

文章编号: 1005-0523(2025)01-0052-09



基于双模态融合的钢轨表面缺陷分割研究

罗 晖, 韩岳霖, 马治伟, 斯成浩

(华东交通大学信息与软件工程学院, 江西 南昌 330013)

摘要: 因长期受反复荷载作用, 高速铁路钢轨会产生表面缺陷, 为了提升复杂场景下多类多尺度钢轨表面缺陷检测的精度与速度, 设计了一种基于双模态融合的钢轨表面缺陷分割网络(DAFNet)。首先构建了一个包含可见光和红外通道的钢轨表面缺陷数据集, 并采用改进的双分支网络架构, 提高了分割速度; 同时, 设计了双模态自适应融合模块(BAFM), 实现了特征的自适应融合, 提高了复杂场景下钢轨表面缺陷的分割精度; 此外, 设计了空间细节提取模块(SDEM)和关键信息增强模块(KIEM), 进一步提高了对缺陷边缘的感知度, 解决了复杂场景下缺陷与背景对比度不高的问题。实验表明, 所设计网络分割的精确度和mIoU分别达到了68.13%, 59.96%, 明显优于其他主流网络; 且FLOPs、参数量和模型大小分别为17.41 GFLOPs, 1.38 M和5.67 MB, 优于大多数主流网络。所设计的网络显著提高了钢轨表面缺陷的分割精度, 并且具有较高分割速度, 对保障高铁的安全运营具有重要意义。

关键词: 语义分割; 钢轨表面缺陷; 深度学习; 红外图像; 可见光图像; 双模态融合

中图分类号: U213; TP39

文献标志码: A

本文引用格式: 罗晖, 韩岳霖, 马治伟, 等. 基于双模态融合的钢轨表面缺陷分割研究[J]. 华东交通大学学报, 2025, 42(1): 52-60.

Research on Rail Surface Defect Segmentation Based on Bimodal Fusion

Luo Hui, Han Yuelin, Ma Zhiwei, Si Chenghao

(School of Information and Software Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: Due to the long-term repeated loading, surface defects occur in high-speed railway steel rails. In order to improve the accuracy and speed of surface defect detection for multiple classes and scales of steel rails in complex scenarios, a steel rail surface defect segmentation network based on multimodal fusion (DAFNet) is designed. Firstly, a steel rail surface defect dataset containing visible light and infrared channels is constructed, and an improved dual-branch network architecture is adopted to increase segmentation speed. Simultaneously, a bimodal adaptive fusion module (BAFM) is designed to achieve adaptive feature fusion, improving the segmentation accuracy of steel rail surface defects in complex scenarios. Additionally, a spatial detail extraction module (SDEM) and a key information enhancement module (KIEM) are designed to further enhance the perception of defect edges and address the low contrast between defects and backgrounds in complex scenarios. Experiments show that the accuracy and mIoU of the designed network segmentation reach 68.13% and 59.96% respectively, which are significantly better than other mainstream networks. Moreover, FLOPs, parameter quantity, and model

收稿日期: 2024-04-30

基金项目: 国家自然科学基金项目(62262021)

size are 17.41 GFLOPs, 1.38 M, and 5.67 MB respectively, which are better than most mainstream networks. The designed network significantly improves the segmentation accuracy of steel rail surface defects and has a high segmentation speed, which is of great significance for ensuring the safe operation of high-speed railways.

Key words: semantic segmentation; rail surface defects; deep learning; infrared image; visible light image; bimodal fusion

Citation format: LUO H, HAN Y L, MA Z W, et al. Research on rail surface defect segmentation based on bimodal fusion[J]. Journal of East China Jiaotong University, 2025, 42(1): 52–60.

在长期的重复荷载影响下,钢轨的健康状况和性能会逐渐下降,同时在其表面形成多种类型和尺度的缺陷。当前钢轨表面缺陷检测主要依赖非破坏性检测技术^[1]和传统图像处理方法^[2-3],这些方法效率低,且存在误检漏检的情况,难以确保准确性和实时性。随着深度学习技术发展,目标检测方法被分为图像级、区域级和像素级3种。图像级方法精度较低,无法精确定位伤损;区域级方法虽能定位,但不足以进行精细分割;相比之下,作为像素级检测方法的语义分割方法可以精确获取目标的类别特征和位置信息,是钢轨表面缺陷检测最有效的方法,对保障高铁的安全运营具有重要意义。

目前,主流的语义分割网络主要包括单模态网络和双模态网络。单模态网络大多利用可见光(RGB)单模态信息,同时使用深度卷积神经网络(CNN)进行分割。Long等^[4]提出了一种将传统的全连接层替换为全卷积层的网络(FCN),并采用不同深度特征提取层的跳级连接,确保了分割的精确性。Yu等^[5]提出了由平滑网络和边界网络组成的判别特征网络(DFN),通过两个协同的子网络来加强类别间的分割性能。在上述单模态网络中,虽然RGB图像含有丰富的细节信息,但易受环境影响。而红外图像(NIR)具有不受外部光照影响、抗干扰能力强等特点,可与RGB图像进行互补,实现对目标更完整、更精确的分割。因此,Ha等^[6]提出了一种双模态特征提取器的网络架构MFNet,并实现了RGB和红外特征融合。Sun等^[7]提出RTFNet,采用编码器-解码器架构来提取和重建双模态特征,同时引入了Upception模块加强特征提取,提高了分割精度。Zhou等^[8]提出了GMNet,其利用分级特征提取策略,设计了针对浅层特征和深层特征的融合

模块,实现了多尺度特征融合。Deng等^[9]提出的FEANet通过FEAM模块从通道和空间视角挖掘和增强多级特征,并设计了边界细化增强模块,可以获取目标边缘信息,提高了分割效果。但在钢轨表面缺陷分割研究中,现有双模态网络存在着复杂场景下背景对比度不高、缺陷边缘的感知度弱、精度和速度低等问题。

综合当前研究现状,本文从基于双模态的编解码网络构建、自适应特征融合、空间细节信息提取及关键信息增强等维度,设计了一种准确、快速的双模态融合的钢轨表面缺陷分割网络(DAFNet)。本文首先构建了一个包含RGB和NIR通道的钢轨表面缺陷分割数据集,并通过数据增强技术^[10]进行了数据集扩充;为充分利用可见光和红外图像特征的互补性,借助改进的双分支架构,设计了DAFNet双模态语义分割网络;结合空间注意力、通道注意力和门控机制,设计了双模态自适应融合模块(BAFM),实现了关键特征的自适应融合;设计了空间细节提取模块(SDEM)和关键信息增强模块(KIEM),提高了分割网络对缺陷边缘的感知度,解决了复杂场景下缺陷与背景对比度不高的问题。

1 DAFNet整体架构

本文所提DAFNet网络采用编码器-解码器架构,具体如图1所示。编码器由两个主干分支组成,一个用于提取RGB特征,另一个用于提取NIR特征。同时,为了自适应地融合双模态关键信息,增强边缘特征,利用注意力机制和门控机制,设计了BAFM模块,该模块可以融合两个分支提取的RGB和NIR特征信息。此外,为了防止下采样丢失空间细节信息,并获取丰富的缺陷边缘信息,设计了

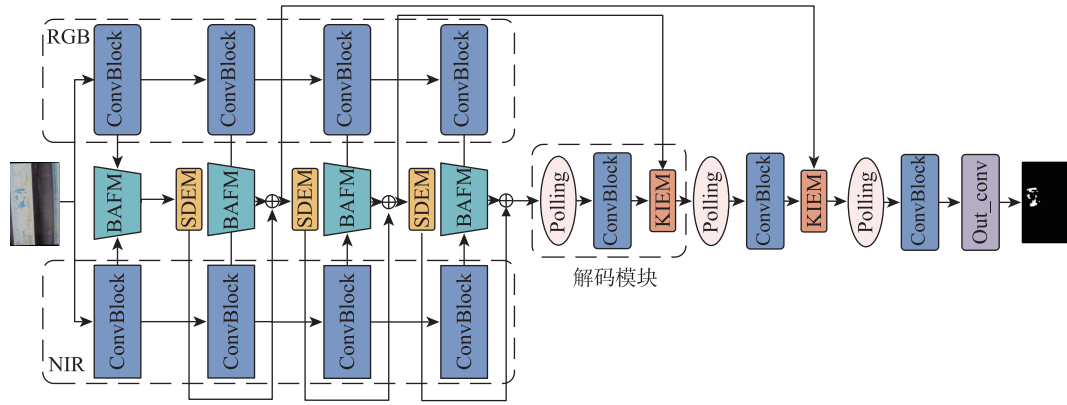


图1 整体框架

Fig. 1 Overall framework

SDEM模块。

在解码器部分,DAFNet采用3层解码层,并设计了KIEM模块,以解决缺陷与背景对比度低的问题;同时,通过引入自注意力机制,扩大了感受野,且捕捉长距离像素之间的依赖性,进而提高表面缺陷分割的精度。图1中,ConvBlock为卷积模块,Pooling为上采样。

2 编码器设计

编码器由两个并行的特征提取分支所组成,其中RGB分支处理3个通道的RGB图像信息,NIR分支处理一个通道的NIR图像信息。两个分支都由4个阶段构成,每个阶段包含2个卷积模块(ConvBlock)、1个激活函数(ReLU)以及1个批量归一化层(BatchNorm),具体结构如表1所示。

2.1 BAFM模块设计

为了能够有效融合RGB和NIR信息,本文提出

了一种基于注意力机制和门控机制的BAFM模块。

在BAFM模块设计中,首先对钢轨表面缺陷图像的RGB特征 f_i^{RGB} 和NIR特征 f_i^{NIR} 进行合并(Concat),并借助一个 1×1 卷积运算(Conv)来减少通道数及计算量,具体如图2所示;然后,为了能够同时关注钢轨表面缺陷的通道和空间特征,实现不同模态特征的有效融合,在通道(ChannelAttn)和空间注意力机制(SpatialAttn)改进过程中,两者同样采用双通道设计,并引入平均池化(AvgPool)和最大池化(MaxPool),使模块可以在保留局部信息的同时,提取出区域内显著的纹理特征,如图3所示。具体如下

$$CA = \text{BN}(f_{\text{avg}} + f_{\text{max}}) \quad (1)$$

$$f_{\text{avg}} = \text{Conv}_{1 \times 1}(\text{ReLU}(\text{Avg}(\text{Conv}_{1 \times 1}(f_i)))) \quad (2)$$

$$f_{\text{max}} = \text{Conv}_{1 \times 1}(\text{ReLU}(\text{Max}(\text{Conv}_{1 \times 1}(f_i)))) \quad (3)$$

$$SA = \text{BN}(f_i + \text{Conv}(\text{Cat}(\text{Avg}(f_i), \text{Max}(f_i)))) \quad (4)$$

表1 编码器结构
Tab.1 Encoder structure

阶段	模块	卷积核	输入尺寸(RGB)	输入尺寸(NIR)	输出尺寸
阶段 1	ConvBlock	3*3	3×784×1 344	1×784×1 344	16×392×672
	BAFM	—	16×392×672	16×392×672	16×392×672
	SDEM	—	16×392×672	16×392×672	32×196×336
阶段 2	ConvBlock	3*3	16×392×672	16×392×672	32×196×336
	BAFM	—	32×196×336	32×196×336	32×196×336
	SDEM	—	32×196×336	32×196×336	64×98×168
阶段 3	ConvBlock	3*3	32×196×336	32×196×336	64×98×168
	BAFM	—	64×98×168	64×98×168	64×98×168
	SDEM	—	64×98×168	64×98×168	128×49×84
阶段 4	ConvBlock	3*3	64×98×168	64×98×168	128×49×84
	BAFM	—	128×49×84	128×49×84	128×49×84

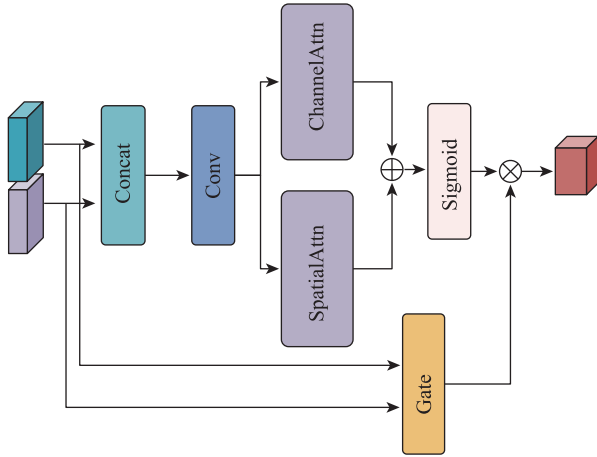


图2 BAFM 模块
Fig. 2 The structure of BAFM module

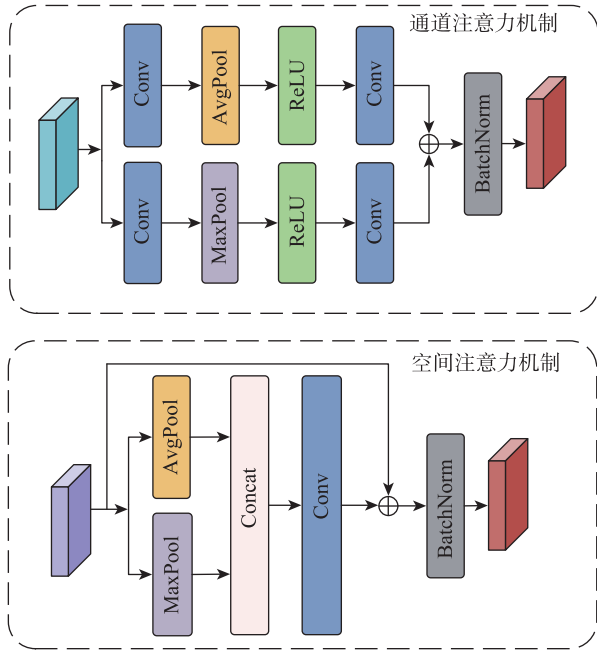


图3 通道注意力机制和空间注意力机制模块
Fig. 3 Channel attention mechanism and spatial attention module

式中: CA 和 SA 分别为通道和空间注意力机制; f_i 为输入特征; Avg 和 Max 分别为平均池化和最大池化运算; Cat 为 Concat 运算; BN 为 BatchNorm 运算。

此外, BAFM 模块还引入了门控机制(Gate), 通过门控权重, 来控制特征信息的传递, 凸显钢轨表面缺陷特征, 提高不规则表面缺陷的分割精度, 其结构如图4所示。

在门控机制中, 分别对钢轨表面缺陷的 RGB 和 NIR 特征进行处理, 合并后经过卷积及激活运算, 最终得到门控权重(Weight)为

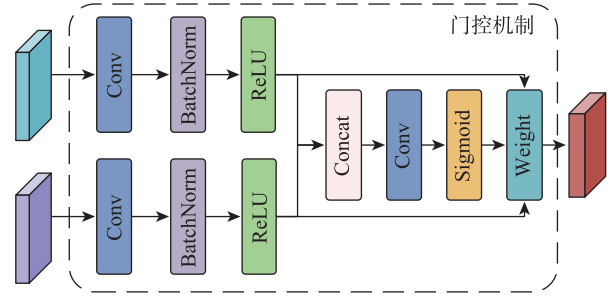


图4 门控机制
Fig. 4 Gating mechanism

$$w_i^{\text{gate}} = zh_1 + (1 - z)h_2 \quad (5)$$

$$z = \text{Sigmoid}(\text{Conv}_{1 \times 1}(\text{Cat}(h_1, h_2))) \quad (6)$$

$$h_1 = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(x_1))) \quad (7)$$

$$h_2 = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(x_2))) \quad (8)$$

式中: w_i^{gate} 为门控权重; x_1 为 RGB 输入特征; x_2 为 NIR 输入特征。由此, 得到的融合特征 f_i^{BAFM} 为

$$f_i^{\text{BAFM}} = \text{Sigmoid}(CA(f_i) + SA(f_i)) \times w_i^{\text{gate}} \quad (9)$$

$$f_i = \text{Cat}(\text{Conv}_{3 \times 3}(f_i^{\text{RGB}}, f_i^{\text{NIR}})) \quad (10)$$

式中: i 为编码器不同阶段 ($i = 1, 2, 3, 4$)。

2.2 SDEM 模块设计

由于在编码器中经过多次下采样操作, 空间细节信息会逐渐丢失, 导致对小尺度缺陷边缘的分割效果不佳。为进一步提高缺陷边缘的精度, 本文采用多尺度卷积和边缘检测算子相结合的方式, 设计了 SDEM 模块, 其结构如图5所示。

在 SDEM 模块设计中: 首先采用不同尺寸的卷积核 (DC_3*3, DC_5*5, DC_7*7) 来提取不同尺度的特征, 增强空间细节信息的提取能力, 同时为了减少参数量, 保持模型性能, 采用深度可分离卷积代替普通卷积核; 然后, 为了提取丰富的边缘细节特征, 引入 Sobel 算子, 相较于传统的 Sobel 算子, 加入 45° 和 135° 方向检测, 能更全面地捕捉图像中的边缘信息; 最后, 最大池化层能捕捉钢轨表面缺陷边缘的主要特征, 提高分割的精度。

SDEM 模块将 BAFM 模块的结果 f_i^{BAFM} 作为输入, 经过深度可分离卷积核, 得到 3 种不同尺度的特征, 然后与 Sobel 算子和最大池化层的特征进行合并, 最终通过卷积得到输出。

3 解码器设计

解码器由 3 个相同的解码层和 1 个输出层

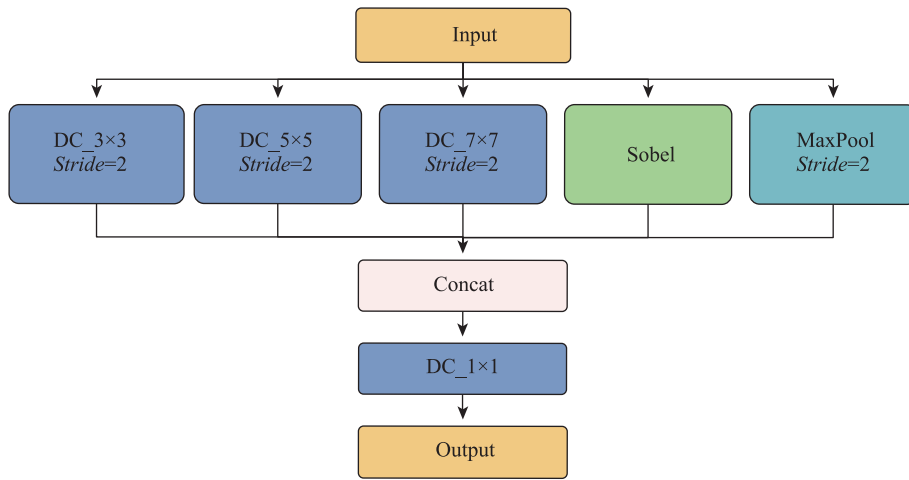


图5 SDEM 模块
Fig. 5 SDEM module

(OutLayer)组成,前2个解码层包含上采样(Pooling)、卷积模块和KIEM模块,第3个解码层只包含上采样、卷积模块。解码器结构如表2所示。为了扩展感受野,捕捉长距离像素依赖关系,充分提取多类多尺度缺陷的类别特征和位置信息,提高了缺陷分割的精度,本文设计了KIEM模块,其结构如图

6所示。

KIEM模块的输入包括卷积的输出特征 f_i^{Conv} 和BFAM模块的特征 f_i^{BAFM} ,经过Concat层合并后,再引入改进的自注意力机制(NonLocal^[11])模块。

在NonLocal机制模块改进过程中,通过捕捉不同位置之间的依赖关系,计算每个像素与其他像素

表2 解码器结构
Tab.2 Decoder structure

解码层	模块	卷积核	输入尺寸	输出尺寸
解码层 3	Pooling	-	128×49×84	128×98×168
	ConvBlock	3*3	128×98×168	64×98×168
	KIEM	-	64×98×168	64×98×168
解码层 2	Pooling	-	64×98×168	64×196×336
	ConvBlock	3*3	64×196×336	32×196×336
	KIEM	-	32×196×336	32×196×336
解码层 1	Pooling	-	32×196×336	32×392×672
	ConvBlock	3*3	32×392×672	16×784×1 344
OutLayer	Conv	1*1	16×784×1 344	3×784×1 344

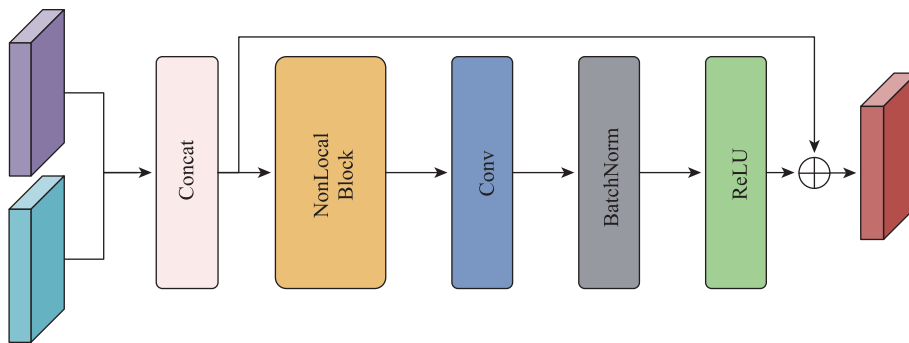


图6 KIEM 模块
Fig. 6 KIEM module

之间的注意力权重,提高了网络对缺陷和背景区分能力,进而提高了分割精度。期间,为了避免 Query, Key, Value 矩阵计算带来的参数量剧增的问题,本文在计算前使用最大池化层进行下采样,同时使用了 1×1 的卷积核进行降维,以降低特征的空间分辨率和计算量。同时在训练前,将输出的 Embedding 层 f^{oc} 权重和偏置初始化为零,有助于网络的训练稳定性。具体计算过程为

$$f^{nlb} = \text{BN}(\text{Conv}_{1 \times 1, \text{stride}=1}(f^{oc})) + f^{Cat} \quad (11)$$

$$f^{oc} = (f^k * f^q) * f^v \quad (12)$$

$$f^{k,q,v} = \text{Max}(\text{Conv}_{1 \times 1, \text{stride}=1}(f^{Cat})) \quad (13)$$

$$f^{Cat} = \text{Cat}(f_i^{\text{Conv}}, f_i^{\text{BAFM}}) \quad (14)$$

式中: * 为矩阵乘法; $f^{k,q,v}$ 分别为 Key、Query 和 Value 矩阵的计算结果; f^{Cat} 为 Concat 的输出; f^{nlb} 为 NonLocal 机制模块的输出。

综上所述, KIEM 模块的计算过程可表示为

$$f_i^{\text{KIEM}} = f^{Cat} + \text{ReLU}(\text{BN}(\text{Conv}(f^{nlb}))) \quad (15)$$

4 数据集

本文实验使用了 MER2-507-23GC 和 MER2-507-23GM NIR 相机,作为可见光和红外图像的采集设备。首先,采集的 3 个通道 RGB 样本和 1 个通道 NIR 样本经过手动匹配校准之后,合并为一张 4 通道图像,如图 7 所示;再将其裁剪为 784×1344 大小,制作出包含 545 张样本的原始数据集;然后,采用随机翻转、添加噪声和亮度变换等数据增强技术,将数据集扩充到 2 180 张,其中 80% 的图像数据作为训练及验证集,20% 的图像作为测试集。

5 实验结果分析

实验采用英特尔 Core i7 处理器和 16 GB 的 NVIDIA TESLA T4 显卡进行网络训练。环境配置为 Py-



图7 钢轨表面缺陷原始样本

Fig. 7 Original image of rail surface defects

thon 3.8, PyTorch 1.7 和 CUDA 10。在训练过程中,网络使用随机梯度下降(SGD)优化器,初始学习率为 0.000 5,并随训练过程呈指数衰减。采用交叉熵损失函数^[12],经过 120 个周期的迭代,损失值趋于平稳。图 8 展示了训练过程中损失的变化情况。

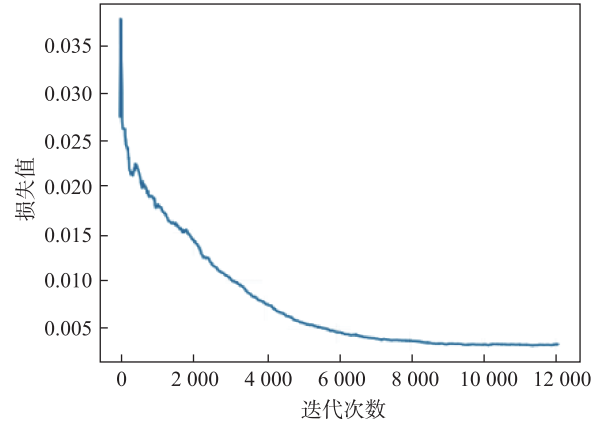


图8 训练损失

Fig. 8 Training loss

5.1 对比实验

为了验证本文所提出网络的先进性,与主流的语义分割网络(FCN, DFN, FuseNet^[13], ACNet^[14], MFNet, RTFNet, FEANet, GMNet, CMX^[15])进行对比,其中包括单模态网络 FCN, DFN。

5.1.1 不同模型的分割精度对比实验

表 3 展示了不同网络在 mAcc 和 mIoU 指标上的对比实验结果。实验表明,面对复杂场景下的多类多尺度钢轨表面缺陷,单模态网络由于缺乏 NIR 图像的支撑,综合分割效果不如双模态网络。同时,本文所提出 DAFNet 网络对比其他双模态网络,

表3 不同模型的精度对比结果

Tab.3 Accuracy comparison results of different models %

网络	类别	mAcc	mIoU
FCN (ResNet-50)		51.21	48.81
DFN		43.34	43.15
FuseNet		55.37	50.96
ACNet		62.98	57.51
MFNet	RGB +NIR	61.74	55.04
RTFNet		48.65	46.18
FEANet (ResNet-50)		54.77	52.39
GMNet		52.53	48.91
CMX		53.02	49.85
DAFNet		68.13	59.96

mAcc 和 mIoU 指标要高出许多。

5.1.2 不同模型的可视化结果展示

为了检测 DAFNet 在实际场景中的效果,本文对实验结果进行了可视化展示,如图9所示。通过对比原始图像、真实标签(GroundTruth)和不同网络的预测结果,可以发现,只有 DAFNet 能够完整覆盖原始图像中表面缺陷,最接近 GroundTruth,同时还能捕获出未在真实标签中标注的缺陷轮廓(灰色标注框区域)。由此证明,本文所设计的 DAFNet 网

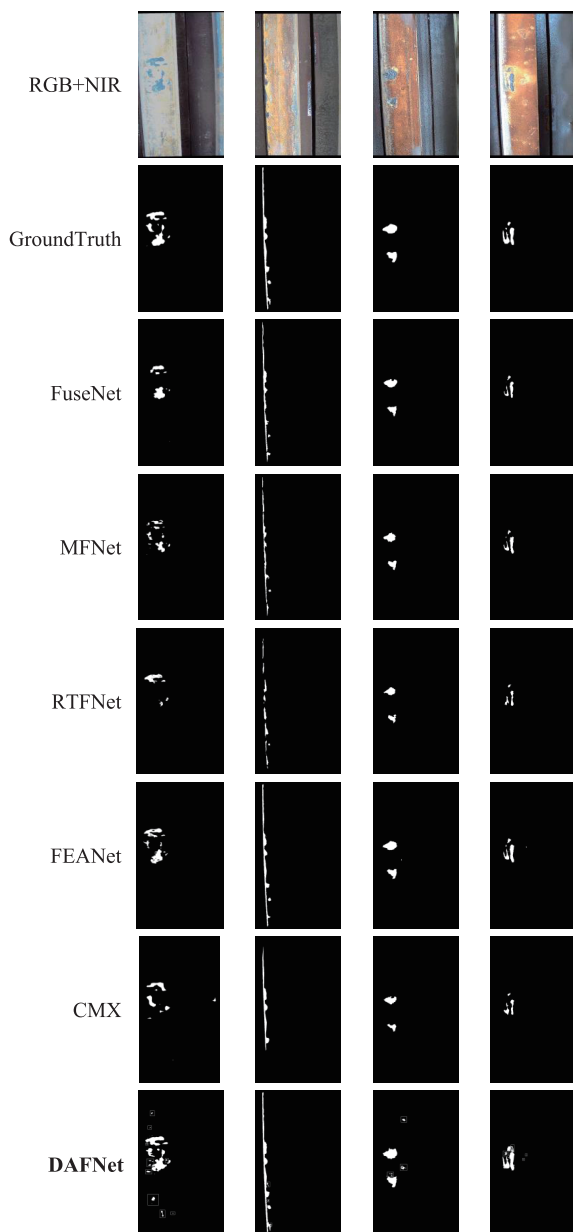


图9 不同模型的可视化结果展示

Fig. 9 Visualization results of different models

络,能够有效适应实际场景的需求。

5.1.3 不同模型的分割速度对比实验

表4展示了本文提出的 DAFNet 与其他双模态网络对比结果。可以发现,DAFNet 网络的各项指标都远远优于大部分双模态网络。但 MFNet 在参数量和模型大小方面略小于 DAFNet,这是由于 BAFM 模块和 KIEM 模块的引入,增加了计算量,但 DAFNet 的 mAcc 和 mIoU 指标比 MFNet 网络分别高出 6.39 个百分点和 4.92 个百分点。

表4 双模态网络的 FLOPs,参数量和模型大小对比
Tab.4 Comparison of FLOPs, parameter count, and model size of dual-modal networks

网络	FLOPs/GFLOPs	参数量/M	模型大小/MB
FuseNet	241.89	42.12	168.65
ACNet	106.04	111.20	445.83
MFNet	6.72	0.70	2.88
RTFNet	29.19	29.55	118.45
FEANet	208.93	177.31	710.05
GMNet	127.86	146.36	586.40
CMX	5.85	10.43	41.89
DAFNet	17.41	1.38	5.67

5.1.4 DAFNet模型鲁棒性对比实验

为了测试 DAFNet 的泛化能力,使用 5 种干扰方式(矩形遮挡、模糊遮挡、噪声遮挡、高斯噪声和高斯模糊)来评估模型的性能。表5展示了实验结果,模型在加入干扰后,平均后的 mAcc 为 62.57%, mIoU 为 54.46%,分别下降了 5.56 个百分点和 5.50 个百分点。

此外,为了增强模型的鲁棒性,本文将这些干扰方式加入到训练集中。经过重新训练后,模型在测试集上的性能有了显著提升,平均后的 mAcc 和 mIoU 分别仅下降了 2.73 个百分点和 2.80 个百分点。这一结果表明,通过在训练过程中引入干扰,可以有效提升 DAFNet 在复杂环境下的鲁棒性,提高实际应用中的稳定性。

5.2 消融实验

为了验证不同模块的有效性,本文设计了 5 组消融实验,表6展示了不同实验下的结果。图表中 NADK 表示去除 BAFM 模块、SDEM 模块和 KIEM 模块的网络;NDK 表示去除 SDEM 模块和 KIEM 模块的网络;NDA 表示去除 SDEM 模块和 BAFM 模块的网络;NAK 表示去除 BAFM 模块和 KIEM 模块的网络。

表5 DAFNet模型鲁棒性对比实验结果

Tab.5 Experimental results of robustness comparison of DAFNet

干扰方式	无干扰训练		有干扰训练		%
	mAcc	mIoU	mAcc	mIoU	
矩形遮挡	60.87 (↓7.26)	52.35 (↓7.61)	64.54 (↓3.29)	56.53 (↓3.34)	
模糊遮挡	62.12 (↓6.01)	53.56 (↓6.40)	65.21 (↓2.92)	56.98 (↓2.98)	
噪声遮挡	61.21 (↓6.92)	52.95 (↓7.01)	64.83 (↓3.30)	56.52 (↓3.44)	
高斯噪声	64.53 (↓3.6)	57.13 (↓2.83)	66.56 (↓1.57)	58.07 (↓1.89)	
高斯模糊	64.14 (↓3.99)	56.34 (↓3.62)	65.89 (↓2.24)	57.74 (↓2.22)	

表6 各个模块消融实验对比

Tab.6 Comparison of ablation experiments for various modules

模块	mAcc	mIoU	%
NADK	54.63	50.48	
NDK	56.97	52.03	
NDA	57.96	52.71	
NAK	56.98	51.75	
DAFNet	68.13	59.96	

1) NDK与NADK相比,mAcc和mIoU分别提高了2.34个百分点和1.55个百分点,说明了BAFM模块可以有效融合RGB和NIR特征,从而提升网络的分割精度。

2) NDA与NADK相比,mAcc和mIoU分别提高了3.33个百分点和2.23个百分点,说明KIEM模块有效解决了缺陷与背景对比度不高的问题,从而提高分割的准确性。

3) NAK与NADK相比,mAcc和mIoU分别提高了2.35个百分点和1.27个百分点,说明了SDEM模块能有效提取空间细节特征,提高网络的分割精度。

4) DAFNet网络与NADK相比,mAcc和mIoU分别提高了13.50个百分点和9.48个百分点,说明了同时引入3个模块时,DAFNet具有最高的分割精度。

6 结论

本文提出的DAFNet分割网络利用可见光和红外图像特征的互补性,提升了复杂场景下多类多尺度钢轨表面缺陷检测的精度与速度。主要结论如下。

1) 改进的双分支网络架构,可以有效提高分割速度。

2) BAFM模块能实现特征的自适应融合,提高不规则表面缺陷的分割精度。

3) SDEM模块可以提高对钢轨表面缺陷特征的空间细节表达能力。

4) KIEM模块能有效解决复杂场景下缺陷与背景对比度不高的问题。

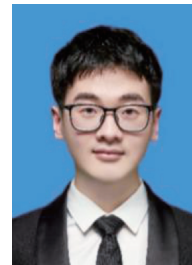
参考文献:

- [1] JIA Y, LIANG K, WANG P, et al. An enhancement-method of magnetic flux leakage signals for rail track surface defect detection[J]. IET Science Measurement & Technology, 2020, 14(2): 711-717.
- [2] PARK J W, LEE T G, BACK I C, et al. Rail surface defect detection and analysis using multichannel eddy current method based algorithm for defect evaluation[J]. Journal of Nondestructive Evaluation, 2021, 40(3): 1-12.
- [3] 周宇, 木东升, 韩延彬, 等. 基于X射线断层扫描的钢轨滚动接触疲劳裂纹形状建模方法[J]. 华东交通大学学报, 2019, 36(1): 41-48.
- [4] ZHOU Y, MU D S, HAN Y B, et al. A modeling method for the shape of rolling contact fatigue cracks in steel rails based on X-ray tomography[J]. Journal of East China Jiaotong University, 2019, 36(1): 41-48.
- [5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [6] YU C, WANG J, PENG C, et al. Learning a discriminative feature network for semantic segmentation[C]//Salt

- Lake: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 1857-1866.
- [6] HA Q, WATANABE K, KARASAWA T, et al. MFNet: towards real-time semantic segmentation for auto-nomous vehicles with multi-spectral scenes[C]//Vancouver: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017.
- [7] SUN Y, ZUO W, LIU M. RTFNet: RGB-thermal fusion network for semantic segmentation of urban scenes[J]. IEEE Robotics and Automation Letters, 2019, 4(3): 2576-2583.
- [8] ZHOU W, LIU J, LEI J, et al. GMNet: graded feature multilabel-learning network for RGB-thermal urban scene semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30(3): 7790-7802.
- [9] DENG F, FENG H, LIANG M, et al. FEANet: feature enhanced attention network for RGB-thermal real-time semantic segmentation[C]//Prague: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021.
- [10] CUBUK E D, ZOPH B, MANE D, et al. Autoaugment: learning augmentation strategies from data[C]//Long Beach: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 113-123.
- [11] TAO Y Z, SUN Q, DU Q, et al. Nonlocal neural networks, nonlocal diffusion and nonlocal modeling[C]//Montréal: Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), 2018: 494-504.
- [12] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back propagating errors[J]. Nature, 1986, 323: 533-536.
- [13] HAZIRBAS C, MA L, DOMOKOS C, et al. FuseNet: incorporating depth into semantic segmentation via fusion-based CNN architecture[C]//Taipei: Proceedings of the Asian Conference on Computer Vision, 2016.
- [14] HU X, YANG K, FEI L, et al. ACNet: attention based network to exploit complementary features for RGBD semantic segmentation[C]//Taipei: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2019.
- [15] LIU H, ZHANG J, YANG K, et al. CMX: cross modal fusion for RGB-X semantic segmentation with transformers[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(12): 14679-14694.



第一作者: 罗晖(1969—), 男, 教授, 硕士生导师, 研究方向为人工智能网络、机器视觉。E-mail: lh_jxnc@163.com。



通信作者: 韩岳霖(1998—), 男, 硕士研究生, 研究方向为机器视觉。E-mail: dj274570787@163.com。

(责任编辑: 吴海燕)