

DOI: 10.13382/j.jemi.B2508166

融合多级注意力与多尺度信息的铁轨缺陷分割网络*

周炜杰^{1,2} 李智² 张绍荣² 唐洪旻² 莫云²

(1. 桂林电子科技大学电子工程与自动化学院 桂林 541004; 2. 桂林航天工业学院电子信息与自动化学院 桂林 541004)

摘要: 铁路轨道缺陷检测技术面临许多挑战。轨道表面纹理复杂、背景噪声干扰严重,使得缺陷难以检出;缺陷种类多样,形态各异,导致检测方法难以同时捕捉所有细节特征;尺寸较小的缺陷由于特征不明显,往往会被漏检。为了精确分割铁路轨道表面缺陷,提出一种融合多级注意力与多尺度信息的铁轨缺陷分割网络。该网络的编码器通过堆叠倒置瓶颈卷积和融合倒置瓶颈卷积有效提高特征提取编码的效率;解码器部分使用多级并行像素级注意力模块辅助模型从大量背景噪声中聚焦定位缺陷区域;金字塔池化模块用于捕获多尺度上下文信息,增强模型对场景中的局部和全局特征的解析能力;多尺度信息融合方法融合像素级注意力模块和金字塔池化模块的输出,充分利用各阶段的特征信息。利用 NRSD-MN 数据集进行实验,在 Craft 和 Real 两类数据上,平均精度分别达到 0.836 4 和 0.725 8;平均交并比分别达到 0.685 8 和 0.634 2。实验结果表明,提出的网络在针对铁路轨道表面缺陷分割任务时,精度上显著优于现有的模型。

关键词: 缺陷检测;语义分割;注意力机制;多尺度信息融合;小目标检测

中图分类号: TM93;TP391.4;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Railway defect segmentation network with multi-level attention and multi-scale information fusion

Zhou Weijie^{1,2} Li Zhi² Zhang Shaorong² Tang Hongkuang² Mo Yun²

(1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China;

2. School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin 541004, China)

Abstract: Railway defect detection faces many challenges. The complex texture of the railway surface, background noise interference is serious, making it difficult to detect defects; defects of various types, different morphology, resulting in the traditional detection methods are difficult to capture all the details of the features at the same time; smaller defects due to the characteristics of the characteristics are not obvious, often missed. To address these issues, this paper proposes a novel semantic segmentation network that integrates a multi-level parallel attention mechanism and multi-scale information fusion to enhance defect segmentation accuracy. In the encoder, feature extraction and encoding efficiency are improved by leveraging stacked Inverted Bottleneck Convolutions and Fused Inverted Bottleneck Convolutions. The decoder incorporates a multi-level parallel pixel attention module (PAM) to enable the network to effectively focus on and localize defect regions amidst considerable background noise. Additionally, a pyramid pooling module (PPM) is introduced to capture multi-scale contextual information, enhancing the model's ability to extract both local and global features. A multi-scale spatial information fusion strategy further integrates the outputs of PAM and PPM, maximizing the utilization of feature representations across different levels. Experimental evaluations on the NRSD-MN dataset demonstrate that the proposed method achieves mPA values of 0.836 4 and 0.725 8 and mIoU scores of 0.685 8 and 0.634 2 on the Craft and Real data subsets, respectively. The results confirm that the proposed network outperforms existing models in railway track surface defect segmentation, offering superior accuracy and robustness.

Keywords: defect detection; semantic segmentation; attention mechanism; multi-perspective feature fusion; small target detection

收稿日期: 2025-02-16 Received Date: 2025-02-16

* 基金项目: 广西重点研发计划项目(桂科 AB24010366)、广西自然科学基金面上项目(2025GXNSFAA069804)、桂林航天工业学院特色优势交叉学科发展战略研究专项(TS2024431)资助

0 引言

钢轨表面的缺陷会加速列车车轮的磨损,这些缺陷在早期主要表现为锈斑、剥落或裂纹^[1-3]。由于轨道长期暴露在复杂环境中,表面附着铁锈、焊缝、磨痕、污渍等多种纹理信息,这些纹理在视觉上真实缺陷区域高度相似,从而导致检测算法误检或漏检。

在过去的几年里,基于机器学习和计算机视觉的表面缺陷检测^[4-7]已成为一个热门的研究方向。王勇等^[8]使用简单的BP神经网络实现了轨道缺陷的分割。马茜等^[9]结合神经网络和二进制SCM分类器分别实现了轨枕和紧固件缺陷检测。传统机器学习虽然能够实现缺陷分类检测,但是在面对多类别、多形态缺陷检测方面存在不足。韦若禹等^[10]改进YOLOv3网络用于轨道缺陷检测;王德伟等^[11]设计一种高效的PConv检测头来改进YOLOv8网络;黄啸等^[12]提出一种轻量级双路并行注意力模块并将该模块引入YOLOX-s。但是上述基于YOLO系列的目标算法只能分类和定位而不能逐像素精细分割出缺陷。付宇彤等^[13]提出一种编解码器架构的语义分割网络来对铸件缺陷进行语义分割。Li等^[14]提出了一种结合归一化注意力机制和双尺度特征交互模块的分割网络。然而对于某些微小的缺陷仍然存在检测不到或分割精度低的问题,造成缺陷与背景边界模糊。任中伟等^[15]设计轻量化算法,Zhang等^[16]采用一种双网络架构,能够在样本有限且标注信息不完全的条件下,实现轨道表面缺陷的检测和分类。

综上,上述算法存在着一些局限性,在缺陷图像中,存在着某些单边尺寸小于10 pixels、面积占比低于5%的小目标缺陷,算法对此识别能力不足。未充分利用不同层级的特征,导致在复杂背景下的分割精度下降;网络训练依赖高质量轨道缺陷图片数据。因此,本文提出了一种用于铁路轨道表面缺陷检测方法,高效地实现轨道表面缺陷的逐像素精细分割任务。本文提出了一种用于铁路表面缺陷检测的基于编解码器框架的语义分割网络。设计了多个像素级注意力机制模块辅助模型对复杂背景下的微小目标缺陷进行精确定位分割。构建了多尺度金字塔池化模块,用于捕捉不同尺度的空间特征,增强网络对局部与全局信息的处理能力。引入一种多尺度特征信息融合方法,将多级特征与多尺度空间信息融合,确保网络能高效利用各阶段的特征信息。

1 语义分割算法关键技术

现阶段针对轨道表面缺陷检测^[17]领域的图像分割算法主要致力于在3个关键方面进行优化与改进:语义

分割网络的结构设计与优化、注意力机制的选取和改进、特征融合方法的优化与应用。首先,语义分割网络的设计优化致力于提高模型的检测准确性,侧重于减少计算复杂度和提升推理速度。其次,注意力机制的引入,旨在增强模型对关键信息的捕捉能力。最后,特征融合方法的优化如何有效整合不同层次和尺度的特征信息,提升模型对细节和全局语义的表达能力,同时提高特征融合的计算效率和自适应性。接下来的内容将详细综述相关领域的研究进展。

1.1 基于卷积神经网络的语义分割方法

FCN^[18]是首个将卷积神经网络应用于像素级别分类的端到端训练方法。然而,由于FCN采用多次下采样策略,小尺寸目标的特征信息在深层特征图中容易被忽略。因此,U-Net^[19]通过对称的编码器-解码器结构,在每个编码器层和解码器层之间建立了跳跃链接使低层的细节信息能够传递到高层特征中。为了高效进行特征优化利用,ENet^[20]利用轻量化编码器-解码器结构、早期下采样等手段,实现了低计算复杂度下的高效实时语义分割。但这在提升计算效率的同时也损失了检测精度。

1.2 注意力机制

在计算机视觉任务中,最常用的注意力机制有空间域、通道域和混合域3种。复杂背景和噪声会导致模型难以区分真实缺陷与背景纹理,空间注意力机制能够帮助模型聚焦于关键区域,忽略无关信息。PiCANet^[21]通过计算特征图中每个位置与其他位置的关系,生成空间注意力图,从而捕获长距离依赖关系。由于缺陷区域的纹理可能与背景相似,通道注意力可以增强关键特征通道的表达,降低纹理噪声对分割的干扰。通道注意力机制提升模型的表达能力,SENet^[22]通过对每个通道进行“压缩”和“激励”操作,来调整每个通道的重要性。混合域的注意力机制是将空间注意力机制和通道注意力机制结合起来,CBAM^[23]整合了空间和通道注意力,以并联的方式进行组合。

1.3 多尺度特征融合

在计算机视觉任务中,不同阶段的网络提取到不同尺度的特征信息。通常来说,浅层网络的感受野比较小,空间几何细节信息表征能力强,提取到的特征图分辨率高,但是语义信息表征能力弱;深层网络的感受野比较大,语义信息表征能力强,提取到的特征图分辨率低,但是空间集合信息表征能力弱。为了融合多尺度特征信息,有两种基本方法,即拼接和元素相加。拼接是指将来自不同尺度的特征图在通道维度上进行融合,这种方法保留了各个尺度特征图的完整信息。而逐元素相加法则需要在融合前对特征图尺寸进行统一操作,其优势在于计算效率高。

2 铁轨表面缺陷分割模型

本文提出的神经网络模型采用 Encoder-Decoder 结构,如图 1 所示,实现端到端的铁路轨道表面缺陷精确分割。编码器部分包含 9 个阶段,这 9 个阶段由浅到深共提取 9 个不同尺度的特征图。解码器部分包含像素级注意力模块 (pixel attention module, PAM)、金字塔池化模块 (pyramid pooling module, PPM) 和特征融合块 (Fusion)。多级并行 PAM 同时关注位置和像素的联系,将第 5~8 阶段的特征图进行加强特征处理;PPM 处

理第 9 阶段的特征信息,进行不同尺度池化处理,获得多尺度空间信息;特征融合块对 PAM 和 PPM 的输出进行融合,得到最终的缺陷分割结果。

本文提出的网络结构具有以下优势:编码器采用逐级特征提取策略,在不同层级上充分捕获局部与全局信息,为后续的网络阶段提供了丰富的特征支持。编码器中有多条数据流与解码器中的 PAM 和 PPM 交互,从而有效增强了特征的流动性和传递性,防止了上采样过程中关键信息的丢失。特征融合块进行二次整合与解析,使得特征信息得以充分利用,从而能显著提升分割效果。

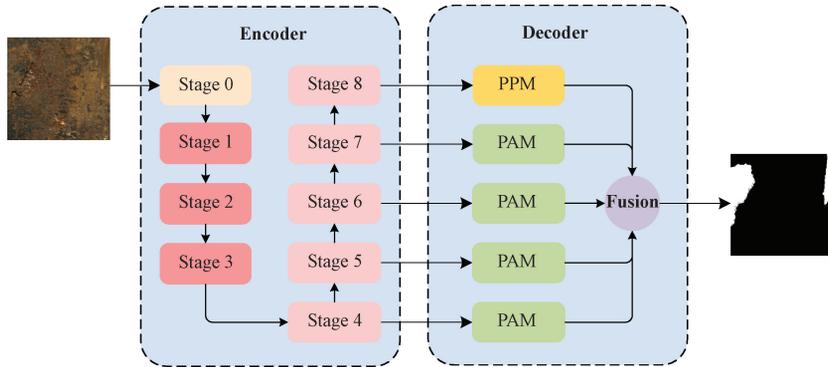


图 1 本文提出的网络结构

Fig. 1 The architecture of our proposed network

2.1 编码器:多尺度特征提取

为了提取有效特征,编码器网络采用多级特征提取的方式,一共包含 9 个特征提阶段 (Stage)。这 9 个阶段可以分为 4 个部分,如表 1 所示。

表 1 特征编码器结构

Table 1 Feature encoder architecture

阶段	模块
Stage 0	Conv2d-Norm-Activation
Stage 1~Stage 3	Fused-MBConv
Stage 4~Stage 7	MBConv
Stage 8	Conv2d-Norm-Activation

Stage 0 和 Stage 8 结构类似,都是包含一个二维卷积、批归一化和 SiLU 激活函数。这两个阶段起到的作用都是升维和通道数转换。MBConv 模块结构和 Fused-MBConv 模块结构如图 2 所示。

MBConv 模块的输入和输出部分都包含一个 1×1 卷积分别进行升维和降维,其核心部分是一个深度可分离卷积 (depthwise convolution, DWConv), 比普通卷积拥有更少的参数量和计算量。虽然 DWConv 结构比普通卷积拥有更少的参数和计算量,但由于其计算密度低、数据访问不连续,反而影响训练或推理速度。因此在 MBConv

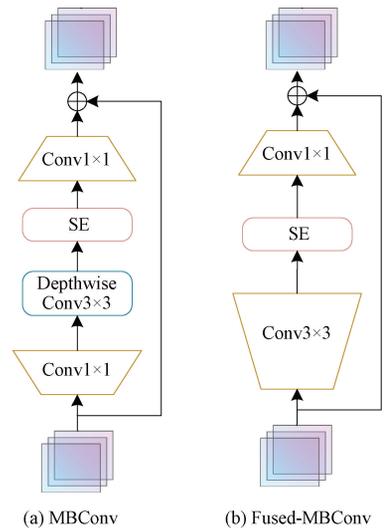


图 2 MBConv 和 Fused-MBConv 模块结构

Fig. 2 The structure of MBConv and Fused-MBConv modules

模块的基础上,使用 Fused-MBConv 模块来解决这一问题,通过将 DWConv 替换成普通的 3×3 卷积,能够明显提升训练速度。在浅层网络,即 Stage 1~Stage 3,通过堆叠 Fused-MBConv 来提升训练速度,而在深层网络阶段,即 Stage 4~Stage 7,通过堆叠 MBConv 来平衡性能、参数和

计算量之间的关系,提升模型的计算效率。

这9个阶段逐级进行多尺度特征进行编码,一共可以得到4个低级特征图 $M_0 \sim M_3$, 5个高级特征图 $M_4 \sim M_8$ 。特征图的输出尺寸如图3所示。

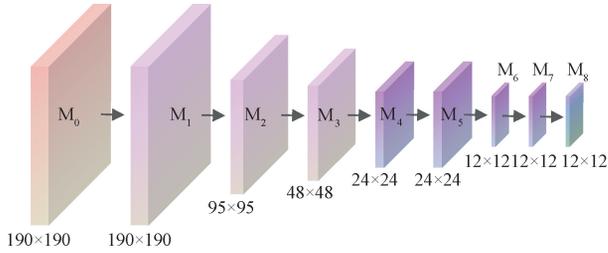


图3 特征编码器输出尺寸

Fig. 3 Output size of feature maps in the encoder

2.2 解码器:多尺度特征融合

1) 像素级注意力机制

低级的特征包含更多的空间信息,而高级的特征包含更多语义信息,为了更好的利用特征提取主干网络提取到的9个不同层级的特征图,使用多级并行像素级注意力机制模块,处理 $M_4 \sim M_7$,其结构如图4所示。

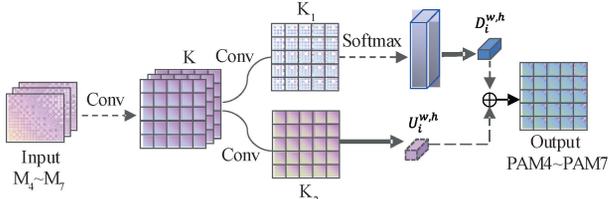


图4 PAM结构示意图

Fig. 4 Structure of the PAM

由于待分割的铁路轨道表面缺陷图片包含大量噪声和复杂纹理信息,而网络模型需要从大量噪声中重点关注缺陷区域的信息,因此设计PAM模块来为特征图中的每一个像素生成局部注意力。多级并行PAM,理论上可以显著提高小目标的检测能力。多层卷积网络在多次下采样过程中,往往会丢失小目标的细节信息,而PAM模块通过局部像素级注意力机制,对每个像素及其邻域信息进行动态调节,使得小目标缺陷区域获得更高的权重,降低了小目标在特征提取过程中被忽略的概率。

该模块由两个卷积层构成,第1层卷积核的大小为 7×7 ,用于生成初步注意力计算核 K ;第2层卷积核的大小为 1×1 ,进行通道数转换。PAM的前向传播过程:第1步,输入的特征图 $M_4 \sim M_7$ 经过第一层卷积处理,产生中间特征图 K ;第2步,对 K 进行第2层卷积操作,得到特征图 K_1, K_2 ;第3步,使用 softmax 激活函数对 K_1 特征图生成注意力权重,得到 D_i ;第4步,将特征图 K_2 进行展开操作转换为多个局部区 U_i ;第5步,将转换展开的特

征图 U_i 与注意力权重 D_i 进行逐个元素相乘,计算过程如式(1)所示。

$$Y_{w,h} = \sum U_i^{w,h} \times D_i^{w,h} \quad (1)$$

式中: w, h 表示特征图的宽和高; Y 为输出的特征图 $AM_4 \sim AM_7$ 。

2) 金字塔池化模块

特征编码器的深层阶段 Stage 8 提取到包含高级语义信息的特征图 M_8 , 使用金字塔池化模块对其处理,利用不同尺度的自适应平均池化,捕捉图像的全局上下文信息。这种方法有助于模型理解视野中各物体的全局和局部特征,从而提高分割的准确率。

金字塔池化模块实现结构如图5所示,将 Stage 8 提取到的特征图 M_8 分为4组,每一组分别对应不同大小的池化操作。PPM的前向传播过程:首先,设定4组池化核大小为 $[1, 2, 3, 6]$,每一组的输出通道数为输入通道数的 $1/4$ 。其次,采用自适应平均池化方法,将 M_8 转换成4个不同尺度的金字塔池化特征图,其尺寸为 $\{n \times n \times d/4\}$ 。最后,这4个不同尺度的金字塔池化特征图经过尺寸统一处理后进行拼接融合,得到最终的输出 PPM_8 。

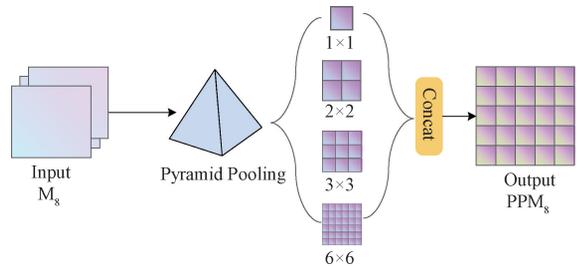


图5 PPM结构示意图

Fig. 5 Structure of the PPM

3) 多尺度空间信息融合方法

特征编码器中 Stage 4~Stage 7 的输出经解码器中的PAM处理后,得到 $PAM_i (i = 4, 5, 6, 7)$ 4个特征图; Stage 8 的输出 M_8 经解码器中的 PPM 操作后,得到一个特征图 PPM_8 。为了融合 PAM_i 和 PPM_8 ,应用了一种多尺度特征融合方法,如图6所示。

首先,对 PPM_8 进行转置卷积操作,将其特征图尺寸大小转换为 48×48 ,保持其通道数不变。其次,对 PAM_i 进行双线性插值的上采样处理,统一其尺寸大小到 48×48 。最后,将其进行融合,这一过程可以表示为:

$$PAM_{all} = \{PAM_4 \oplus PAM_5 \oplus PAM_6 \oplus PAM_7\} \quad (2)$$

$$PAM_{out} = f\{PAM_{all}\} \quad (3)$$

$$Output = f\{PAM_{out} \oplus PPM_8\} \quad (4)$$

式中: \oplus 表示为逐元素相加; f 表示为卷积核为 1×1 的卷积操作。

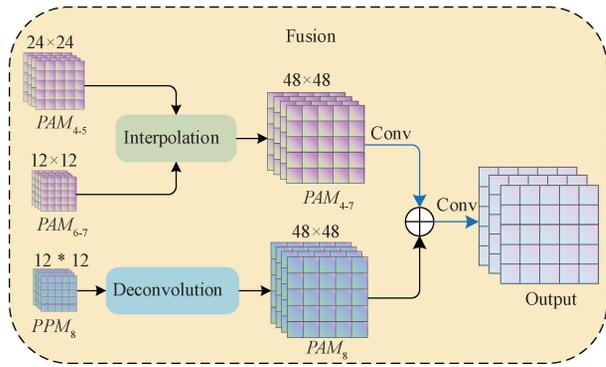


图 6 多尺度特征融合方法

Fig. 6 Multi-perspective feature fusion method

这样操作后,输出尺寸大小为 48×48 ,经过一次双线性插值后,尺寸扩大到与原输入尺寸一样的大小,得到最终的缺陷分割结果。

3 实验

3.1 数据集

本文使用公开数据集:NRSD (no-service rail surface defects) 来进行实验。包含 Man-made NRSD 和 Natural NRSD 两个部分。其中,测试集分为 Craft 和 Real 两个子集,分别代表 Man-made NRSD 和 Natural NRSD。

3.2 模型训练

使用 GeForce RTX 3090 24 G GPU,基于 Pytorch

2.2.1 深度学习框架。所有的训练图像和标签大小统一

为 380×380 pixels,并归一化均值和标准差。设置训练集批次大小为 8,验证集为 4,损失函数使用交叉熵损失。优化器采用 Adam,学习率调度器为 StepLR。网络总共训练 50 个 Epoch。

3.3 性能评价指标

本任务是针对轨道表面的缺陷进行像素分割,只需要分出缺陷和背景,因此是一个二分类问题。本文主要使用 mIoU、PA、Dice 系数来评估模型的性能,推理时间、浮点运算性能作为辅助参考指标。

3.4 实验结果对比

本文在 NRSD-MN 数据集上进行了实验,对比评估了 8 种较为先进的语义分割算法,包括 FCN^[18]、Unet^[19]、U2Net^[24]、DeeplabV3+^[25]、PSPNet^[26]、GCN^[27]、YOLO11-Seg^[28]、YOLO12-Seg^[29]。最好的结果用粗体表示。实验结果如表 2 所示。本文提出的算法在 crafts 子集上 PA 指标优于次好的 U2Net 约 1.59%,在 real 子集上,PA 指标优于次好的 U2Net 约 4.55%,mIoU 指标优于次好的 DeeplabV3+约 6.11%。在推理时间和计算效率上,本文方法相较于其他网络展现出较低的推理时间和适中的浮点运算量。

图 7 所示为这些方法的分割结果。从图 7 可以看出,FCN 和 PSPNet 在针对小目标缺陷时,无法分割。GCN、Unet 在复杂区域分割会出现空洞区域。U2Net 和 DeeplabV3+在某些情况下则会分割出余的缺陷区域。本文方法不论是在大面积缺陷还是分散的小面积缺陷情况下,均能够精确地分割出缺陷区域和背景区域。

表 2 与其他 SOTA 方法的 PA 和 mIoU 对比结果

Table 2 Comparison results of various methods

模型	NRSD-crafts		NRSD-real		推理时间/ms	浮点数/GFLOPs
	PA	mIoU	PA	mIoU		
FCN	0.795 5	0.645 5	0.612 2	0.529 5	25	190.7
Unet	0.790 6	0.653 2	0.686 5	0.588 7	73	91.6
U2net	0.823 3	0.687 4	0.694 2	0.589 7	87	92.1
PSPNet	0.765 6	0.588 2	0.620 5	0.519 7	67	179.5
GCN	0.524 6	0.676 7	0.650 0	0.576 2	70	120.8
DeeplabV3+	0.801 9	0.672 8	0.668 5	0.597 7	66	75.1
YOLO11-Seg	0.809 0	0.605 4	0.639 8	0.521 8	51	142.7
YOLO12-Seg	0.795 0	0.613 3	0.685 6	0.557 9	54	144.8
本文	0.836 4	0.685 8	0.725 8	0.634 2	60	93.1

3.5 消融实验

为了验证各模块对模型性能的有效性,本文进行了消融实验。通过移除或替换模型中的特定模块,观察其对模型性能的影响,从而确定每个模块在整体模型中的贡献。本文主要对多尺度特征融合方法、PAM 和 PPM

进行了消融实验。4 组实验模型结构的示意图如图 8 所示。图 8(a)为 Baseline 模型,仅包含特征编码器和基础解码器;图 8(b)在 Baseline 基础上加入了 PAM 模块;图 8(c)在 Baseline 基础上加入了 PPM 模块;图 8(d)则同时集成了 PAM 和 PPM 模块,并进行特征融合。

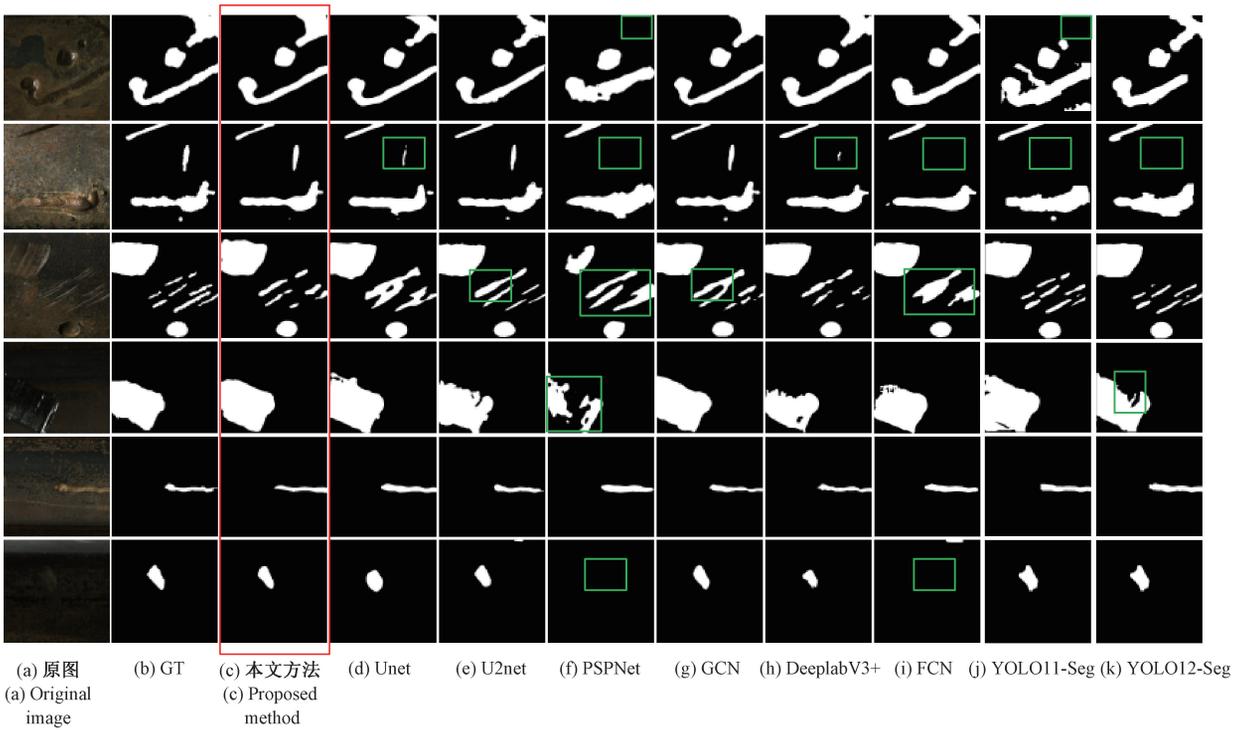


图 7 不同方法的分割结果比较

Fig. 7 Comparison results of different methods

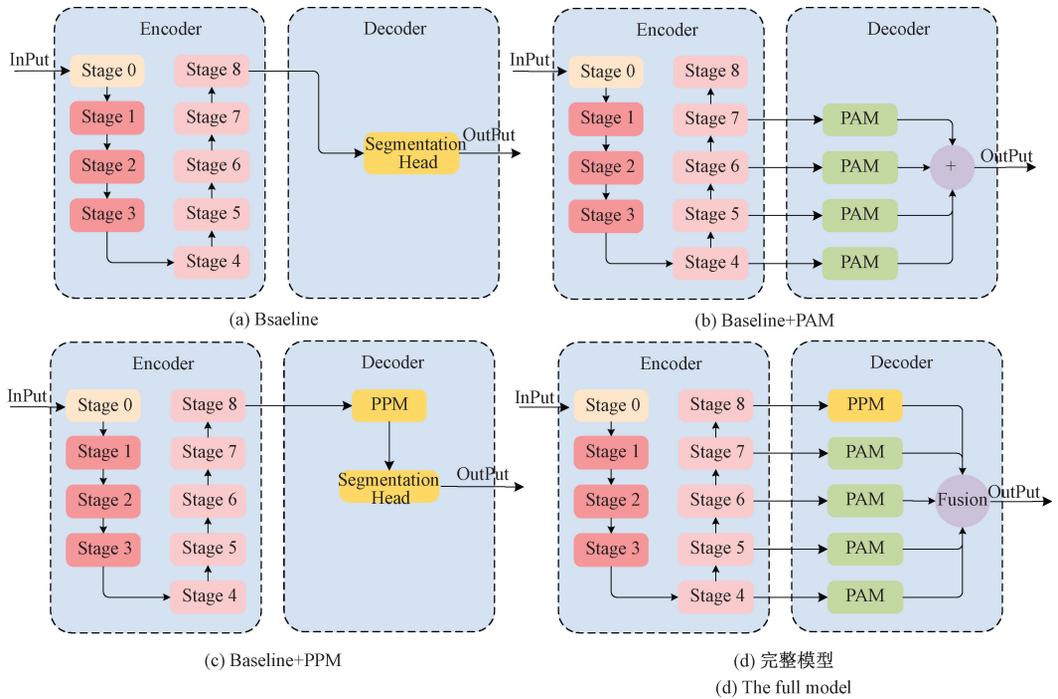


图 8 消融实验对比模型结构示意图

Fig. 8 Structures of ablation model

实验结果如表 3 所示。从表 3 来看,仅包含 PPM 模块的 Baseline+PPM 模型在 Craft 和 Real 子集上的表现较

差,PA 和 mIoU 指标均有所下降;而包含 PAM 模块的 Baseline+PAM 模型在 Craft 和 Real 子集的所有指标上均

表现出显著提升。此外,完整模型在解码器中融合了 PPM 和 PAM 的输出,并引入了多尺度特征融合方法,分割性能相较于 Baseline 显著提高。

表 3 消融实验结果

Table 3 Results of the ablation experiments

模型	NRSD-crafts		NRSD-real	
	PA	mIoU	PA	mIoU
Baseline	0.727 9	0.600 9	0.683 3	0.601 5
Baseline+PPM	0.704 0	0.585 0	0.609 9	0.550 6
Baseline+PAM	0.841 8	0.681 0	0.717 7	0.619 6
完整模型	0.836 4	0.685 8	0.725 8	0.634 2

具体分析其原因,Baseline+PAM 模型的 PPM 模块在捕获上下文信息方面未能被充分利用,导致模型对缺陷区域的关注不足,从而产生了错误的分割结果。而 Baseline+PPM 模型中引入的 PAM 注意力机制能够有效引导模型聚焦于目标区域,同时抑制图像中的噪声和无关特征,从而提升模型在 NRSD 复杂场景下的表现。对于完整模型,PAM 通过为目标区域的每个像素分配不同的注意力权重,使模型更加关注缺陷区域,并抑制背景噪声;PPM 则通过多尺度池化操作获取丰富的多尺度上下文信息,帮助模型解析局部和全局的语义信息。多尺度特征融合方法将来自 PAM 和 PPM 的输出融合,从而进一步增强模型的分割性能。

综上所述,PPM 提供了多尺度的上下文信息,PAM 突出重要区域,多尺度特征融合方法则有效整合这些信息,使模型在分割时拥有更丰富的特征表示。

3.6 注意力机制的选择

在设计本文的铁路轨道表面缺陷分割模型时,注意力机制模块的选择对模型分割性能起到了至关重要的作用。注意力机制模块通过为输入的数据分配不同的权重,使得模型更聚焦于重要信息而忽略无关信息。注意力机制模块的选择,需要综合考虑任务需求、计算资源、数据分布特点等因素,并通过实验验证。

为了找到更适合本方法的注意力机制模块,实验对比几种注意力机制模块:SE、ECA、CBAM、Non-Local、CA。在 NRSD 数据集上实验对比了各模块的效果,粗体表示最佳结果,实验结果如表 4 所示。PAM 在 5 个注意力模

表 4 不同注意力机制模块下的分割结果

Table 4 Results of different attention mechanisms

注意力机制	NRSD-crafts		NRSD-real	
	PA	mIoU	PA	mIoU
SE	0.830 6	0.693 0	0.690 1	0.622 7
ECA	0.824 9	0.689 7	0.697 2	0.625 6
CBAM	0.822 8	0.680 3	0.693 0	0.608 5
Non-Local Block	0.822 3	0.682 0	0.683 3	0.607 0
CA	0.832 5	0.691 2	0.707 9	0.624 5
PAM	0.836 4	0.685 8	0.725 8	0.634 2

块中,在数据集的 craft 和 real 两个部分都取得了很好的表现。

为了更直观的展示注意力机制模块对模型的影响,本文使用热力图可视化结果。如图 9 所示,颜色色调表示模型在不同区域的注意力强度,图像中的红色区域表示模型对这些区域的响应较强,蓝色或绿色区域表示响应较弱,显然不同的注意力机制对模型的聚焦能力有显著的影响。可以看到,PAM 在针对背景复杂的 NRSD 数据集时,更能有效捕捉图像中的关键区域,为缺陷的分割提供了更加准确的特征表示。

3.7 多级注意力机制融合方案

在解码器中,需要融合 4 个 PAM 模块的解码输出特征图。常规的特征图融合方法有逐元素相加、拼接等。为了更好地利用 PAM 模块的输出特征图,本研究尝试确定一种最优方案。图 10(a)为对 PAM 的输出进行逐元素相加融合,图 10(b)为对 PAM 的输出进行拼接融合处理,图 10(c)为在逐元素相加的基础上,进行改进,对每次逐元素相加融合后,进行一次卷积处理。

表 5 为 3 种不同方法在 NRSD 上的实验结果。显然,图 10(c)在 NRSD 分割中表现最佳。从实验结果分析,逐元素相加方案虽然直接地结合了不同特征图的信息,但是缺乏对特征的进一步提取和优化。拼接融合方案则没有充分考虑不同通道特征之间的关系,不同特征之间缺乏交互和优化。本文的方案最终在逐元素相加的基础上进行卷积处理,能够有效地对融合后的特征进行进一步的提取和优化,更好地捕捉到特征图中的关键信息。

表 5 不同融合方法下的分割性能

Table 5 Segmentation performance with different fusion methods

方法	NRSD-crafts			NRSD-real		
	PA	mIoU	Dice	PA	mIoU	Dice
Addition	0.835 9	0.681 7	0.797 3	0.713 8	0.623 5	0.720 0
Concatenation	0.817 1	0.653 7	0.775 2	0.675 1	0.577 4	0.676 0
本文	0.836 4	0.685 8	0.799 8	0.725 8	0.634 2	0.729 8

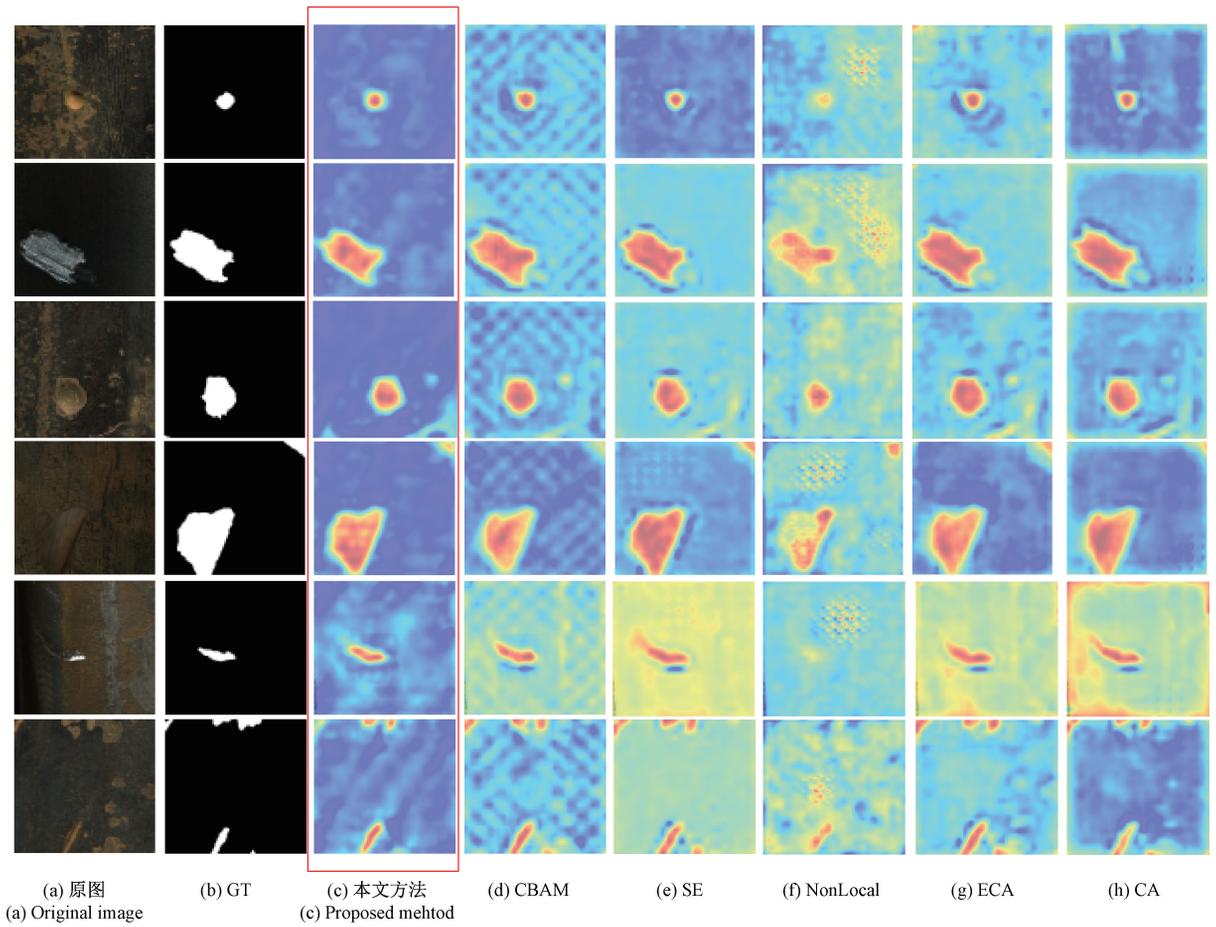


图 9 不同注意力机制作用下的可视化结果

Fig. 9 Visualization results of different attention mechanisms

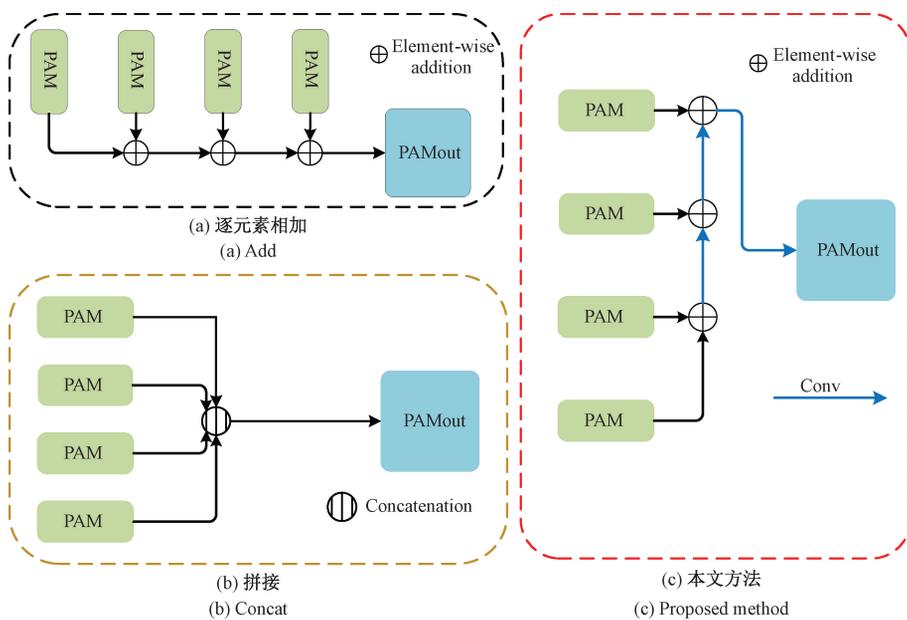


图 10 多级注意力机制融合方法

Fig. 10 Multi-level attention mechanism fusion methods

4 结 论

本文提出了一种基于多级并行注意力机制和多尺度空间信息融合的语义分割网络,并在铁路轨道表面缺陷分割任务中进行了有效验证。针对 NRSD 分割任务中缺陷纹理复杂、缺陷种类多样的挑战,模型通过高效编码器提取充分的特征;在解码器中引入多级并行注意力机制以进一步增强特征处理能力;最后通过多尺度上下文信息融合方法,实现了 PAM 和 PPM 提取特征的有效整合。实验结果表明,本文模型在多个评价指标上均取得了优异的性能,验证了多尺度空间信息融合与多级并行注意力机制在语义分割任务中的有效性。本文模型在一定的计算资源下取得了良好的效果,但其训练和推理过程对计算性能要求较高。因此,未来可以进一步研究轻量化的注意力机制和融合策略,以提升模型的效率和部署适应性。此外,由于铁路缺陷数据具有特定性,模型在其他应用场景中的推广性仍需探索,未来可以通过数据扩展和迁移学习增强模型的鲁棒性。

参考文献

- [1] 李杰,张楠,李智,等. 铁路轨道检测技术研究的进展[J]. 电子技术,2024,53(6):208-209.
LI J, ZHANG N, LI ZH, et al. Research progress in railway track inspection technology [J]. Electronics Technology, 2024, 53(6): 208-209.
- [2] 田贵云,高斌,高运来,等. 铁路钢轨缺陷损伤巡检与监测技术综述[J]. 仪器仪表学报,2016,37(8):1763-1780.
TIAN G Y, GAO B, GAO Y L, et al. Review of inspection and monitoring technologies for railway rail defect damage [J]. Chinese Journal of Scientific Instrument, 2016, 37(8): 1763-1780.
- [3] 张琳娜,岑翼刚. 铁轨图像的低秩矩阵分解缺陷检测[J]. 信号处理,2019,35(4):667-675.
ZHANG L N, CEN Y G. Defect detection of rail images using low-rank matrix decomposition [J]. Signal Processing, 2019, 35(4): 667-675.
- [4] 张祯祥. 基于图像特征的高速铁路轨道焊缝缺陷检测[J]. 现代城市轨道交通,2023(7):27-31.
ZHANG ZH X. High-speed railway rail weld defect detection based on image features [J]. Modern Urban Rail Transit, 2023(7): 27-31.
- [5] 刘传洋,吴一全,刘景景. 基于视觉的输电线路金具锈蚀缺陷检测方法研究进展[J]. 仪器仪表学报,2024,45(3):286-305.
LIU CH Y, WU Y Q, LIU J J. Research progress on visual-based detection methods for corrosion defects of power transmission line fittings [J]. Chinese Journal of Scientific Instrument, 2024, 45(3): 286-305.
- [6] 周建新,李忠泽,郝英杰. 基于改进 YOLOv9 的钢板表面缺陷检测的方法[J]. 电子测量技术,2024,47(22):181-188.
ZHOU J X, LI ZH Z, HAO Y J. Steel surface defect detection method based on improved YOLOv9 [J]. Electronic Measurement Technology, 2024, 47(22): 181-188.
- [7] 梁海波,王怡. 基于深度学习的天然气钢制管道缺陷检测方法研究[J]. 电子测量与仪器学报,2022,36(9):148-158.
LIANG H B, WANG Y. Research on defect detection method for natural gas steel pipelines based on deep learning [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(9): 148-158.
- [8] 王勇,贺锐,蔡志贤. 基于计算机视觉技术的轨道缺陷检测研究与应用[J]. 工程技术研究,2022,7(23):13-16.
WANG Y, HE R, CAI ZH X. Research and application of track defect detection based on computer vision technology [J]. Engineering Technology Research, 2022, 7(23): 13-16.
- [9] 马茜. 基于图像识别技术的轨道交通缺陷检测研究[J]. 计算技术与自动化,2022,41(1):117-122.
MA X. Research on track defect detection based on image recognition technology [J]. Computer Technology and Automation, 2022, 41(1): 117-122.
- [10] 韦若禹,李舒婷,吴松荣,等. 基于改进 YOLO V3 算法的轨道扣件缺陷检测[J]. 铁道标准设计,2020,64(12):30-36.
WEI R Y, LI SH T, WU S R, et al. Track fastener defect detection based on improved YOLOv3 algorithm [J]. Railway Standard Design, 2020, 64(12): 30-36.
- [11] 王德伟,刘小芳. 基于改进 YOLOv8n 的带钢表面缺陷检测方法[J]. 国外电子测量技术,2024,43(7):158-169.
WANG D W, LIU X F. Steel strip surface defect detection method based on improved YOLOv8n [J]. Foreign Electronic Measurement Technology, 2024, 43(7): 158-169.
- [12] 黄啸,吴龙,黎尧,等. 基于改进 YOLOX-s 的轻量级型钢表面缺陷检测算法[J]. 计算机应用,2023,43(S2):

- 201-208.
- HUANG X, WU L, LI Y, et al. Lightweight steel surface defect detection algorithm based on improved YOLOX-s[J]. *Computer Applications*, 2023, 43(S2): 201-208.
- [13] 付宇彤,李敏,黄及远,等. 铸件缺陷语义分割的编码器-解码器网络[J]. *制造业自动化*, 2023, 45(5): 207-214.
- FU Y T, LI M, HUANG J Y, et al. Encoder-decoder network for cast defect semantic segmentation [J]. *Manufacturing Automation*, 2023, 45(5): 207-214.
- [14] LI G, HAN C, LIU Z. No-service rail surface defect segmentation via normalized attention and dual-scale interaction[J]. 2023, DOI:10.48550/arXiv.2306.15442.
- [15] 任中伟,方维维,许文元,等. 面向轨道缺陷检测的联邦学习轻量化模型训练技术研究[J]. *铁道学报*, 2023,45(4):77-83.
- REN ZH W, FANG W W, XU W Y, et al. Research on federated learning lightweight model training technology for track defect detection[J]. *Journal of Railway Science and Engineering*, 2023, 45(4): 77-83.
- [16] ZHANG D, SONG K, WANG Q, et al. Two deep learning networks for rail surface defect inspection of limited samples with line-level label [J]. *IEEE Transactions on Industrial Informatics*, 2020, 17(10): 6731-6741.
- [17] 陈科圻,朱志亮,邓小明,等. 多尺度目标检测的深度学习研究综述 [J]. *软件学报*, 2021, 32(4): 1201-1227.
- CHEN K Q, ZHU ZH Q, DENG X M, et al. A review of deep learning for multi-scale object detection[J]. *Journal of Software*, 2021, 32(4): 1201-1227.
- [18] 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.
- [19] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]. *Medical Image Computing and Computer-assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer International Publishing, 2015: 234-241.
- [20] PASZKE A, CHAURASIA A, KIM S, et al. Enet: A deep neural network architecture for real-time semantic segmentation [J]. *ArXiv preprint arXiv: 1606.02147*, 2016.
- [21] LIU N, HAN J, YANG M H. Picanet: Learning pixel-wise contextual attention for saliency detection [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 3089-3098.
- [22] 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.
- [23] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 3-19.
- [24] QIN X, ZHANG Z, HUANG C, et al. U2-Net: Going deeper with nested U-structure for salient object detection [J]. *Pattern Recognition*, 2020, 106: 107404.
- [25] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 801-818.
- [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] PENG C, ZHANG X, YU G, et al. Large kernel matters—improve semantic segmentation by global convolutional network [C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017: 4353-4361.
- [28] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements [J]. *ArXiv preprint arXiv:2410.17725*, 2024.
- [29] TIAN Y, YE Q, DOERMANN D. YOLOv12: Attention-centric real-time object detectors [J]. *ArXiv e-prints, arXiv: 2502.12524*, 2025.

作者简介



周炜杰,2022年于天津职业技术师范大学获得学士学位,现为桂林电子科技大学硕士研究生,主要研究方向为智能感知技术与仪器。

E-mail: vijayz@foxmail.com

Zhou Weijie received his B. Sc. degree from Tianjin University of Technology and Education in 2022. He is now a M. Sc. student at Guilin University of Electronic Technology. His main research interests include intelligent sensing technology and instruments.



李智, 1987 年于东南大学获得学士学位, 1993 于上海理工大学获得硕士学位, 2003 年于电子科技大学获得博士学位, 现为教授, 博士生导师, 桂林航天工业学院副校长, 主要研究方向为自动测试总线与系统、现代测试理论与技术。

E-mail: cclizhi@guat.edu.com

Li Zhi received his B. Sc. degree from Southeast University in 1987, M. Sc. degree from Shanghai University of Science and Technology in 1993, and Ph. D. degree from University of Electronic Science and Technology in 2003. He is now a professor, doctoral supervisor and vice president of Guilin Institute of Aerospace Science and Technology. His main research interests include automatic test bus and system and modern test theory and technology.



张绍荣(通信作者), 2021 年于桂林电子科技大学获得博士学位, 深圳大学博士后。现为桂林航天工业学院副教授, 桂林电子科技大学硕士生导师。主要研究方向为脑机接口与生理监测技术、智能感知与仪器系统。

E-mail: zsrong@guat.edu.cn

Zhang Shaorong (Corresponding author) received his Ph. D. from Guilin University of Electronic Technology in 2021 and completed his postdoctoral research at Shenzhen University. He is now an associate professor at Guilin University of Aerospace Technology and a master's supervisor at Guilin University of Electronic Technology. His main research interests include brain-computer interface and physiological monitoring technology, intelligent perception and instrumentation systems.