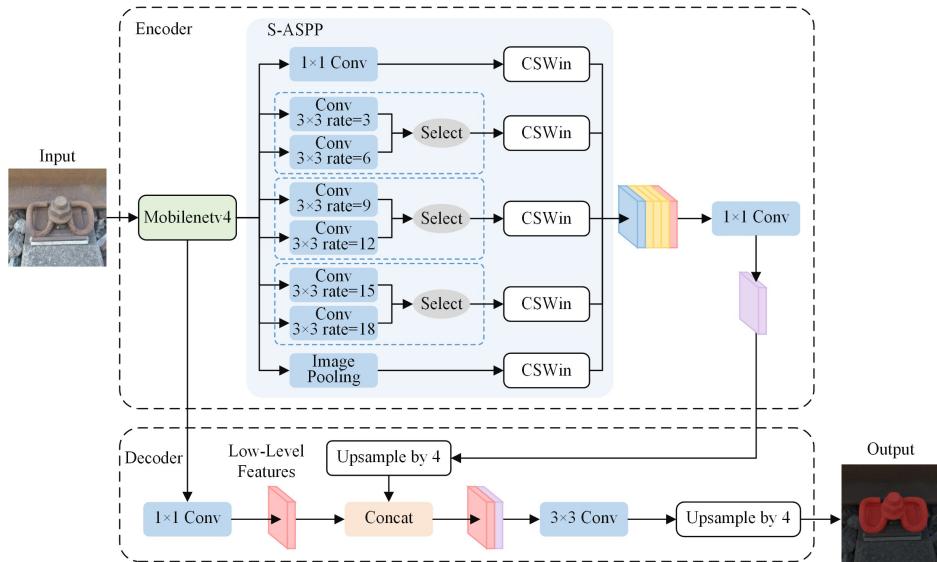


图 3 Sc-DeepLabV3+ 模型结构

Fig. 3 Structure of Sc-DeepLabV3+ model

3.1 轻量化的主干特征网络

铁轨扣件的背景环境复杂,但其扣件形状较为固定,若原模型中的 DCNN 对于扣件分割任务过于繁琐,易导致局部信息丢失、运算时间长和细节信息表征能力不足等问题。本文采用轻量级的 MobileNetV4 结构作为主干特征提取网络,MobileNet V4(MNv4)的核心在于通用倒置瓶颈结构(Universal Inverted Bottleneck, UIB)(见图 4)和 Mobile MQA 两大模块,搭配优化后的神经网络架构搜索(Neural Architecture Search, NAS)配方。在原有 MobileNet 的经典组件中,如可分离的深度卷积和逐点扩展及投影倒置瓶颈块,引入了 UIB。

在倒置瓶颈块中引入了两个可选的深度卷积,一个位于扩展层之前,另一个位于扩展层和投影层之间。这两个深度卷积的“保留与否”被纳入神经架构搜索的决策过程,最终会生成更优的架构。本文方案巧妙地整合了多个重要模块,以构建一个高效的模型架构。首先,原始的倒置颈结构与 ConvNext 及 Vision Transformer(ViT)中的前馈网络相结合,使得特征提取过程更加灵活和强大。其次,引入了一种新变体——ExtraDW,该模块通过增强深度可分离卷积的特性,进一步提升了模型的表现力和计算效率。此外,设计了一种专为加速器优化的 Mobile MQA 注意力块,该模块实现了高达 39%

的推理加速,大幅提升了处理速度。最后,为了优化模型性能,提出了一种改进的神经网络架构搜索(NAS)方法,显著提高了 MNv4 搜索的有效性,确保了最佳结构的选择和应用。

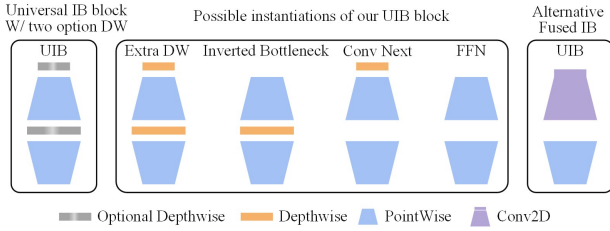


图 4 通用倒置颈结构

Fig. 4 Universal Inverted Bottleneck blocks

3.2 ASPP 模块优化

DeepLabV3+模型中的 ASPP 模块利用空洞卷积来获取更大的感受野,但是对于铁轨扣件这种背景环境复杂的提取任务,容易出现细节特征提取不足、特征丢失等现象,从而造成漏分、误分等问题。因此,本文对 ASPP 模块进行改进,提出一种可控制的的空洞空间金字塔池化(Switchable Atrous Spatial Pyramid Pooling, S-ASPP)模块,使用控制函数来组合不同空洞率卷积的结果(见图 5),再进行特征的多尺度融合(见图 6)。

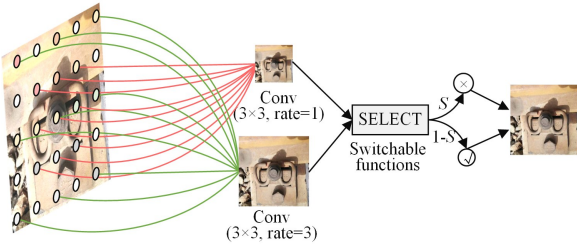


图 5 可控制函数示意图

Fig. 5 Schematic diagram of switchable functions

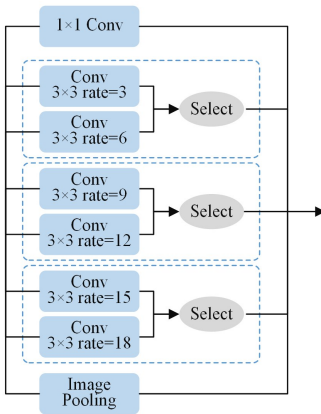


图 6 S-ASPP 结构示意图

Fig. 6 Structural diagram of S-ASPP

对于钢轨扣件的复杂背景环境来讲,通过这种“双重观察并结合”的策略,能够有效地处理复杂的特征模式。这种方法不仅提高了特征提取的灵活性和适应性,而且还提升了扣件检测和分割任务中的准确性和效率。

3.3 注意力模块 CSWin

铁轨扣件背景环境受到多种因素的影响,如光线的变化、

角度的不同、涂装状态、灰尘覆盖以及腐蚀程度等。而且,不同地区的背景也各有不同,如隧道、道岔等环境,以及雨雪、雾霾、冻雨等恶劣天气都会对扣件的状态产生影响。这些对模型的分割任务带来了巨大的挑战。CSWin (Cross-Shaped Window Self-attention)是一种十字形状的自注意力机制模块,能够并行地计算横向和纵向的注意力,如图 7 所示。对于大量的钢轨扣件图像检测来讲,全局自注意的计算成本非常高,而局部自注意通常会限制每个词向量的交互领域。为了解决这个问题,使用 Cross-Shaped Window 自注意力机制,计算水平和垂直平行条纹的自注意,这些条纹形成十字形窗口,每个条纹通过将输入特征分割成等宽的条纹获得。

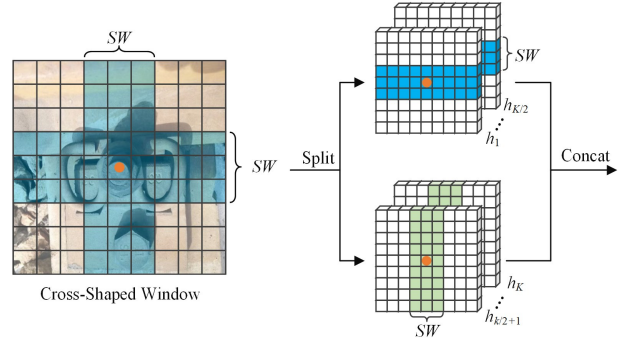


图 7 十字形状的自注意力机制模块结构

Fig. 7 Structure of CSwin selfattention module

横向条纹自注意力的计算方式为:

$$X = [X^1, X^2, \dots, X^M]$$

$$Y_k^i = \text{Attention}(X^1, X^2, \dots, X^M) \quad (1)$$

$$H\text{-Attention}_k(X) = [Y_k^1, Y_k^2, \dots, Y_k^M]$$

纵向自注意力 V-Attention 和 H-Attention 的计算方式类似,不同的是它取的是宽度为 SW 的竖条。

最终,这个 block 的输出表示为:

$$\text{CSWin-Attention}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_K)W^O$$

$$\text{head}_k = \begin{cases} H\text{-Attention}_k(X), & k = 1, 2, \dots, K/2 \\ V\text{-Attention}_k(X), & k = K/2 + 1, \dots, K \end{cases} \quad (2)$$

将 CSWin 加入 ASPP 模块的每层卷积后,可以充分利用 CSWin 提取到不同尺度的高阶特征,进一步增强模型对于多尺度信息的感知能力。

在 ASPP 模块中,将 CSWin 模块添加到每个卷积操作之后,以获得多尺度的上下文信息。通过在每个卷积操作后引入 CSWin 模块,模型能够在不同尺度上有效地捕获局部和全局的语义信息,从而提高模型在复杂场景下的分割性能。此外,CSWin 注意力机制可以帮助网络更好地理解铁轨扣件图像中的关键特征,使网络能够更准确地分割出扣件,在不同光照条件、角度、涂装状态等背景环境下,减少由背景干扰或恶劣环境条件引起的误分割。考虑到不同地区和不同环境的差异,CSWin 注意力机制可以使网络更加鲁棒,能够适应各种环境条件下的铁轨扣件分割任务,包括隧道、道岔以及各种恶劣天气条件。

3.4 模型创新点

Sc-DeepLabV3+在语义分割领域的创新性主要体现在模块整合、细节特征提取和环境适应性 3 个方面。首先,基于

MobileNetV4 的深度可分离卷积设计,使模型在计算量和特征捕捉上达到平衡,适合铁路场景中资源受限的设备。其次, S-ASPP 模块通过自适应调整空洞率,有效提升了多尺度信息的捕捉能力,尤其在物体边缘和细节上表现优异。此外, CSWin 注意力机制的引入使得模型能够在复杂环境中优先聚焦重要特征,减少背景干扰,从而增强分割效果。总的来说, Sc-DeepLabV3+ 模型能显著提高铁轨扣件分割的精度和效率,解决由于背景环境相似、钢轨扣件腐蚀特征复杂,无法准确分割出扣件区域的问题,从而为后续的疾病检测和养护决策提供可靠的基础,具有一定的工程应用价值和场景创新性。

4 实验环境及评价指标

4.1 实验环境

本实验在 Windows 操作系统下搭建的 PyTorch 深度学习框架中进行,其软硬件详细环境配置如表 1 所列。

表 1 实验环境配置

Table 1 Experimental environment configuration

实验环境	配置说明
硬件环境	CPU: Intel(R) Core(TM) i5-14600KF 3.49 GHz GPU: NVIDIA GeForce RTX 4070 Ti SUPER(16 GB) 内存: 32 GB
操作系统	Windows 操作系统
开发环境	Python 3.8
开发框架	PyTorch 深度学习框架

训练过程采用了如下参数设置:学习率为 5×10^{-4} , 每批次输入大小为 8, 训练迭代次数为 100, 使用 cos 学习率下降方式。选择 Adam 优化器, 相比于其他优化器, Adam 优化器能够同时考虑一阶动量和二阶动量, 并动态调整每个参数的学习率, 从而使模型更快收敛至性能最优。

4.2 实验数据集

铁轨扣件图像由宝晟科技公司装载高清摄像头的铁轨作业机器人拍摄, 上海铁路局和武汉铁路局的工务部门提供, 共计 2 208 张图像, 其中包含不同光线、角度、地区(隧道、道岔等)、天气(雨天、晴天、雪天、冻雨等)的扣件图像。使用 labeme 软件人工标注铁轨图像中的扣件区域, 制作成模型训练所需数据集。为了适应铁轨图像分析领域的特点, 并结合铁路轨道的实际场景, 提出了 RailAugment 数据增强方法。首先, 从常见的数据增强操作开始; 然后根据铁轨领域的特点进行筛选, 排除不适合扣件图像的操作, 如反转、均衡化和反转等操作, 这些操作可能会破坏扣件图像中的细节和特征; 之后将数据增强操作分为像素级和空间级操作, 并构建了两个增强空间, 即像素增强空间 S_p (Pixel Augmentation Space) 和几何增强空间 S_g (Geometric Augmentation Space)。其中 $S_p = \{ \text{gaussian blur}, \text{gaussian noise}, \text{sharpness}, \text{color transformations}, \text{brightness}, \text{contrast}, \text{posterize} \}$, $S_g = \{ \text{translate}_x, \text{translate}_y, \text{shear}_x, \text{shear}_y, \text{Scale}, \text{rotate}, \text{horizontal flip}, \text{vertical flip} \}$ 。

随机采样每个分支的 M 个数据增强操作, 本文给定 $M = \{2, 3\}$, 生成了 4 种从 S_p 和 S_g 中采样的组合方式, 分别是 $1+2, 0+3, 1+1$ 和 $0+2$ 。RailAugment 数据增强方法模拟

了隧道内外、自然光暴晒、阴影遮挡、特殊天气、腐蚀等实际场景(见图 8), 再进行人工筛选, 获得数据集包含 6 832 张图像。按照 72:1 的比例随机划分数据集, 得到 4 782 张训练图像、1 366 张验证图像和 684 张测试集图像, 形成用于实验的铁轨扣件图像分割(Rail Fastener image Segmentation, RFS)数据集。

为了确保本文模型的泛化性和鲁棒性, 除了在 RFS 数据集上进行测试外, 还引入了公共 DeepGlobe 数据集进行泛化性实验。DeepGlobe 数据集涵盖了多种遥感图像和场景, 其中包含 6 226 张大小为 1024×1024 像素的带标签的图像, 空间分辨率为 0.5 m。按照 7:2:1 的比例随机划分数据集, 得到 4 314 张训练图像、1 289 张验证图像和 623 张测试集图像。DeepGlobe 数据集具有不同的地理区域和复杂的背景变化, 可以更有效地评估 Sc-DeepLabV3+ 模型在多样化数据环境下的综合表现。

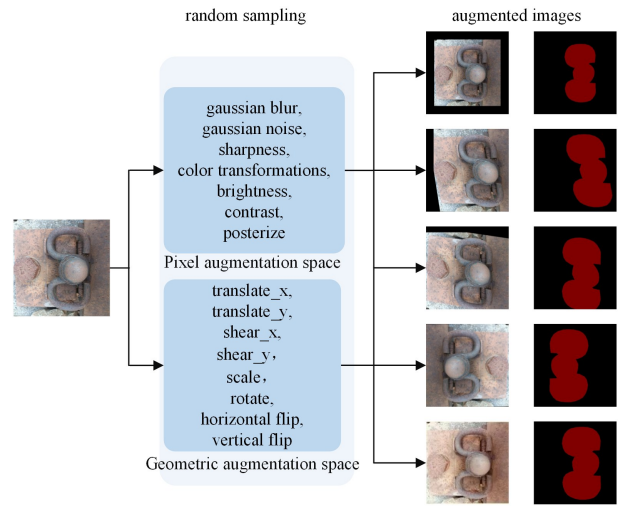


图 8 RailAugment 数据增强方法效果图

Fig. 8 Effect diagram of RailAugment data augmentation method

4.3 评价指标

在图像语义分割领域, 常用的评价指标包括准确率(Accuracy)、平均像素精度(mPA)、平均交并比(mIoU)等。其计算式分别为:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

其中, TP 代表被正确识别的扣件的像素点数量, TN 代表被正确识别的背景的像素点数量, FP 代表被错误识别为扣件的背景像素点数量, FN 代表被错误识别为背景的扣件像素点数量。

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (4)$$

其中, $k+1$ 为类别数, p_{ii} 为正确分类的像素数, p_{ij} 为 i 类被预测为 j 类的像素数, p_{ji} 为 j 类被预测为 i 类的像素数。PA 表示计算每个类中正确分类的像素数量与像素总数之间的比率, mPA 则表示计算每个类的 PA, 然后再累加求平均。

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (5)$$

5 实验结果与分析

5.1 主干网络替换对比实验

为了验证 MobileNetV4 网络的有效性,在 RFS 数据集上,选用不同网络作为 DeepLabV3+模型的主干特征提取网络进行对比实验。如表 2 所列,MobileNetV4 的 mPA 为 96.35%,仅比 Xception 低 0.43%,但高于 ResNet101 和 VGG16 的 7.88%和 12.05%。MobileNetV4 的 Accuracy 和 mIoU 分别为 95.11%和 91.50%,略低于 Xception 的 95.67%和 92.36%,同样远高于 ResNet101 和 VGG16。在参数量上,MobileNetV4 的参数量仅为 1.392×10^7 ,显著低于其他网络,比 VGG16 少 1.659×10^7 ,远低于 ResNet101 和 Xception。在推理时间上 MobileNetV4 相比 ResNet101, Xception 和 MobileNetV2 的速度更快,最明显的是比 Xception 快了 21.56 ms。MobileNetV4 在精度、参数量和推理时间之间实现了良好的平衡,精度表现基本接近 Xception,同时具有远低于这些网络的参数量和推理时间,使其成为资源受限环境中的理想选择。特别是在铁路场景中,需要实时响应且对存储和计算资源有严格要求,MobileNetV4 具有一定的优势。

表 2 不同主干网络对比结果

Table 2 Comparison results of different backbone networks

主干网络	mPA/%	Accuracy/%	mIoU/%	参数量	推理时间/ms
VGG16	84.30	87.34	85.93	3.051×10^7	16.22
ResNet101	88.47	90.71	88.49	2.949×10^7	20.93
Xception	96.78	95.67	92.36	2.537×10^7	37.28
MobileNetV2	95.94	94.20	90.79	1.643×10^7	17.34
MobileNetV4	96.35	95.11	91.50	1.392×10^7	15.72

5.2 可控制的空洞空间金字塔池化对比实验

可控制的空洞空间金字塔池化(S-ASPP)是将提取到的高阶特征经过不同空洞率的空洞卷积后,再使用控制函数选择更佳的结果。在铁轨扣件分割任务中,需要根据扣件具体情况选择合适的空洞率,以得到最佳分割结果。将原 ASPP 中的(6,12,18)作为一组空洞卷积,为了确定备选的空洞卷积的空洞率,设置了以下空洞率组合来进行对比实验,实验使用 RFS 数据集作为训练数据集,使用 MobileNetV4 作为主干网络。

由表 3 可得,(3,9,15)的空洞率组合表现最佳,mPA 达到 96.71%,Accuracy 为 97.07%,mIoU 为 93.74%,因此将其作为 S-ASPP 中备选的空洞卷积率。

表 3 S-ASPP 模块性能对比实验

Table 3 Performance comparison experiment of S-ASPP module

序号	卷积率	mPA/%	Accuracy/%	mIoU/%
1	(2,5,7)	95.53	96.49	92.62
2	(3,9,15)	96.71	97.07	93.74
3	(4,16,24)	95.93	96.71	92.43
4	(5,15,25)	95.64	96.10	92.13

5.3 模块优化对比实验

使用 RFS 数据集作为训练数据集,使用 MobileNetV4 主干网络,对 S-ASPP 模块、CSWin 注意力机制,以及 S-ASPP 模块和 CSWin 注意力机制的结合进行分析对比,如表 4 所列。

表 4 改进模块的性能对比结果

Table 4 Performance comparison results of improved modules

	S-ASPP	CSWin	mPA/%	Accuracy/%	mIoU/%
			96.14	95.32	91.57
✓			96.71	97.07	93.74
		✓	96.43	96.79	93.42
✓	✓	✓	97.14	97.86	95.17

根据表 4 的结果,S-ASPP 模块的引入提升了分割任务中小物体和细节部分的处理能力,尤其在复杂背景下,其多尺度特征提取能力提升了分割精度。CSWin 注意力机制增强了模型对目标物体的注意力集中能力,抑制了背景噪声的影响,虽然其单独效果不如 S-ASPP,但对于提升模型整体的背景抑制和目标区分能力有积极作用。S-ASPP 和 CSWin 的结合使用实现了最佳分割性能。在铁轨扣件分割任务中,二者的互补性发挥了重要作用,提升了模型的整体性能,达到了最佳效果。

5.4 语义分割模型对比实验

在同一环境下,将 U-Net, SegNet, DenseNet, PSPNet 和 DeepLabV3+ 这 5 种经典语义分割模型以及近期提出的 Swin-UNet^[18]和 Segmenter^[19]模型,与本文 Sc-DeepLabV3+模型在 RFS 数据集上进行训练,所有实验均在相同的实验环境和数据集上进行。得到的不同模型的分割精度评价指标以及参数量如表 5 所列。

表 5 RFS 数据集上不同模型的性能对比结果

Table 5 Performance comparison results of different models on

RFS dataset

Algorithm	mPA/%	Accuracy/%	mIoU/%	参数量
U-Net	85.62	95.22	82.04	7.850×10^6
SegNet	86.62	95.84	83.39	2.736×10^7
DenseNet	91.91	96.47	89.02	2.868×10^7
PSPNet	96.21	94.18	90.97	4.671×10^7
DeepLabV3+	96.78	95.67	92.36	5.471×10^7
Swin-UNet	97.02	97.27	94.08	5.965×10^7
Segmenter	97.11	97.50	93.80	5.350×10^7
Sc-DeepLabV3+	97.14	97.86	95.17	3.048×10^7

从表 5 中可以看出,Sc-DeepLabV3+的 mIoU 达到了 95.17%,mPA 为 97.14%,比基础的 DeepLabV3+提高了 2.81个百分点和 0.36 个百分点,说明在 DeepLabV3+的基础上创新性地引入 S-ASPP 和 CSWin 起到了明显的效果,提升了对扣件细节和边缘的准确分割能力。与其他模型相比,Sc-DeepLabV3+的精度表现同样突出,其 mPA, Accuracy 和 mIoU 3 个指标皆与 Swin-UNet 和 Segmenter 接近,表明该模型在现阶段具有一定竞争力。Sc-DeepLabV3+的参数量为 3.048×10^7 ,显著低于基础 DeepLabV3+,Swin-UNet 和 Segmenter,这是因为 Sc-DeepLabV3+采用的是轻量级 MobileNetV4 主干网络,所以该模型在计算资源有限的环境下,能够高效执行任务,降低了人力和算力成本,且进一步加快了计算速度并减少了内存消耗,使其在铁路场景移动端应用成为可能。

从图 9 的分割效果图可以看出,不同模型在处理 RFS 数据集时表现各异。U-Net 和 SegNet 在处理铁轨扣件与背景相似的场景时,出现了明显的漏分现象,原因在于这两个网络在面对复杂背景时,未能有效提取细节特征,导致扣件区域被误判为背景。具体来说,U-Net 和 SegNet 的特征提取能力较

为局限,无法充分处理扣件区域和背景的相似性,进而导致分割不完整。相比之下, DenseNet, PSPNet, DeepLabV3+, Swin-UNet 和 Segmenter 虽然在分割精度上有所提高,但依然存在一定的漏分和误分现象,尤其在边缘细节的处理上较为模糊。这是由于这些模型虽然引入了多尺度信息和更深的网络结构,但在面对铁轨扣件与复杂背景的分隔时,仍然无法有效消除背景噪声的干扰。Sc-DeepLabV3+ 通过引入改进的 S-ASPP 模块,能够实现更密集的像素采样,捕捉到更精细

的局部特征;同时,CSWin 注意力机制通过并行计算横向和纵向的注意力,显著减少了背景的干扰,确保了分割结果的准确性。

综上所述,Sc-DeepLabV3+ 在铁轨扣件分割任务中,不仅有效降低了漏分和误分问题,还显著提升了分割精度和细节提取能力,尤其在复杂背景下具有明显的优势。这些结果在图 10 的直方图中得到了直观体现,充分显示了 Sc-DeepLabV3+ 在复杂场景中的细粒度分割能力。

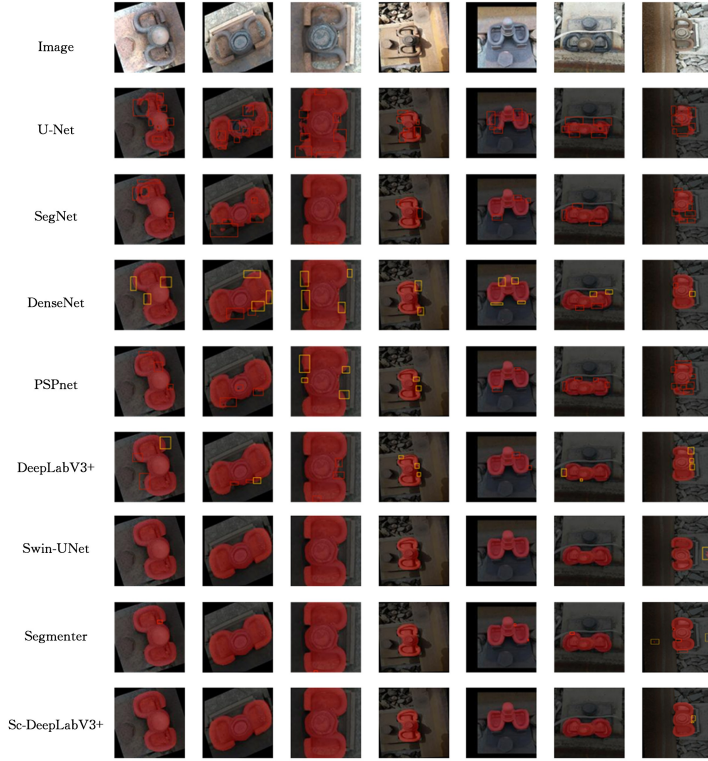


图 9 RFS 数据集上不同模型分割效果

Fig. 9 Segmentation effect of different models on RFS dataset

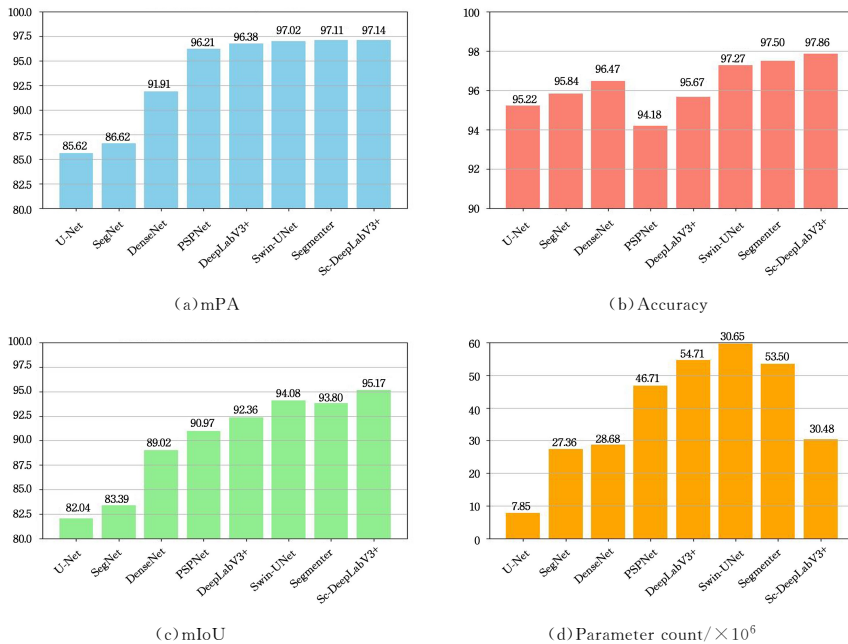


图 10 不同分割模型的性能比较直方图

Fig. 10 Performance comparison histogram of different segmentation models

5.5 数据增强前后模型性能对比实验

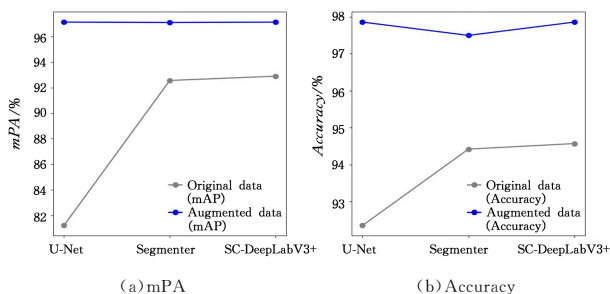
RFS 原始数据集包含了铁路场景下的未处理图像,背景复杂性、环境条件变化或拍摄角度不同等因素,导致图像中的细节特征提取受到限制。表 6 中的数据显 示,原始数据集上的各个模型性能相对较低,尤其是在 mPA 和 mIoU 的指标上。通过采用 RailAugment 数据增强技术,数据集的多样性和覆盖度得到了有效提升,表 5 中 Sc-DeepLabV3+ 在 mPA 和 mIoU 上分别达到了 97.14% 和 95.17%,相较于表 6 中使用原始数据的模型提升了 4.24 个百分点和 6.19 个百分点。

表 6 RFS 原始数据集上不同模型性能的对比结果

Table 6 Comparison results of the performance of different models on the original RFS dataset

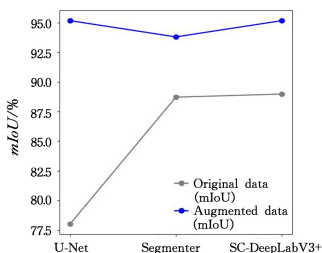
Algorithm	mPA/%	Accuracy/%	mIoU/%
U-Net	81.20	92.35	78.01
SegNet	82.35	93.12	79.43
DenseNet	87.45	94.20	84.13
PSPNet	90.15	92.75	85.42
DeepLabV3+	91.87	93.35	87.1
Swin-UNet	92.05	94.12	88.12
Segmenter	92.57	94.42	88.72
Sc-DeepLabV3+	92.90	94.57	88.98

由图 11 可以得出,RailAugment 数据增强在提升模型性能、增强其泛化能力和适应铁路复杂环境方面具有显著优势。本研究在数据增强技术和改进的模型架构的共同作用下,Sc-DeepLabV3+ 模型在铁轨扣件分割任务中的表现得到了显著提升,使其具备了更强的实用性和应用潜力。



(a) mPA

(b) Accuracy



(c) mIoU

图 11 数据增强前后模型性能对比

Fig. 11 Model performance comparison before and after data augmentation

5.6 不同数据集上的泛化性实验

为了进一步验证所提模型的鲁棒性和泛化性,将 Sc-DeepLabV3+ 模型应用于 DeepGlobe 数据集进行实验。

实验结果如表 7 所列,Sc-DeepLabV3+ 模型在精度上超越了基础的 DeepLabV3+,其中 mPA 提升了 0.4 个百分点,mIoU 提升了 1.28 个百分点,但 mPA 略低于 Swin-UNet 等

主流方法。而 Sc-DeepLabV3+ 在轻量化方面具有显著优势,其推理时间比 Swin-UNet 快 51.4ms,比 Segmenter 快 66.5ms,展现了良好的效率与实时性。得益于 MobileNetV4,S-ASPP 模块和 CSWin 注意力机制的创新设计,Sc-DeepLabV3+ 在 DeepGlobe 数据集上的表现依然优异,验证了其在不同领域和场景下的鲁棒性以及良好的泛化能力。

表 7 在 DeepGlobe 数据集上不同模型性能的对比结果

Table 7 Comparison results of the performance of different models on DeepGlobe dataset

Algorithm	mPA/%	Accuracy/%	mIoU/%	推理时间/ms
U-Net	85.23	96.43	76.81	63.3
SegNet	83.76	95.98	74.89	156.5
DenseNet	87.45	97.21	77.88	103.5
PSPNet	87.01	97.68	79.33	90.8
DeepLabV3+	88.81	97.81	78.77	104.4
Swin-UNet	89.35	97.92	80.62	95.2
Segmenter	88.92	97.85	79.20	110.3
Sc-DeepLabV3+	89.21	98.19	80.05	43.8

结束语 本文提出了一种基于改进 Sc-DeepLabV3+ 模型的铁轨扣件分割方法,并通过制作铁轨扣件数据集 RFS 以及引入 RailAugment 数据增强技术,显著提升了分割模型在复杂背景环境中的表现。实验结果表明,采用轻量化的 MobileNetV4 网络和改进的 S-ASPP 模块,结合 CSWin 注意力机制,能够有效提高扣件病害检测的准确性,并在细节特征提取和背景干扰处理方面有明显改善。此外,Sc-DeepLabV3+ 模型在公共 DeepGlobe 数据集上的验证进一步证明了其良好的鲁棒性与泛化能力,同时展现了出色的推理速度和实时性。

所提方法虽然取得了一些成果,但仍面临一些挑战。首先,模型在复杂场景下的适应能力仍有待提高,特别是在恶劣天气和隧道内缺少光照的条件下的检测效果。其次,尽管增强了数据集,但不同环境下的扣件病害样本仍较为稀缺,数据集的多样性有待进一步丰富。未来的研究将重点探索更高效的特征融合方法、数据增强策略,以及更多先进的注意力机制,以进一步提升模型在复杂场景下的鲁棒性和泛化性。

参 考 文 献

- [1] GAO S G, WANG A B, XIAO J, et al. Research on Damage and Structural Optimization Analysis of Spring Bars in High Speed Railway Fasteners [J]. Noise and Vibration Control, 2018, 38(6):190-193.
- [2] LU Z W. Overview of high-speed railway track technology[J]. Journal of Railway Engineering, 2007(1):41-54.
- [3] WEI X K, SUO D, WEI D H, et al. Overview of the Application of Machine Vision in State Detection of Rail Transit Systems [J]. Control and Decision Making, 2021, 36(2):257-282.
- [4] HSIEH H Y, CHEN N, LIAO C L. Visual recognition system of elastic rail clips for mass rapid transit systems [C]//IEEE Joint Rail Conference & Internal Combustion Engine Division Spring Technical Conference. IEEE, 2007:234-237.
- [5] LI J L, MA H F, ZHANG W Z, et al. Research on edge detection of rail fasteners based on machine vision [J]. Journal of Northwest Normal University: Natural Science Edition, 2017, 53(5):45-48.

- [6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015;3431-3440.
- [7] NASH W, DRUMMOND T, BIRBILIS N. Quantity beats quality for semantic segmentation of corrosion in images [J]. arXiv: 1807.03138, 2018.
- [8] KATSAMENIS I, DOULAMIS N, DOULAMIS A, et al. Simultaneous precise localization and classification of metal rust defects for robotic-driven maintenance and prefabrication using residual attention U-Net [J]. Automation in Construction, 2022, 137:104182.
- [9] HAN Q, ZHAO N, XU J. Recognition and location of steel structure surface corrosion based on unmanned aerial vehicle images [J]. Journal of Civil Structural Health Monitoring, 2021, 11(5):1375-1392.
- [10] QIAN C. Evaluation of deep learning-based semantic segmentation approaches for autonomous corrosion detection on metallic surfaces [D]. West Lafayette: Purdue University, 2019.
- [11] LIN G, MILAN A, SHEN C, et al. Refinenet: Multipath refinement networks for high-resolution semantic segmentation [C] // Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017. 5168-5177.
- [12] 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.
- [13] RAHMAN A, WU Z Y, KALFARISI R. Semantic deep learning integrated with RGBfeature-based rule optimization for facility surface corrosion detection and evaluation [J]. Journal of Computing in Civil Engineering, 2021, 35(6):04021018.
- [14] ZHANG H, LIAO D, LI C. Rail Surface Defect Detection Model Based on Attention Module and Hybrid-supervised Learning [EB/OL]. <https://www.jsjx.com/CN/article/openArticlePDF.jsp?id=21188>.
- [15] QIN D F, LEICHTNER C, DELAKIS M, et al. MobileNetV4—Universal Models for the Mobile [C] // Computer Vision-ECCV 2024; 18th European Conference. 2024;78-96.
- [16] DONG X Y, BAO J M, CHEN D D, et al. CSWin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows [J]. arXiv:2107.00652, 2021.
- [17] WANG Q, HE L, WANG Z Q, et al. Road Extraction Algorithm for Remote Sensing Images Based on Improved DeepLabv3+ [J]. Computer Science, 2024, 51(8):168-175.
- [18] CAO Z, YU L, XU D, et al. Swin-Unet: Unifying self-attention and convolutional networks for medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2021, 41(5):1382-1392.
- [19] STRUDEL R S, CARION X, CORD I, et al. Segformer: Transformer-based Semantic Segmentation [J]. Theoretical Computer Science, 2021, 730(19):1-20.



HUANG Kun, born in 2000, postgraduate. His main research interest is image processing.



HE Lang, born in 1974, professor, Ph.D. His main research interests include intelligent calculation and image processing.

(责任编辑:何杨)