

doi: 10.3969/j.issn.1672-6073.2024.05.011

# 钢轨表面缺陷检测 Mask R-CNN 算法研究与优化

孟瑞锋<sup>1</sup>, 梁 桢<sup>1</sup>, 贾 超<sup>2</sup>, 乔 志<sup>3</sup>, 赵 晨<sup>1</sup>

(1. 内蒙古工业大学航空学院, 呼和浩特 010051; 2. 内蒙古灵奕高科技(集团)有限责任公司, 呼和浩特 010010;  
3. 内蒙古高速公路集团有限责任公司, 呼和浩特 010051)

**摘要:** 为有效防止城市轨道交通事故发生, 更好地保障运行安全, 钢轨表面缺陷检测技术在巡检工作中发挥着重要作用。针对现有钢轨缺陷检测技术中检测精度差、小目标敏感度低等问题, 在 Mask R-CNN(mask region-based convolutional neural network)算法模型基础上, 提出一种融合注意力机制的模型改进方案。该方案在特征提取网络中引入通道-空间复合注意力机制(channel-wise spatial module, CSM)用于实例分割缺陷检测, 有效剔除干扰信息, 获得多尺度特征表达, 得到更多空间信息以及更优的浅层信息, 从而提升对钢轨表面缺陷边缘检测能力。在相同的实验环境下, 相对于 Mask R-CNN 算法, 加入 CSM 后, Mask R-CNN 模型的平均精度均值(mean average precision, *mAP*)提高了 6.5%。其中, 对钢轨“凹陷”“裂纹”以及“疲劳磨损”缺陷识别的平均精度(average precision, *AP*)分别提高了 6.3%、6.9%和 6.1%。横向对比发现, 加入 CSM 后的 Mask R-CNN 模型, 相较于 Fast R-CNN 模型, 三种缺陷的分割效果分别提升了 11.6%、12.5%和 12.9%。同时, 相较于 Faster R-CNN 模型, 三种缺陷的分割效果分别提升了 8.8%、10.0%和 10.3%。加入 CSM 后的 Mask R-CNN 模型可以更好地识别三类缺陷, 提升检测精度和小目标敏感度, 为轨道智能巡检提供更安全有力的技术支持和保障。

**关键词:** 城市轨道交通; 通道-空间注意力机制; 钢轨缺陷; 实例分割; Mask R-CNN 算法

中图分类号: U231

文献标志码: A

文章编号: 1672-6073(2024)05-0068-10

## Study and Optimization of Mask R-CNN Algorithm for Surface Defect Detection on Railway Tracks

MENG Ruifeng<sup>1</sup>, LIANG Zhen<sup>1</sup>, JIA Chao<sup>2</sup>, QIAO Zhi<sup>3</sup>, ZHAO Chen<sup>1</sup>

(1. School of Aeronautics, Inner Mongolia University of Technology, Hohhot 010051; 2. Inner Mongolia Lingyi High-Tech Group Co., Ltd., Hohhot 010010; 3. Inner Mongolia Expressway Group Co., Ltd., Hohhot 010051)

**Abstract:** Surface defect detection technology plays an essential role in railway inspections by effectively preventing railway accidents and ensuring operational safety. This study addresses the issues of poor detection accuracy and low sensitivity to small targets in existing railway defect detection technologies. For the mask region-based convolutional neural network (Mask R-CNN) algorithm model, a model improvement scheme was proposed by incorporating an attention mechanism. This scheme introduces a Channel-wise Spatial Module (CSM) into the feature extraction network for segmentation defect detection, effectively eliminating interference, obtaining multiscale feature representations, and acquiring more spatial and shallow information, thereby enhancing the edge detection capability for surface defects on railway tracks. In the same experimental environment, compared with the Mask R-CNN algorithm, after adding the CSM, the *mAP* value of the Mask R-CNN model increased by

收稿日期: 2023-12-26 修回日期: 2024-05-23

第一作者: 孟瑞锋, 男, 博士, 副教授, 研究方向为智能交通、深度学习, mrfnmgn@imut.edu.cn

基金项目: 内蒙古自治区科技计划项目(2022YFSJ0040); 内蒙古自治区直属高校基本科研业务费项目(JY20220170)

引用格式: 孟瑞锋, 梁桢, 贾超, 等. 钢轨表面缺陷检测 Mask R-CNN 算法研究与优化[J]. 都市轨道交通, 2024, 37(5): 68-77.

MENG Ruifeng, LIANG Zhen, JIA Chao, et al. Study and optimization of Mask R-CNN algorithm for surface defect detection on railway tracks[J]. Urban rapid rail transit, 2024, 37(5): 68-77.

6.5%。Among them, the AP values for the recognition of “depression,” “crack,” and “fatigue wear” defects on railway tracks increased by 6.3%, 6.9%, and 6.1%, respectively. The horizontal segmentation effects of the three defects improved by 11.6%, 12.5%, and 12.9%, respectively, compared with the Faster R-CNN model, and the segmentation effects of the three defects enhanced by 8.8%, 10.0%, and 10.3%, respectively. This study demonstrates that the Mask R-CNN model with CSM can better recognize three types of defects, enhance the detection accuracy and sensitivity to small targets, provide more secure and robust technical support, and guarantee intelligent track inspection.

**Keywords:** urban rail transit; channel-wise spatial attention mechanism; rail defects; instance segmentation; mask R-CNN algorithm

我国的城市轨道交通飞速发展, 钢轨表面缺陷检测与识别是目前轨道运维作业中重要环节。《中华人民共和国铁道行业标准》<sup>[1]</sup>针对钢轨表面缺陷指出缺陷分类, 主要有磨损、裂纹、压陷和剥落。这些缺陷可能威胁列车安全行驶并且影响钢轨的正常使用。有缺陷的钢轨将导致列车运行不稳、加速设备损伤, 甚至引发严重事故, 直接威胁其正常运营。为避免钢轨缺陷带来的风险, 轨道交通领域引进了各种检测技术, 美国纽约 KLD(knowledge lab for digital finance)实验室和 Beena Vision 公司研发的钢轨检测系统主要使用超声波、磁粉探伤和光学成像等无损检测技术, 能够检测钢轨的内部和表面缺陷, 包括裂纹、夹杂物等。瑞士 ELAG 公司研发的 Railmonitor 检测器结合声学传感器, 可实现检测包括磨耗、裂纹、疲劳等多种缺陷在内的钢轨整体状态, 检测效果明显, 检测技术成熟<sup>[2]</sup>。

早期钢轨缺陷检测主要依赖人工物探, 效率低且缺乏统一标准, 逐渐被其他方法取代。随后, 超声波、射线、渗透和涡流等无损探伤技术提高了检测精度和速度<sup>[3]</sup>, 这些方法虽穿透力强、操作安全, 却易受外部干扰, 结果抽象且难处理。

基于机器视觉的钢轨缺陷检测方法通过先进的视觉设备采集钢轨表面图像, 根据算法对图像进行处理, 具有实时性、非接触式等特点, 能够很好地运用于钢轨缺陷检测领域, 闵永智<sup>[4]</sup>提出将平滑滤波器与阈值分割相结合的钢轨表面缺陷检测方法, 减轻了光照变化、轨面不平对检测结果的影响, 但该方法对背景图像的自适应平滑过程运算量过大, 实时性不强。

刘泽等<sup>[5]</sup>设计了动态阈值分割算法和缺陷区域提取算法, 该算法能够准确提取缺陷区域, 标定缺陷位置并统计缺陷特征。上述研究将传统图像处理技术与机器学习的方法相结合, 设计了适用于特定场景下的钢轨缺陷检测方法, 该类方法的检测性能易受外部环境的影响, 检测速度难以满足实时检测要求。

与传统图像处理技术相比, 深度学习能够从采集的钢轨图像中学习钢轨缺陷的轮廓及语义规律, 从而

对缺陷进行智能检测, 具有较强的泛化能力和鲁棒性。罗晖等<sup>[6]</sup>提出一种基于图像增强与深度学习的钢轨表面缺陷视觉检测算法, 对图像进行去噪, 并使用颜色空间变换方法, 最终基于空洞卷积来改进 Faster R-CNN (faster region-based convolutional neural) 卷积神经网络, 有效提高了检测准确率, 但检测速度较慢, 难以满足实时要求。苏焯等<sup>[7]</sup>提出基于 Faster R-CNN 的缺陷检测方法, 首先将预处理后的图像进行反转, 利用投影变换实现钢轨图像的投影, 然后通过区域建议网络提取候选区域, 完成检测。翟鑫等<sup>[8]</sup>将新的特征提取结构与 Mask R-CNN 相结合, 强化图像边缘处的精度, 完成铁路区域进行识别分割。马宇超等<sup>[9]</sup>采用迁移学习权重, 并在 Mask R-CNN 的主干特征提取网络中加入自适应层, 从而提升检测效率和精度。

针对工业表面缺陷多样化难以精确检测的问题, 王海云等<sup>[10]</sup>提出改进特征金字塔网络(feature pyramid network, FPN)的 Mask R-CNN 表面缺陷检测算法, 该算法增加自底向上的反向连接路径, 对各阶段特征图进行充分融合, 得到具有更强语义信息和定位信息的多尺度特征图, 平均精度均值可达到 98.25%。

上述研究对两阶段实例分割算法进行改进后, 模型的检测精度得到一定的提高, 但存在检测速度慢、预处理步骤复杂、边缘检测精度不足以及模型的泛化性差等问题。因此, 本文针对常见钢轨缺陷, 开展图像智能识别技术研究。相较于原 Mask R-CNN 模型识别精度欠佳和小目标敏感度低的问题, 探索通道-空间复合注意力机制在提升模型分割性能方面的机理和作用效果, 以期提高钢轨缺陷识别的能力。

## 1 结合 CSM 的 Mask R-CNN 算法

### 1.1 Mask R-CNN 算法

目前 Mask R-CNN 算法广泛应用于医学、农业、工业、交通、信息等领域。Mask R-CNN 除识别边界框外, 还可预测目标掩码, 可有效检测目标对象的同时, 对每个实例生成高质量的分割掩模。Mask R-CNN

的主干网络采用了特征金字塔网络 FPN 和残差神经网络(residual neural network, ResNet)的结合。在原始特征金字塔网络中添加  $3 \times 3$  卷积。从特征图像上取样并和分辨率更高的底层特征图像横向链接,在空间大小一致的情况下结合,使混合后的特征图像在特征定位上表现更出色。

Mask R-CNN 使用轻量的区域候选网络(region proposal network, RPN)获取候选区域,对每块扫描获得的锚(anchor-box)输出前景或背景类别,并同时评估输出精细调整前景锚框与目标中心位置关系;在多个锚框重叠的情况下使用非极大值抑制(non-maximum suppression, NMS)选择前景分数最高锚框,最终输出图片内最佳的候选区域。

Mask R-CNN 算法执行过程是,第 1 步,对输入图片进行尺寸归一化处理,将结果导入预训练神经网络中 ResNet 等获得对应的特征图;第 2 步,设定候选区域(region of interest, ROI),得到多个相符的区域候选框,并将获得的多个候选框送入区域候选网络做二值分类和边界框回归处理,剔除掉一部分候选区域的回归框;第 3 步,对筛选后较优的 ROI 候选框进行对齐(ROI Align)、分类、回归操作,最终所有 ROI 送入到全卷积神经网络(fully convolutional networks, FCN),生成掩膜 mask。

## 1.2 注意力机制

注意力机制是实现网络自适应注意的方法之一,在计算机视觉领域,注意力机制的应用尤其广泛<sup>[11]</sup>。

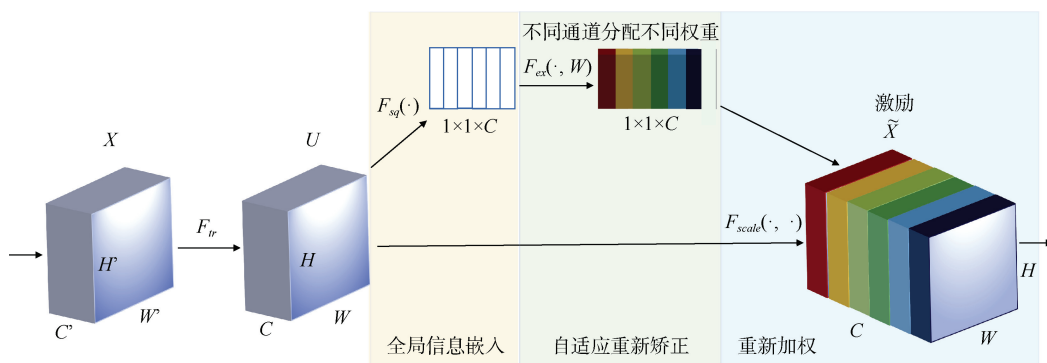
最初的研究目的是解决深度神经网络(deep neural network, DNN)在计算机视觉任务中如何从所有输入特征角度入手来提升计算效率的问题<sup>[12]</sup>。神经网络中的注意力机制<sup>[13]</sup>可以解决信息超载问题,它的实质是将更重要的运算任务优先级提高。

对改进 R-CNN 系列算法和单次检测系列算法的成功案例进行分析<sup>[14]</sup>,添加注意力机制是目前可行的算法优化手段。为改善原 Mask R-CNN 存在边界框回归的收敛较慢和边缘检测性能欠佳的问题,引入注意力机制<sup>[15]</sup>,提高模型的收敛速度和对小目标的识别敏感度。

### 1.2.1 通道注意力机制

通道注意力机制是一种机器学习技术,其模型将自主学习一组权重,用于计算输入数据中每个特征的重要性。通道注意力机制的网络结构通常由一个编码器和一个解码器组成,编码器用于提取输入信息的特征,解码器根据此特征生成相应的输出。注意力机制用来实现编码器和解码器之间的连接,并控制编码器和解码器之间的信息流,提高网络性能。

2017 年提出的压缩和激励网络(squeeze and excitation net, SENet)是通道注意力机制的经典模型(见图 1)。SENet 重点是获得输入进来的特征层中每一个通道的权值<sup>[16]</sup>。SENet 可使网络关注它最需要关注的通道,使得靠近数据输入的特征也可以具有全局感受野,通过全连接网络的处理,在降低了网络计算量的同时,也增加了网络的非线性能力。



$H$  为特征图宽度,  $W$  为特征图高度,  $C$  为通道数,  $X$  为输入特征,  $U$  为经  $F$  函数处理后特征

图 1 通道注意力机制经典模型

Figure 1 The classic model of channel attention mechanism

其实现过程如图 1 所示。网络对输入进来的特征层进行全局平均池化(GAP)后,进行两次全连接;再接上述过程结果做一次 sigmoid 激活运算,获得所述输入特征层每一个通道的权值(0~1 之间);最后,只

需将该权重与原始输入特征层进行逐元素相乘即可。

### 1.2.2 空间注意力机制

空间注意力机制是一种在深度学习模型中使用的技术,用于提高模型对输入数据中不同位置的关注程

度。不同于通道注意力，空间注意力(spatial attention mechanism, SAM)聚焦于特征图上有效信息的位置<sup>[17]</sup>，如图 2 所示。

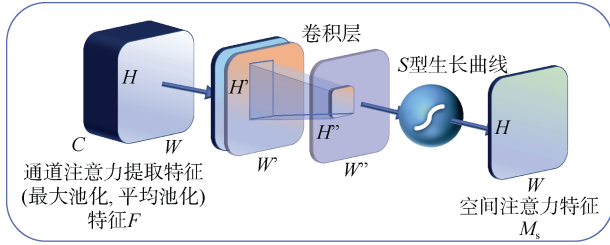


图 2 空间注意力模块 SAM

Figure 2 Spatial Attention Module (SAM)

空间注意力模型通过将通道中注意力模型输出的特征图  $F$  作为本模型的输入特征图，而特征图  $H \times W \times C$  则经过基于通道的最大池化方法(MaxPool)和平均池化(AvgPool)获得了 2 个  $H \times W \times 1$  的特征图，基于通道的拼接；在经过  $7 \times 7$  卷积操作之后降维到  $H \times W \times 1$ ，然后再经过 sigmoid 生成空间的注意力特征；最后再对该向量与该模型中的输入特征图做乘运算，以得出最终的特征图。其过程运算方法如下：

$$M_s(F) = \sigma\left(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])\right) \quad (1)$$

$$= \sigma\left(f^{7 \times 7}([F_{avg}^s; F_{max}^s])\right)$$

式中， $\sigma$  为激活函数 sigmoid。其优势在于：参数量和计算量都能够通过输入图像的尺寸进行控制，忽略杂乱图像中的噪声部分，网络对输入图像采样的尺度可控，以适应不同尺寸的图像。经上述分析，空间注意力机制是有利于 Mask R-CNN 网络降低参数处理的体量，精准提取特征的。

### 1.2.3 结合 CSM 的 Mask R-CNN

混合注意力机制的代表模型是卷积注意力模块(convolutional block attention module, CBAM)，其中包括通道注意力模块(channel attention module, CAM)和空间注意力模块 SAM。CAM 和 SAM 可以通过串联或者并联的方式进行组合，并且可以嵌入多尺度模块，实现准确有效的特征提取。CBAM 模型将输入特征分别经过 CAM 模块和 SAM 模块处理，最终获得调整后特征(见图 3)。

通过以上论述，结合通道注意力和空间注意力网络优势，本文使用通道-空间复合注意力机制，添加到 Mask R-CNN 算法结构中，来达成改进算法预期效果。

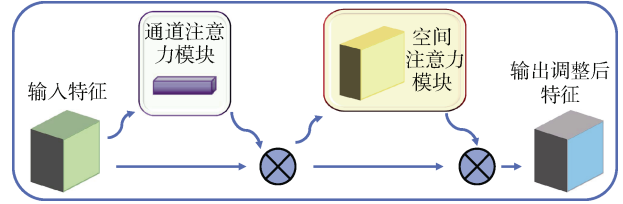


图 3 卷积网络注意力模块 CBAM

Figure 3 Convolutional Block Attention Module (CBAM)

通道-空间复合注意力模型<sup>[18]</sup>CSM 基于 VGG-16 框架，去掉了 VGG-16 末尾的全连接层，最终得到更契合图像空间特征的全卷积网络<sup>[19]</sup>。其输出结构把主干网络中 5 个层的特征提取结果输入空洞卷积(dilated convolution)，获得多尺度特征提取(multi-scale feature extraction, MSFE)模块，从而得到更多的空间信息<sup>[20]</sup>。由  $3 \times 3$  的感受野进行归一化卷积操作，输出特征图。依照语义细节信息的互补知识和卷积层间的信息性质差异，通过单向反馈传递的方式将主干网络中由 5 个通道-空间复合注意力机制加权融合获得的特征图传送到浅层中，从而获得优化的浅层信息，得到最终结果图。主干网络提取的 5 种不同尺寸的特征图，能够从不同维度(空间、尺寸、通道等)进行特征提取后的叠加，获取更为丰富的特征信息。

模型中采用空洞卷积并联进行上下文特征提取，形成特征图。如式(2)所示， $\text{Conv}(d_i)$  为空洞卷积运算。

$$Y = \sum_{i=1}^k \text{Conv}(d_i) \quad (2)$$

在多尺度特征提取模块之后插入“ $3 \times 3$  卷积层-Batch Normalization 卷积层- $3 \times 3$  Rectified Linear Unit 层”(见图 4)，输出特征图  $T$ ，并与后一个特征图  $T$  经通道-空间复合注意力模块取得的特征图  $S_i$  与上一卷积层取得的  $T$  链接，形成上一个通道-空间注意力模块的输入，最终对结果进行归一化。反复上述过程，形成单向反馈传递，对得到的  $S_i$  进行上采样操作，最终形成特征图  $S$ 。

本文改进方案采用添加通道-空间注意力机制的方法，通道注意力模块采用经典的 SENet，空间注意力模块采用 SAM，通过将二者混合，可以剔除干扰信息，增强模型的泛化能力，并精准提取到有效特征信息。基于 Mask R-CNN 和通道-空间复合注意力机制的改进后算法，如图 5 所示。将该复合注意力机制模型分别添加在整体网络模型第一层卷积之后和最后一层卷积之后。图像不同区域的像素点经头部网络进入卷积网络注意力模块，屏蔽初次卷积的噪声特征，对特征参数量和计算量进行控制，并对每一个通道的权重

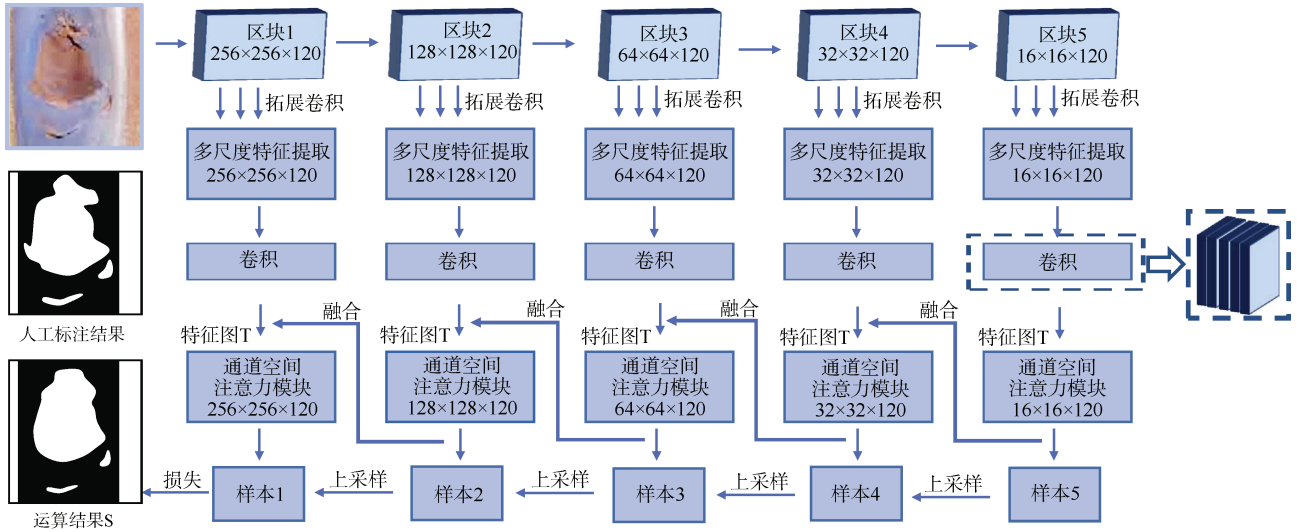


图 4 通道-空间注意力模块  
Figure 4 Channel-Spatial Attention Module

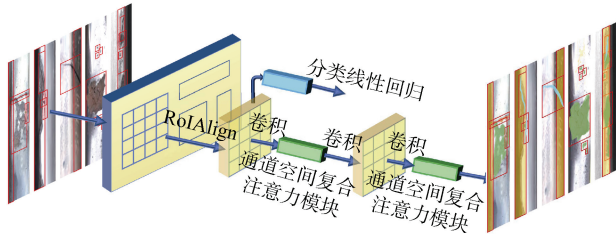


图 5 改进的 Mask R-CNN 整体框架

Figure 5 The enhanced framework of Mask R-CNN

进行关注, 以此更好收敛识别钢轨缺陷的锚框(Anchor Box), 对其边缘特征更加敏感。

## 2 实验及数据分析

### 2.1 实验环境配置

Mask R-CNN 算法训练平台采用 15G 显存 NVIDIA TeslaT4 GPU, 12 核心 Xeon Gold 6271 CPU, 64 位系统 Ubuntu18.04。环境配置为 Python3.7, 深度学习框架为 TensorFlow1.14, 初始学习率为 0.02、迭代次数为 200,  $IoU$  阈值为 0.5。TensorFlow2.x 框架使用 Python3.8, 其余环境配置与 TensorFlow1.14 相同。主干网络选取 ResNet-50, 它是 ResNet 系列中比较轻量级的模型之一, 包含了 50 个卷积层, 具有较低参数数量和计算复杂度。相比于 ResNet-101, ResNet-50 更适合于本文计算资源受限的情况。

### 2.2 缺陷数据集

通过自主采集和网络获取两种途径获得数据集, 数据集中各类别样例如图 6 所示。数据集预处理分为两步, 首先进行筛选并剔除坏样本, 然后进行数据增强。

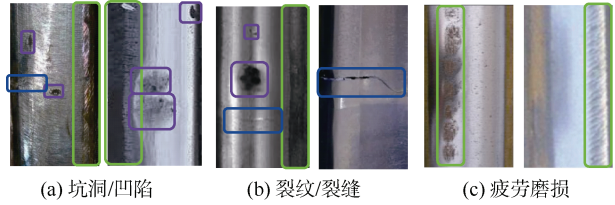


图 6 钢轨表面缺陷实例  
Figure 6 Instances of rail surface defects

在原始数据集中的损坏样本(指实例样本扭曲畸变无法准确归类识别的样本)和模糊样本(指实例样本无法达到算法训练对像素要求的样本)均需要手动进行剔除, 损坏样本和模糊样本如图 7 所示。

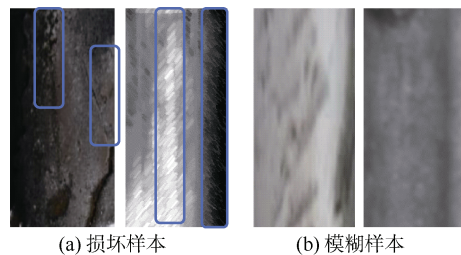


图 7 损坏样本和模糊样本  
Figure 7 Damaged samples and blurred samples

经筛选剔除后, 原始数据集涵盖 3 类缺陷, 每类 2 500 张, 均由位深度为 8 的 RGB 图片组成。通常在标注时会出现样本标注标准不统一、混标、漏标的情况, 如图 8 所示。本文采取规范的数据标注流程, 得到更准确的数据标签。

随后, 将筛选好的原始数据集进行人工标注并通过调整图片的亮度、对比度, 以及锐化和增加噪声的

方法, 扩充原始数据集, 扩充后的数据集如图 9、图 10 所示。将由 lablème 工具标注生成的 json 通过数据增强代码将其扩充, 并使标注文件名称与数据增强后的图片相互对应。

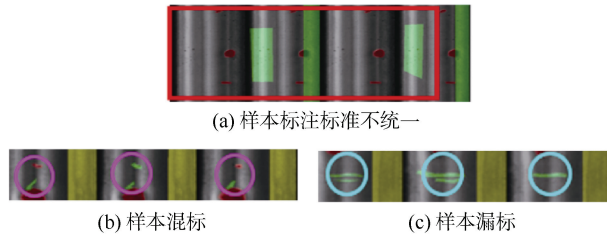


图 8 样本的标注  
Figure 8 Labeling of samples

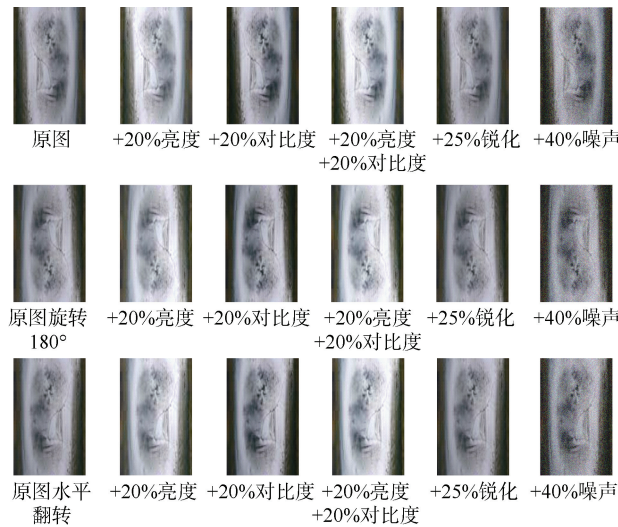


图 9 原始图片的数据增强  
Figure 9 Data enhancement of the original image

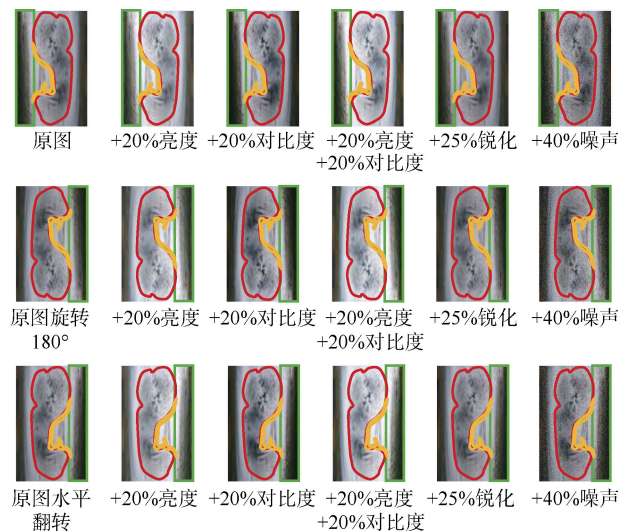


图 10 对应标注图片的数据增强  
Figure 10 Data enhancement corresponding to annotated images

最终得到符合训练要求的 COCO 格式数据集图片数量为 22 500 张。经上述方法处理后, 可进一步增强原始数据集的多样性。

### 2.3 实验评估指标

1) 实例分割的评价指标为: 准确率(precision,  $P$ )、召回率(recall,  $R$ )、平均准确率(average precision,  $AP$ )、 $P$ - $R$  曲线(precision-recall Curve)、平均精度均值(mean average precision,  $mAP$ ), 其中,  $P$ - $R$  曲线围成的面积为  $AP$ ,  $AP$  为每个单一缺陷类别识别的平均准确率,  $mAP$  为所有缺陷类别识别准确率的均值。上述评估指标计算如下:

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \int_0^1 P(R) dR \quad (5)$$

$$mAP = \frac{\sum_{i=1}^{N_c} AP_i}{N_c} \quad (6)$$

式(3)、(4)中, 正确检测框  $TP$ (true positive)表示预测框正确地与标签框匹配; 误检框  $FP$ (false positive)表示将背景预测成了物体; 漏检框  $FN$ (false negative)表示需模型检测出的物体, 模型没有检测出; 正确背景  $TN$ (true negative)表示目标是背景, 模型没有检测出。式(5)中  $AP$  需要同时考虑准确率(Precision)和召回率(Recall), 式(6)中  $N_c$  代表类别检测数。

2) 损失函数(loss function)是用来评估模型的预测值与真实值的差异, 损失函数表示为  $L_{mask}$ , 损失函数越小, 说明模型的预测与真实值越接近, 模型的检测精度越好。损失函数的公式如下:

$$L = L_{cls} + L_{bbox} + L_{mask} \quad (7)$$

即每个候选区域上的多任务损失函数等于分类损失  $L_{cls}$ 、边框回归损失  $L_{bbox}$ 、掩码损失  $L_{mask}$  之和。

### 2.4 钢轨表面三种缺陷识别结果分析

#### 2.4.1 实验结果

改进前后训练过程中  $loss$  变化曲线如图 11 所示, 从图中分析得出: 原 Mask R-CNN 算法在训练过程中  $loss$  曲线震荡严重, 在依次加入通道 ca、空间 sa 以及混合域 ours 注意力机制后, 随着训练轮数的增加, 训练过程中的损失值逐步下降, 最终稳定在 0.08, 说明模型学习到了更为准确的特征表示, 并且在整个训练过程中, 模型的收敛速度加快, 具体表现为,  $loss$  值

在 0~1 000 步时下降幅度相对于原模型更大。

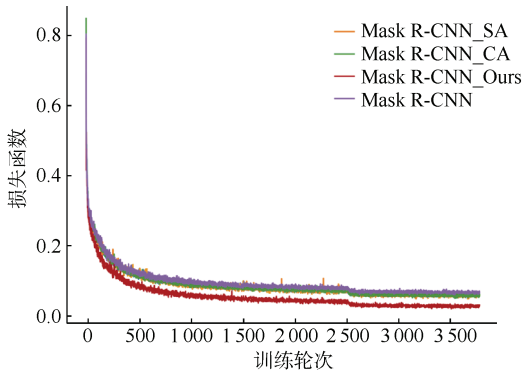


图 11 损失函数变化曲线  
Figure 11 Loss function evolution curve

改进前后模型的  $mAP$  变化曲线如图 12 所示,从图中可以看出,加入复合注意力机制后,Mask R-CNN\_Ours 模型在  $IoU$  阈值设置为 0.5 时,模型的  $mAP$  获得了显著提升,上升幅度约为 6.5%。

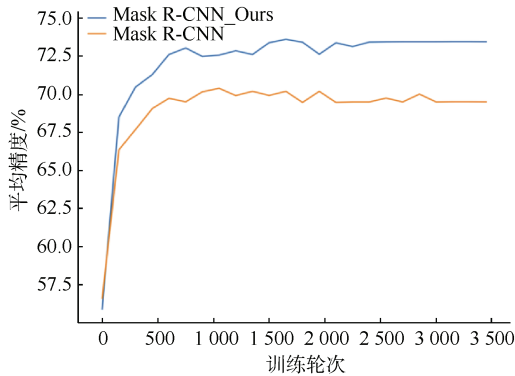


图 12  $mAP$  变化曲线  
Figure 12 The curve of  $mAP$  variation

改进前后  $P$ - $R$  曲线如图 13 所示,从图中可以看出,改进后的模型曲线将原 Mask R-CNN  $P$ - $R$  曲线完全

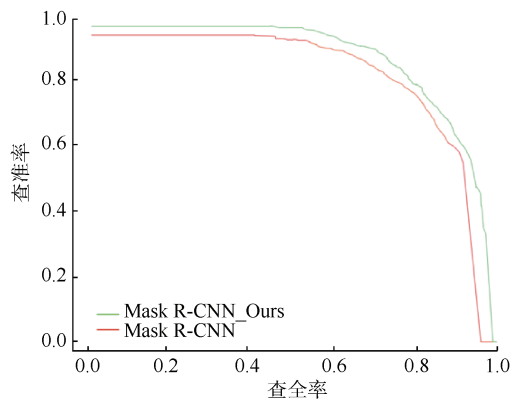


图 13  $PR$  曲线  
Figure 13  $PR$  curve (Precision-Recall curve)

包围,说明改进后模型 Mask R-CNN\_Ours 在各个类别的  $AP$  值都获得了提升。

### 2.4.2 实验过程分析

整体实验设计分为 3 步:①探究不同版本的 TensorFlow 框架对模型训练是否有影响,同时探究加入通道和空间注意力先后顺序对模型检测精度的影响;②通过消融实验验证单独加入通道、空间以及混合域注意力机制对模型的影响效果;③通过横向对比其他算法,验证改进后模型的效果。

表 1 为在不同 TensorFlow 版本下进行模型训练所获得的数据。由表 1 可知,Mask R-CNN 实例分割网络在 TensorFlow1.X 和 TensorFlow2.X 版本下的模型精度在坑洞、裂纹以及疲劳磨损类别的  $AP$  值均没有明显变化。这表明深度学习框架的版本对模型识别精度的影响可以忽略不计。

表 1 不同 TensorFlow 版本检测精度对比  
Table 1 Comparison of detection accuracy across different TensorFlow versions

网络名称	平均准确率 $AP$ /%			$mAP$ @0.5/%	运算 耗时/h
	坑洞/ 凹陷	裂纹/ 裂缝	疲劳 磨损		
Mask R-CNN_tf1.X	70.6	56.8	77.7	68.3	7.0
Mask R-CNN_tf2.X	70.4	57.3	78.1	68.6	6.7

表 2 为不同注意力机制添加顺序下的实验结果,由表中数据可知,CA+SA 相比于 SA+CA 所产生的模型对 3 种钢轨表面缺陷检测的  $AP$  值分别提高 1.2%、0.8%和 1.7%,模型的  $mAP$  提高 1.2%,因此选择 CA+SA 的组合顺序作为本文的改进方案。

表 2 不同注意力机制组合顺序下的实验结果  
Table 2 Experimental results with various attention mechanism orders

网络名称	平均准确率 $AP$ /%			$mAP$ @0.5/%
	坑洞/ 凹陷	裂纹/ 裂缝	疲劳 磨损	
Mask R-CNN	70.6	56.8	77.7	68.3
Mask R-CNN+SA+CA	73.6	60.7	80.2	71.5
Mask R-CNN+CA+SA	74.8	61.5	81.9	72.7

同时,对通道注意力机制和空间注意力机制不同组合顺序所得到的结果进行可视化。可视化结果如图 14 所示,CA+SA 对该数据集的训练效果更佳,相较 SA+CA 在目标边缘收敛性更好,进而不同程度提升了识别 3 种钢轨表面缺陷的能力。

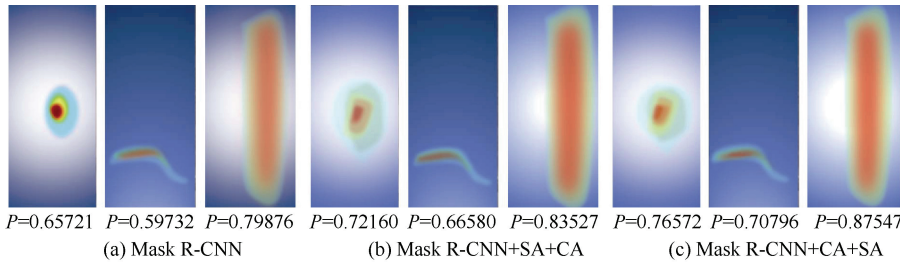


图 14 不同组合顺序下的检测结果可视化

Figure 14 Results visualization for various combination orders

表 3 是改进模型的实验结果，分析表 3 中的数据可得：分别加入通道、空间以及混合域的注意力机制后，模型对 3 类钢轨表面缺陷的实例分割精度均有一定的提升。其中，对钢轨表面缺陷类别“疲劳磨损”的识别精度分别提升了 1.3%、1.5%和 6.1%，对“凹陷”类别的识别精度分别提升了 0.7%、0.6%和 6.3%，对“裂纹”类别的识别精度分别提升了 1.4%、1.2%、6.9%。这表明，单独添加通道或者空间注意力机制对模型的提升效果有限，而加入混合域 CSM 注意力机制后，模型对各个类别的缺陷检测精度得到了显著提高。最终，改进算法对钢轨表面缺陷的平均识别精度达到 74.8%，相较于原始算法，虽然网络结构相对复杂，但  $mAP$  值提高了 6.5%。

表 3 消融实验结果

Table 3 Results of the ablation experiment

网络名称	CA	SA	CSM	平均准确率 AP/%			$mAP@0.5/\%$
				坑洞 凹陷	裂纹 裂缝	疲劳 磨损	
Mask R-CNN				70.6	56.8	77.7	68.3
Mask R-CNN_1	√			71.3	58.2	79.0	69.5
Mask R-CNN_2		√		71.2	58.0	79.2	69.4
Mask R-CNN_Ours			√	76.9	63.7	83.8	74.8

缺陷检测  $mAP$  值的提升，与回归框精度和实例分割精度正相关。在交并比  $IoU$  为 0.5 的前提下，所有识别目标生成回归框的平均精度曲线如图 15 所示，显然，改进后的算法的性能优于原始算法，改进后算法对所有预设识别目标生成回归框的平均精度为 71.2%，相比于原始算法高出 5 个百分点。这表明模型对小目标样本，如裂纹等的识别准确度获得显著提升。

同样在交并比  $IoU$  为 0.5 的前提下，所有预设识别目标进行实例分割的平均精度曲线如图 16 所示。显然，本文改进后的算法优于原始算法的性能，改进后

算法对所有预设识别目标实例分割的平均精度为 70.2%，高出原始算法 5.5 个百分点。这表明模型对边缘检测的能力，如裂纹的识别有着正向的影响。

对加入通道、空间以及混合域的注意力机制后模型检测效果分别进行可视化及精度标注，检测效果如图 17 所示(其中  $P$  为

检测精度)。

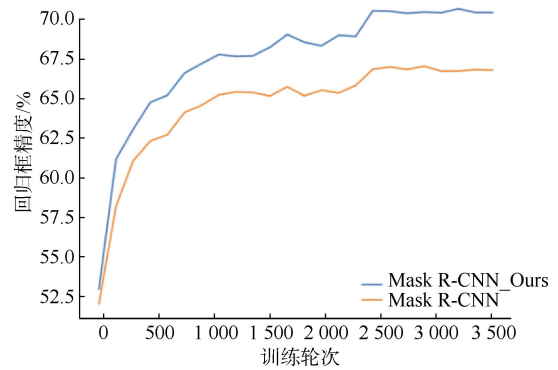


图 15 回归框精度曲线

Figure 15 Bounding box accuracy

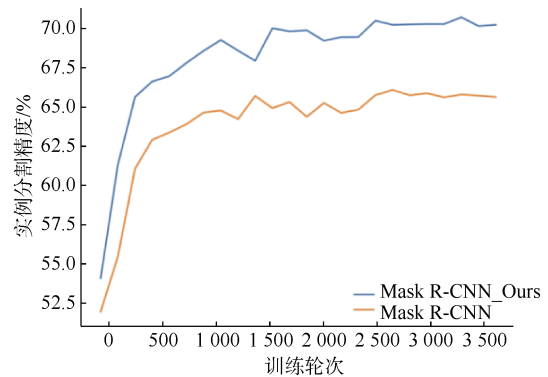


图 16 实例分割精度曲线

Figure 16 Instance segmentation accuracy

相同的实验环境下，应用不同 R-CNN 算法训练本文数据集，得到的实验数据如表 4 所示，改进后的网络模型性能优于原始算法以及该原始算法的两个前身 Fast R-CNN 和 Faster R-CNN。横向对比，从各个单类缺陷识别检测的平均精度  $AP$  值来看，改进后 Mask R-CNN\_Ours 模型，相对 Fast R-CNN 模型分别提升了 11.6%、12.5%、12.9%，相对 Faster R-CNN 模型提升效果依次是 8.8%、10.0%、10.3%，相对 Mask R-CNN 模型提升效果依次是 6.3%、6.9%、6.1%；纵

向对比, 从各个单类缺陷识别检测的整体平均精度  $mAP$  值来看, Mask R-CNN\_Ours 模型, 相对 Fast R-CNN 模型有 12.3% 的提升, 相对 Faster R-CNN 模型有 6.2% 的提升, 相对 Mask R-CNN 模型有 6.5% 的提升。实验结果表明, 改进后的模型在识别检测精度和小目标敏感度方面有显著优势。

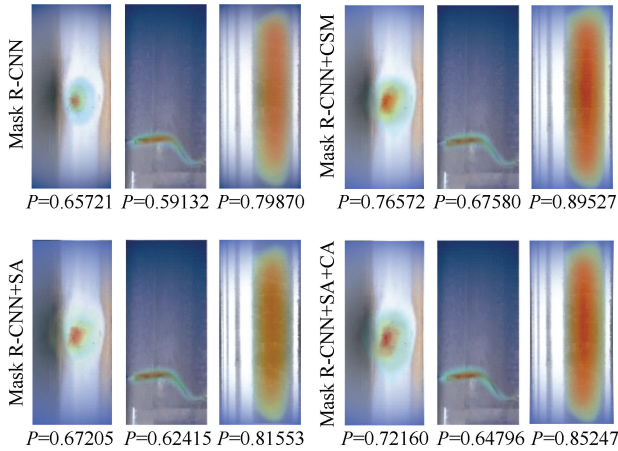


图 17 可视化检测效果及精度

Figure 17 Visualization of detection results and accuracy

表 4 改进的 Mask R-CNN 模型与其他不同算法的比较

Table 4 Comparison of improved Mask R-CNN model with other different algorithms

网络名称	平均准确率 AP/%			mAP@0.5/%
	坑洞 凹陷	裂纹 裂缝	疲劳 磨损	
Fast R-CNN	65.3	51.2	70.9	62.5
Faster R-CNN	68.1	53.7	73.5	68.6
Mask R-CNN	70.6	56.8	77.7	68.3
Mask R-CNN_Ours	76.9	63.7	83.8	74.8

对改进后的模型检测结果进行可视化, 如图 18 所示, 改进后的模型对凹陷、裂纹和磨损的识别效果得到了显著的提升。

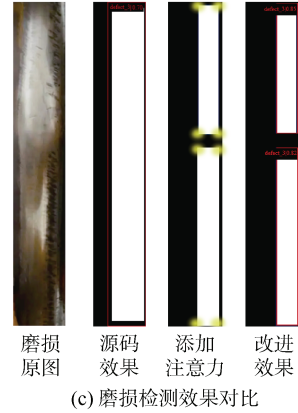
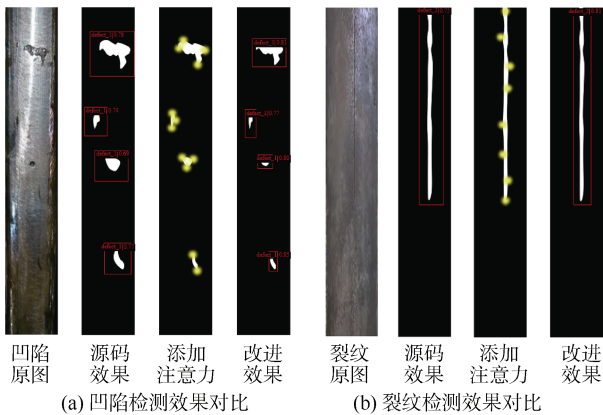


图 18 改进后模型实例分割效果

Figure 18 Improved model instance segmentation performance

图 18 中改进算法对 3 类钢轨表面缺陷图片的边缘敏感度有明显提升, 黄色标记为改进后模型对钢轨边缘的识别效果, 相比原始 Mask R-CNN 算法, 在边缘细节处理上更加细致准确。通过可视化识别结果表明, 在原始算法的残差网络之后添加通道-空间复合注意力机制, 进一步提高了模型的识别精度和泛化能力。

### 3 结论

本文针对钢轨缺陷检测实例分割算法 Mask R-CNN, 通过添加通道-空间复合注意力机制模块, 使得模型收敛速度更快, 回归框更加准确, 候选区域对识别对象边缘更加敏感, 最终生成的加权融合特征图精确地反映了像素点特征层与权重之间的关联。主要研究结论如下。

1) 验证了不同版本的 TensorFlow 深度学习框架对算法本身训练过程的影响较小, 排除了不同版本框架给实验结果带来影响的可能性, 同时, 设计实验比较不同组合顺序下注意力机制对模型精度的影响, 选定先通道后空间作为本文的改进方案。

2) 对比于单独添加通道和空间注意力机制, 加入混合域注意力机制后, 模型在钢轨缺陷数据集上的边缘检测能力获得显著提升, 并且预测掩膜形状更接近真实化。

3) 相较于 Fast R-CNN 和 Faster R-CNN 模型, 加入 CSM 后的 Mask R-CNN 模型对钢轨表面 3 种缺陷的识别效果得到了显著提升。

### 参考文献

[1] 中华人民共和国铁道部. 钢轨伤损分类: TB/T 1778—2010[S]. 北京: 中国铁道出版社, 2011.

- [2] 龙珍, 唐曼玲, 李静静. 基于机器视觉的钢轨缺陷检测方法综述[J]. 电子技术与软件工程, 2015, (12): 133.
- [3] 张辉, 宋雅男, 王耀南, 等. 钢轨缺陷无损检测与评估技术综述[J]. 仪器仪表学报, 2019, 40(2): 11-25.  
ZHANG Hui, SONG Yanan, WANG Yaonan, et al. Review of rail defect non-destructive testing and evaluation[J]. Chinese journal of scientific instrument, 2019, 40(2): 11-25.
- [4] 闵永智, 岳彪, 马宏锋, 等. 基于图像灰度梯度特征的钢轨表面缺陷检测[J]. 仪器仪表学报, 2018, 39(4): 220-229.  
MIN Yongzhi, YUE Biao, MA Hongfeng, et al. Rail surface defects detection based on gray scale gradient characteristics of image[J]. Chinese journal of scientific instrument, 2018, 39(4): 220-229.
- [5] 刘泽, 王崑, 王平. 钢轨表面缺陷检测机器视觉系统的设计[J]. 电子测量与仪器学报, 2010, 24(11): 1012-1017.  
LIU Ze, WANG Wei, WANG Ping. Design of machine vision system for inspection of rail surface defects[J]. Journal of electronic measurement and instrumentation, 2010, 24(11): 1012-1017.
- [6] 罗晖, 徐广隆. 基于图像增强与深度学习的钢轨表面缺陷检测[J]. 铁道科学与工程学报, 2021, 18(3): 623-629.  
LUO Hui, XU Guanglong. Rail surface defect detection based on image enhancement and deep learning[J]. Journal of railway science and engineering, 2021, 18(3): 623-629.
- [7] 苏焯, 李筠, 杨海马, 等. 基于 Faster R-CNN 的钢轨表面缺陷识别研究[J]. 电子科技, 2020, 33(9): 63-68.  
SU Ye, LI (J/Y), YANG Haima, et al. Research on rail surface defect recognition based on faster R-CNN[J]. Electronic science and technology, 2020, 33(9): 63-68.
- [8] 翟鑫, 李昕. 基于深度学习的铁路实例分割研究[J]. 佳木斯大学学报(自然科学版), 2021, 39(6): 116-118.  
ZHAI Xin, LI Xin. Research on railway instance segmentation based on deep learning[J]. Journal of Jiamusi University (natural science edition), 2021, 39(6): 116-118.
- [9] 马宇超, 付华良, 吴鹏, 等. 深度网络自适应优化的 Mask R-CNN 模型在铸件表面缺陷检测中的应用研究[J]. 现代制造工程, 2022(4): 112-118.  
MA Yuchao, FU Hualiang, WU Peng, et al. Research on the Mask R-CNN model of deep network adaptive optimization in the detection of casting surface defects[J]. Modern manufacturing engineering, 2022(4): 112-118.
- [10] 王海云, 王剑平, 张果, 等. 改进 FPN 的 Mask R-CNN 工业表面缺陷检测[J]. 制造业自动化, 2020, 42(12): 35-40.  
WANG Haiyun, WANG Jianping, ZHANG Guo, et al. Industrial surface defect detection based on improved FPN Mask R-CNN[J]. Manufacturing automation, 2020, 42(12): 35-40.
- [11] 钟瑞, 蒋斌, 李南星, 等. 嵌入注意力机制残差网络的人脸表情识别方法[J]. 计算机工程与应用, 2023, 59(11): 88-97.  
ZHONG Rui, JIANG Bin, LI Nanxing, et al. Facial expression recognition method embedded with attention mechanism residual network[J]. Computer engineering and applications, 2023, 59(11): 88-97.
- [12] AMBARTSOUMIAN A, POPOWICH F. Self-attention: a better building block for sentiment analysis neural network classifiers[EB/OL]. 2018: arXiv: 1812.07860. <http://arxiv.org/abs/1812.07860>.
- [13] HE Xiangnan, HE Zhankui, SONG Jingkuan, et al. NAIS: neural attentive item similarity model for recommendation[J]. IEEE transactions on knowledge and data engineering, 2018, 30(12): 2354-2366.
- [14] Wu X, Wen S, Xie Y. Improvement of Mask-RCNN object segmentation algorithm[C]//Intelligent Robotics and Applications: 12th International Conference, ICIRA 2019, Shenyang, China, August 8-11, 2019, Proceedings, Part I 12. Springer International Publishing, 2019: 582-591.
- [15] LI Haifeng, QIU Kaijian, CHEN Li, et al. SCAttNet: semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images[J]. IEEE geoscience and remote sensing letters, 2021, 18(5): 905-909.
- [16] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [17] GUO Menghao, XU Tianxing, LIU Jiangjiang, et al. Attention mechanisms in computer vision: a survey[J]. Computational visual media, 2022, 8(3): 331-368.
- [18] 刘嘉敏, 谢文杰, 黄鸿, 等. 基于空间和通道注意力机制的目标跟踪方法[J]. 电子与信息学报, 2021, 43(9): 2569-2576.  
LIU Jiamin, XIE Wenjie, HUANG Hong, et al. Spatial and channel attention mechanism method for object tracking[J]. Journal of electronics & information technology, 2021, 43(9): 2569-2576.
- [19] 郭倩. 嵌入多尺度模块和注意力机制的图像语义分割[D]. 烟台: 山东工商学院, 2021.  
GUO Qian. Image semantic segmentation embedded with multi-scale module and attention mechanism[D]. Yantai: Shandong Technology and Business University, 2021.
- [20] 陈维婧, 周萍, 杨海燕, 等. 通道-空间联合注意力机制的显著性检测模型[J]. 计算机工程与应用, 2021, 57(19): 214-219.  
CHEN Weijing, ZHOU Ping, YANG Haiyan, et al. Salient detection model based on channel-spatial joint attention mechanism[J]. Computer engineering and applications, 2021, 57(19): 214-219.

(编辑: 傅依萱)