

基于改进 Mask R-CNN 的钢铁表面缺陷检测与分割

薛殿龙¹, 李琳¹, 周子杰¹, 常永胜¹, 向勇², 陈德阳²

(1. 中国航空发动机集团有限公司 哈尔滨东安发动机有限公司, 哈尔滨 150000;

2. 重庆大学 机械与运载工程学院, 重庆 400044)

摘要: 针对金属表面缺陷存在的问题, 提出一种基于改进 Mask R-CNN 的表面缺陷检测与分割方法; 使用先进的 ConvNeXt-T 替换 ResNet-50 以改进用于特征提取的骨干网络, 在特征金字塔部分添加交错稀疏自注意力模块增强模型的全局建模能力, 同时通过多级区域特征融合以加强模型的上下文信息表达能力; 在钢铁表面缺陷数据集上开展了对比和校验验证, 结果显示骨干网络改进的效果最明显, 其 mAP_{bbox} 指标和 mAP_{mask} 指标分别提升了 8.2% 和 6.3%, 相较于对比方法, 所提方法对钢铁表面缺陷的检测和分割精度最高, mAP_{bbox} 指标和 mAP_{mask} 指标分别达到了 0.690 和 0.662。

关键词: ConvNeXt-T; 交错稀疏自注意力; 多级区域特征融合; 改进 Mask R-CNN

Metal Surface Defect Detection and Segmentation Based on Improved Mask R-CNN

XUE Dianlong¹, LI Lin¹, ZHOU Zijie¹, CHANG Yongsheng¹, XIANG Yong², CHEN Deyang²

(1. Harbin Dongan Engine Co., Ltd., AECC, Harbin 150000, China;

2. College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China)

Abstract: To address metal surface defects, a surface defect detection and segmentation method based on an improved Mask R-CNN is proposed. The method replaces ResNet-50 with advanced ConvNeXt-T to improve the backbone network for feature extraction, add interleaved sparse self-attention modules to enhance the global modeling ability of the model in the feature pyramid section, and strengthen the contextual information representation ability of the model through multi-level regional feature fusion. Comparisons and validations are conducted on a dataset of steel surface defects. The results show that the backbone network has the most significant improvement effect, with increases of 8.2% and 6.3% in the mAP_{bbox} and mAP_{mask} indicators, respectively. Compared with other methods, the proposed method has the highest detection and segmentation accuracy for steel surface defects, with the mAP_{bbox} and mAP_{mask} indices reaching up to 0.690 and 0.662, respectively.

Keywords: ConvNeXt-T; interleaved sparse self-attention; multi-level regional feature fusion; improved Mask R-CNN

0 引言

表面缺陷是钢铁产品制造过程中一种常见的问题, 例如划痕、麻点、孔洞等, 这些表面缺陷可能会对钢铁产品的性能和质量产生不利影响, 因此, 对钢铁表面缺陷进行有效检测和控制非常重要^[1-2]。由于表面缺陷的多样性与复杂性, 目前在表面缺陷自动检测领域在技术上还面临诸多的挑战。传统图像处理法利用阈值法^[3]、结构法^[4]、滤波器法^[5]、基于模型^[6]等方法将图像局部

异常反映的原始特性来分割和检测缺陷, 但缺陷检测的性能受光照、摄像机角度和背景影响很大。基于传统机器学习的钢铁表面缺陷识别方法主要分为基于纹理特征的方法和基于形状特征的方法, 显著性线性扫描获得油污区域, 识别硅钢中的各种缺陷^[7]。近些年研究者习惯代替传统的机器学习方法来检测钢铁表面缺陷, 一种改进的 Faster R-CNN 网络^[8]被用于识别钢材表面的微小目标缺陷, 一种带背景约束的全局低秩非负重建算法^[9]实现钢轨表面缺陷的检测, 一种采用 OCR-UNet 网

收稿日期: 2024-06-11; 修回日期: 2024-08-23。

作者简介: 薛殿龙(1996-), 男, 硕士研究生, 工程师。

引用格式: 薛殿龙, 李琳, 周子杰, 等. 基于改进 Mask R-CNN 的钢铁表面缺陷检测与分割[J]. 计算机测量与控制, 2025, 33(6): 47-53, 75.

络^[10]实现金属表面缺陷分割。

语义分割应用于钢铁表面缺陷识别领域,可以将钢铁表面图像分割成不同的区域,并将每个像素分配到不同的类别中(背景和各类缺陷)。钢铁表面缺陷检测的链式空洞空间金字塔池化网络(CASPPNet)^[11],通过引入改进的全局注意力模块和残差边界细化模块有效提升了缺陷分割的精度。深度学习语义分割方法无法区分同一缺陷类别中不同的实例,实例分割不仅能将图像中的像素分类,还可以将缺陷实例与其他实例分开,实例分割可以更加准确地检测出钢铁表面的每个缺陷实例并分类。使用 Mask R-CNN 和迁移学习的方法^[12],成功测量了高碳钢中的金属粉末颗粒的含量。Mask R-CNN 等图像分割方法^[13]识别钢铁缺陷,但该方法受大量无缺陷样本的干扰比较严重。本文提出一种基于改进的 Mask R-CNN 的两阶段算法模型,实现对钢铁表面缺陷的实例级检测和像素级分割,不仅能对缺陷进行分类识别,还能够确定缺陷的具体形状和位置,对于钢铁表面缺陷的精细化检测和分析具有很大的价值和应用前景。

1 Mask R-CNN 算法

1.1 Mask R-CNN 算法流程

Mask R-CNN^[14]是一种实例分割方法,它在两阶段目标检测网络 Faster R-CNN^[15]的基础上添加了一个掩码分割分支,弥补了目标检测网络只能定位和分类目标而不能实现像素级分割目标的不足,Mask R-CNN 同时实现目标检测与分割的算法流程如图 1 所示,该方法具有多任务学习的能力,可以实现将钢铁表面缺陷图像中的缺陷同时检测、分类和分割。

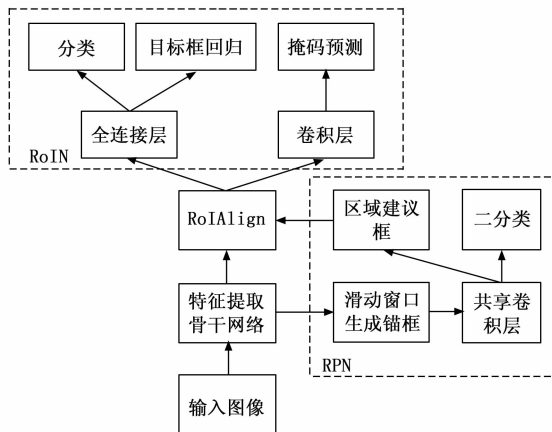


图 1 Mask R-CNN 算法流程图

整个流程分为 3 个部分,首先使用特征提取骨干网络从输入图像生成初始特征;然后经过 RPN (Region Proposal Network) 生成区域建议框,具体来说:第一步通过滑动窗口生成密集的预设锚框,第二步使用共享卷积层提取每个锚框的特征,第三步对每个锚框进行前

后景的二分类和通过回归分支生成区域建议框;最后使用感兴趣区域对齐 (RoIAlign) 将 RPN 阶段生成的区域建议框重映射回骨干网络的特征层中提取相对应的区域特征,并经过 RoIN (Region of Interest Network) 对区域建议框进行目标前景分类、精细地检测框回归和分割掩码预测。

1.2 FPN 多尺度预测模型

Mask R-CNN 算法中,由于最高层特征相较原始输入图像尺寸缩小了几十倍,导致小目标难检测甚至无法检测的问题,为此需要引入了特征金字塔网络 (Feature Pyramid Network)。经典检测模型,当遇到多尺度目标检测尤其是小目标检测问题时,需要使用 FPN 多尺度预测模型^[16],其模型结构如 2 所示。使用带 FPN 多尺度预测模型结构的 Mask R-CNN 能够实现同时对钢铁表面缺陷不同尺寸目标的预测,其性能较单尺度预测的模型更好。

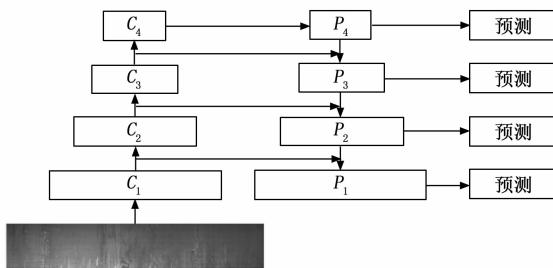


图 2 FPN 多尺度预测模型

2 基于改进 Mask R-CNN 的缺陷检测与分割方法

2.1 改进的 Mask R-CNN 结构

Mask R-CNN 算法是一个典型的两阶段实例分割算法,卷积神经网络在不断的上采样和层数的堆叠过程中,高层特征能够获得较大的感受野,但相较于输入图像的尺寸而言,仍缺少长距离依赖,掩码预测需要对区域建议框内的像素做密集的逐像素分类,而缺陷在一张所采集的图像中的分布是杂乱无章的,因此在基本模型 Mask R-CNN 的基础上增加计算长距离依赖的注意力模块来增强特征的长距离信息交互能力。原始的区域建议框特征重映射是只使用了单级特征,而高级语义信息和低级细节信息对密集的目标检测和分割任务都很重要,因此,采用多级特征映射并融合的方法使得 RoIAlign 所提取的特征更综合,适应多种多样的缺陷识别。所提出的改进 Mask R-CNN 结构如图 3 所示。

假设输入的缺陷图像大小为 $H \times W$, H 、 W 分别为图像的高和宽,首先使用骨干网络提取基本特征,获得 4 个尺度的特征图 C_1 、 C_2 、 C_3 、 C_4 ,特征图大小分别为 $H/4 \times W/4$ 、 $H/8 \times W/8$ 、 $H/16 \times W/16$ 、 $H/32 \times W/32$,原始的 Mask R-CNN 使用的特征提取网络为

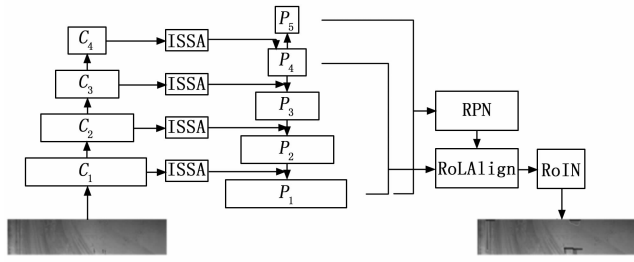


图 3 改进的 Mask R-CNN 结构

ResNet-50, 而 ConvNeXt 网络具有比 ResNet 网络更好的特征提取能力, 因此, 使用 ConvNeXt-T 作为提出的改进 Mask R-CNN 的骨干网络进行特征提取; 紧接着使用交错稀疏自注意力 (ISSA, interlaced sparse self-attention) 给每个尺度的特征增加对长距离依赖的表示能力, 提升模型密集检测与分割的性能; 之后使用 FPN 结构融合不同尺度的特征 P_1 、 P_2 、 P_3 、 P_4 , 将高级语义特征自顶向下融合, 融合过程如图 4 所示, 对同级特征采用 1×1 卷积调整通道数至一致 (此处设为 256), 对上级特征采用 2 倍上采样将特征图的分辨率提高到原来的 2 倍, 上采样方法为最近邻插值法, 然后使用特征逐点相加的方式融合特征, 融合之后便能使高分辨率的低级特征对缺陷的高级语义信息有更准确的表示, 为了满足更大尺度的缺陷目标的检测和分割需求, 对融合后的特征 P_4 使用步长为 2, 核大小为 2×2 的最大池化以生成更小分辨率与更大感受野的特征图 P_5 , 其特征图大小为 $H/64 \times W/64$; 然后使用 RPN 网络做第一阶段的缺陷检测并生成一系列的区域建议框, RoIAlign 使用这些区域建议框重映射回 FPN 结构的特征图的具体层级和位置并生成特定大小 (检测分支设为 7×7 , 分割分支设为 14×14) 的区域特征图, 由于特征图 P_5 的分辨率较小, 容易造成密集的缺陷检测和分割任务的不准确, 因此, 重映射区域建议框的特征时不使用该层特征图; 最后使用 RoIN 对 RoIAlign 输出的区域特征图进行缺陷分类、缺陷目标框回归和缺陷前景分割。

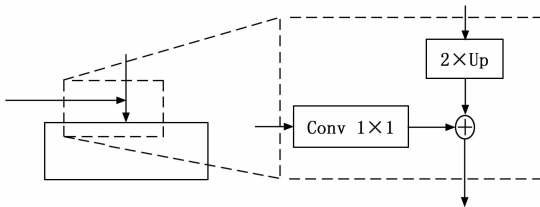


图 4 FPN 特征融合过程

2.2 交错稀疏自注意力

给模型增加长距离依赖能有效增强所提取特征的全局表达能力, 但其计算量太大, 导致模型效率大幅度降

低^[17], 不适用于在工业领域有实时性要求的钢铁表面缺陷检测场景。为此引入交错稀疏自注意力模块, 使用稀疏矩阵计算自注意力来替代原始自注意力的密集矩阵计算。

图 5 为交错稀疏自注意力的计算过程, 其计算过程主要包括长距离注意力计算和短距离注意力计算, 自注意力的计算公式如式 (1) 所示:

$$z = \text{Softmax} \left[\frac{\theta(x) \varphi(x)^T}{\sqrt{d}} \right] \psi(x) \quad (1)$$

其中: $x, z \in R^{N \times C}$ (N 表示特征所含的像素点数, C 表示特征的通道数) 分别为自注意力的输入和输出特征块, θ 函数和 φ 函数分别为降维函数, 主要目的是给输入特征降维以减小计算量, 减小计算量后 $\theta(x)$, $\varphi(z) \in R^{N \times (C/2)}$, $d=C/2$ 是规模因子, 主要是为了解决 Softmax 函数的小梯度问题, $\psi(x) \in R^{N \times C}$ 是为了给自注意力函数学习一个更好的嵌入值。

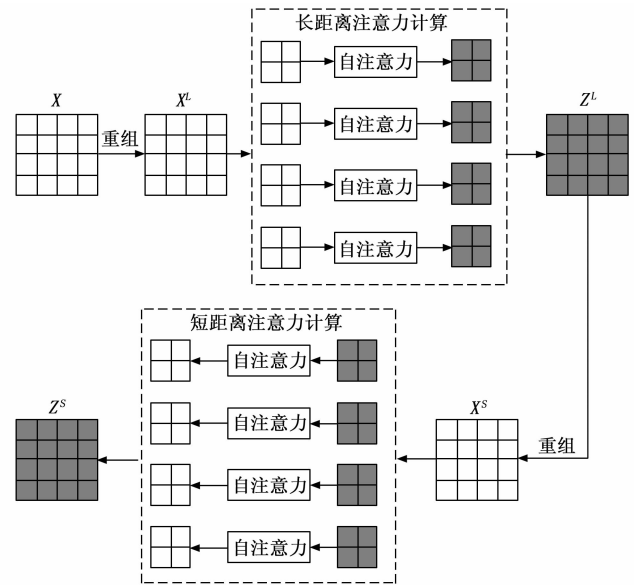


图 5 交错稀疏自注意力

假设输入特征图为 X , 首先将特征图进行重组, 将较远位置的像素块组合到一个子区域特征块, 以便进行后续的分块自注意力计算, 得到重组后的特征 X^L 。然后, 将重组后的特征 X^L 分成 P 个不同的区域块, 每个区域块包含 Q 个像素点 ($N=P \times Q$), 之后, 对每个小区块都使用公式 (2) 来计算对应的自注意力权重:

$$Z_i^L = \text{Softmax} \left[\frac{\theta(X_i^L) \varphi(X_i^L)}{\sqrt{d}} \right] \psi(X_i^L), i = 1, 2, \dots, P \quad (2)$$

其中: $X_i^L, Z_i^L \in R^{Q \times C}$ 分别为长距离注意力计算中的输入和输出的小区块, 将分块计算自注意力后的含有注意力权重的小区块重新组合就得到长距离注意力

计算的输出特征图 Z^L 。

计算完长距离的自注意力之后, 将特征再次重组回与输入特征像素一一对应的位置, 得到特征 X^S 以便进行短距离注意力的计算, 然后使用和长距离注意力计算一样的分块方式将特征图分成 P 个小区块并分别计算各自的自注意力权重, 如式 (3) 所示:

$$Z_i^S = \text{Softmax}\left(\frac{\theta(X_i^S)\varphi(X_i^S)}{\sqrt{d}}\right)\psi(X_i^S), i = 1, 2, \dots, P \quad (3)$$

其中: $X_i^S, Z_i^S \in R^{Q \times C}$ 分别为短距离注意力计算中的输入和输出的小区块。最后将各个小区块重新对应组合, 得到最终的特征图 Z^S , 此时, 特征图既具有长距离的注意力权重依赖, 又含有短距离的自注意力, 且整个注意力计算过程仅使用了两次稀疏矩阵计算, 相较于非局部自注意力机制的全特征图矩阵的密集计算, 极大地减少了计算量。

2.3 多级区域特征融合

传统的 Mask R-CNN 在使用 RoIAlign 提取 RPN 生成的区域建议框所对应的区域特征时, 需要判断进行采样的最佳特征层级, 如公式 (4)^[18] 所示:

$$k = 4 - \log_2 \lfloor (\sqrt{wh} / \sqrt{w_s h_s}) \rfloor \quad (4)$$

其中: k 为计算的特征层级, w, h 分别为区域建议框的宽和高, $\sqrt{w_s h_s}$ 为最小分辨率特征图的尺寸。

虽然可以计算出最佳的特征层级, 但没有综合应用从低级的细节特征到高级的语义特征, 对细节信息和语义信息多变的缺陷目标鲁棒性不高。为此, 采用多级区域特征融合的方式, 进行更精确的缺陷分类、检测与分割, 此过程如公式 (5) 所示:

$$F = \sum_{i=1}^n A[f_i(r)] \quad (5)$$

其中: $f_i(r)$ 表示对区域建议框在特征层级 i 重映射后的特征区域, A 表示使用 RoIAlign 将特征调整为特定大小的新特征, n 表示使用的特征层级总数 (使用 P_1 到 P_4 特征图, 故 n 取 4), F 为融合多级特征信息并固定输出大小之后的综合特征。

2.4 网络损失函数

Mask R-CNN 损失函数包括 RPN 阶段的分类与回归损失和 RoIN 阶段的分类、回归与掩码预测损失, 复合损失函数的计算如式 (6) 所示:

$$L = L_{\text{RPN}} + L_{\text{RoIN}} \quad (6)$$

其中: $L, L_{\text{RPN}}, L_{\text{RoIN}}$ 分别为总复合损失、RPN 的总损失和 RoIN 的总损失。RPN 网络的总损失由二分类损失和回归损失组成, 如式 (7) 所示:

$$L_{\text{RPN}} = L_{\text{RPN}}^{\text{cls}} + L_{\text{RPN}}^{\text{reg}} \quad (7)$$

RoIN 的总损失由缺陷类别的分类损失、缺陷边界框的回归损失和缺陷的分割掩码损失组成, 如式 (8)

所示:

$$L_{\text{RoIN}} = L_{\text{RoIN}}^{\text{cls}} + L_{\text{RoIN}}^{\text{reg}} + L_{\text{RoIN}}^{\text{mask}} \quad (8)$$

3 实验验证结果与分析

3.1 实验数据集分析

数据集为 Kaggle “Severstal: Steel Defect Detection” 比赛中的数据集^[17], 该数据集一共有 6 666 张含缺陷的图像 (其中仅含麻点缺陷的图像有 769 张, 仅含开裂缺陷的图像有 195 张, 仅含划痕缺陷的图像有 4 759 张, 仅含斑块缺陷的图像有 516 张, 含两种以上缺陷的图像有 427 张), 图像分辨为 $1\,600 \times 256$ 。对数据集各缺陷分别进行缺陷大小和缺陷高宽比的统计, 统计结果如 (6) 所示。图 6 (a) 中各种缺陷的尺寸中大部分是小面积缺陷, 采用 FPN 多尺度特征提取与预测的方式来检测与分割缺陷更适应钢铁表面缺陷检测的场景。图 6 (b) 中大部分缺陷高宽比主要分布在 $0 \sim 2.5$ 范围内, 详细分布概率情况如图 6 (c) 所示, 为了提高检测与分割的精度, 将 Mask R-CNN 算法中的预设锚框的高宽比值从 $[0.5, 1.0, 2.0]$ 更改为 $[0.25, 0.5, 1.0, 2.0, 5.0]$, 使密集的预设锚框更符合此处的缺陷尺寸需求。

3.2 缺陷检测与分割评价指标

检测和分割中一个比较重要的基本指标是交并比 (IoU, intersection over union), 交并比的计算如式 (9) 所示, 表示预测区域与实际区域的交集与并集的比值, 当预测区域与实际区域完全重合时, 交集和并集也重合, 此时交并比为 1:

$$IoU = \frac{A \cap B}{A \cup B} \quad (9)$$

对于缺陷检测与分割任务, 通常使用 0.5 作为判断的阈值, 即当对某缺陷的预测区域与人工标定的实际区域的交并比大于 0.5 时, 则认为该缺陷被正确检测出, 否则就属于检测错误的实例。对于某一类缺陷而言, 模型的预测存在 4 种情况: 1) 缺陷被正确的识别出来, 即 TP (True Positive); 2) 非该类别缺陷被正确的识别出, 即 TN (True Negative); 3) 该类别缺陷被错误的识别为其他类别的缺陷或背景, 即 FN (False Negative); 4) 非该类别缺陷被错误地识别为该类别缺陷, 即 FP (False Positive)。目标检测和实例分割任务中的两个基本指标是精确率 (Precision) 和召回率 (Recall), 其计算公式分别如式 (10) 和 (11) 所示, 精确率表示预测为缺陷的目标中被正确预测的比例, 召回率表示真实缺陷中实际被正确预测出的比例:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

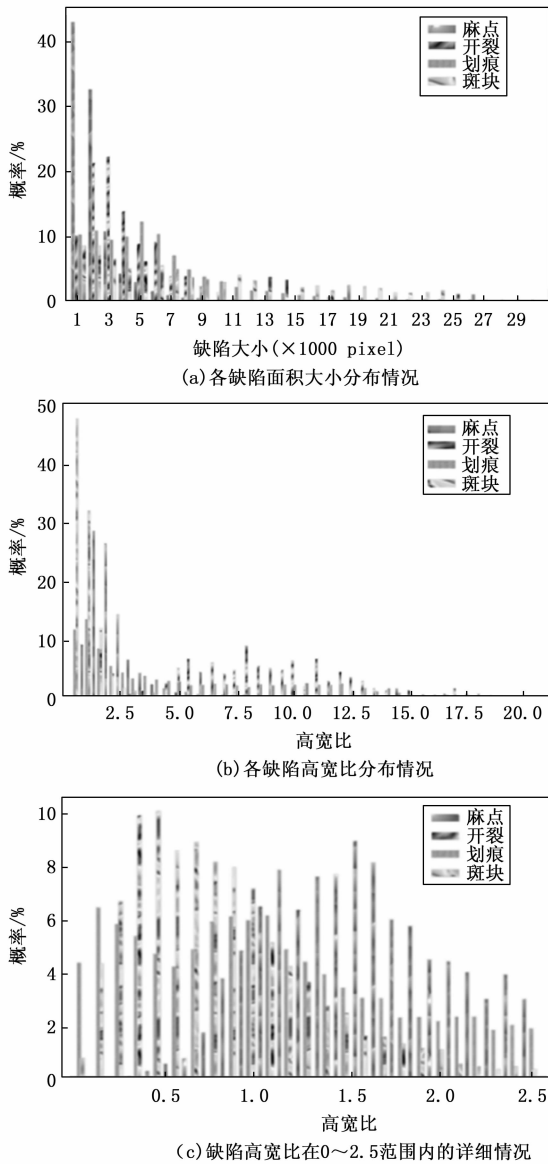


图 6 数据集各缺陷统计

PR (Precision-Recall) 曲线与坐标轴围成的面积表示模型预测的平均精确率 (AP, average precision), AP 值的计算使用 101 点插值法计算, 即以 0.01 为间隔取 0~1 之间所对应的召回率值所对应的精确率值的插值的平均值, 计算过程如式 (12) 所示:

$$AP = \frac{1}{101} \sum_{r \in \{0.0, 0.01, \dots, 1\}} \rho_{\text{interp}}(r) \quad (12)$$

其中: $\rho_{\text{interp}}(r)$ 表示在召回率 r 处精确率值的插值。

模型使用目标检测和实例分割领域的经典评估指标评价模型的性能, 即平均精确率均值 (mAP, mean average precision), 表示所有缺陷类别的平均精确率的平均值, 计算如公式 (13) 所示:

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (13)$$

其中: C 代表缺陷的总类别数, 由于算法同时实现缺陷检测和分割, 因此, 分别使用 mAP_{bbox} 和 mAP_{mask} 表示检测和分割的平均精确率均值。

除了使模型的预测精确度尽量高之外, 钢铁表面缺陷的检测与分割还要求算法满足工业领域缺陷检测的实时性要求。使用每秒帧率 (FPS, frame per second) 作为缺陷检测与分割的实时性指标。

3.2 模型训练及超参数设置

使用在 ImageNet 上预训练的 ConvNeXt-T 模型的参数作为骨干网络的初始参数, 以减少数据量不足带来的影响。将数据集图像按类别以 80% 与 20% 的比例分为训练集和测试集。训练时设置批处理大小 batch_size 为 8, 总共训练 50 次 epoch, 迭代批次共 33 000 多个。使用随机梯度下降 (SGD, stochastic gradient descent) 算法作为优化函数, 并选择动量, 动量参数 momentum 设置为 0.9, 同时设置 weight decay 参数为 0.000 1, 预热批次设置为 500 次, 初始学习率设为 0.01, 分别在训练 30 和 40 次 epoch 之后将学习率降低为原来的 1/10, 设置梯度的欧几里得范数 (又称为 L2 范数) 的阈值为 35, 当模型训练时计算的梯度的 L2 范数超过阈值时, 将相应的梯度缩放到预设的阈值之下, 以此有效地限制梯度的大小。

图 7 为本文方法训练过程的损失变化图, Mask R-CNN + ConvNeXt-T 表示改用 ConvNeXt-T 作为 Mask R-CNN 的骨干网络。由图 7 可以看出, 所提方法在网络训练了 30 个 epoch 之后, 损失值有一个明显的下降, 然后趋于收敛的平稳状态, 证明了改进方法的有效性。

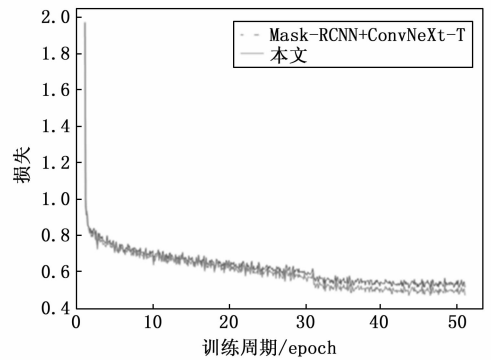


图 7 网络训练损失

3.3 实验结果与分析

为了验证所提方法的有效性和先进性, 除训练所提模型之外, 还使用 Mask R-CNN、YOLACT 和 SOLOv2 等主流实例分割模型进行对比, 此外, 为验证所提改进点的有效性, 进行了不同改进点的消融实验, 设置的对比方法和消融实验的代号设置如表 1 所示, 其中, C1、C2、C3 表示 3 种对比方法, 而 A1、A2、A3、A4、A5、A6、A7 表示 7 组消融实验。

表 1 对比方法及消融实验代号

方法		代号
对比方法	YOLACT(文献[19])	C1
	SOLOv2(文献[20])	C2
	Mask R-CNN(文献[21])	C3
消融实验	Mask R-CNN + ISSA	A1
	Mask R-CNN + RFF	A2
	Mask R-CNN + ConvNeXt-T	A3
	Mask R-CNN + ISSA + RFF	A4
	Mask R-CNN + ConvNeXt-T + ISSA	A5
	Mask R-CNN + ConvNeXt-T + RFF	A6
	Mask R-CNN + ConvNeXt-T + ISSA + RFF	A7

各对比算法和各组消融实验的性能对比如表 2 所示。相较于典型的两阶段实例分割算法 Mask R-CNN 而言, YOLACT 和 SOLOv2 是单阶段的实例分割算法, 其算法流程更简单快速, 但精度会低于两阶段的算法, YOLACT 模型的训练对训练样本的要求较高, 但钢铁表面缺陷样本较少, 由表 2 可以看出, YOLACT 方法的缺陷检测和分割精度都较低, SOLOv2 是一种直接通过位置来分割物体的方法, 但其对目标形状的鲁棒性不强, 所用数据集存在的缺陷的高宽比的变化范围很广, 尤其是开裂和划痕类型的缺陷多以长条形存在, 更容易造成 SOLOv2 方法的错分割情况。Mask R-CNN 作为两阶段实例分割算法, 在 RoIN 阶段能对 RPN 阶段生成的感兴趣区域进行更精细化的回归, 因此, 其精度相较于单阶段实例分割算法通常更高, 对的钢铁表面缺陷的检测精度和分割精度分别达到了 0.571 和 0.570。

表 2 不同模型算法的性能

代号	mAP_{bbox}	mAP_{mask}	参数量/M	计算量/GFLOPs	FPS
C1	0.454	0.431	34.81	193.78	34.5
C2	—	0.557	46.24	139.62	34.8
C3	0.571	0.570	43.76	129.67	40.4
A1	0.598	0.599	82.19	192.95	28.5
A2	0.608	0.584	43.76	129.76	38.9
A3	0.653	0.633	47.69	130.47	40.1
A4	0.627	0.614	82.19	193.03	26.9
A5	0.673	0.657	54.93	150.18	30.3
A6	0.671	0.642	47.69	130.55	38.1
A7	0.690	0.662	54.93	150.27	28.9

对比实验 C3 和 A3, 使用 ConvNeXt-T 替换 Res-Net-50 作为 Mask R-CNN 的骨干网络进行特征提取, 由于其强大的特征提取能力, 使模型的缺陷检测和分割性能大幅度提升, 和分别达到了 0.653 和 0.633, 相较于原始的 Mask R-CNN 模型, 检测性能提升了 8.2 个百分点, 且分割性能提升了 6.3 个百分点, 同时模型实时性指标 FPS 两者相当, 说明检测和分割能力提升的

同时还能保持算法实时性能几乎没有降低, 证明了 ConvNeXt-T 作为骨干网络的先进性。对比实验 C3 和 A1, 通过在基准网络 Mask R-CNN 的基础上添加 ISSA 改进点, 通过计算长距离和短距离注意力并融合, 提高缺陷检测与分割的精度, 最终实现检测精度达到 0.598 和分割精度达到 0.599。对比实验 C3 和 A2, 通过在基准网络上添加 RFF 改进点, 多级区域特征的融合将模型的检测精度提高 3.7 个百分点, 分割精度也提高了一个多百分点, 同时, 由 C3 与 A2 的对比、A1 与 A4 的对比、A3 与 A6 的对比、A5 与 A7 的对比看出, 添加 RFF 改进策略基本不增加模型的参数量, 少量计算量的增加 FPS 只有轻微下降, 即使用较小的模型运行速度的代价获得较大的缺陷识别精确度的提升。从实验 C3、A1、A2、A3 四组实验可以看到, 仅给基准网络添加单个改进点的策略中, 骨干网络的改进是最有效的, 能够极大地提高缺陷识别精度的同时模型实时性能也基本没有降低。A4、A5、A6、A7 的对比实验中, 使用组合改进点对基准网络进行优化, 能更有效地提升基准网络的缺陷识别性能, 同时使用 3 种改进策略之后的模型具有强大的特征提取能力, 且既能使特征具有含长距离依赖的全局表达能力, 又使特征具有不同层级的上下文信息表达能力, 3 者优势的融合使得提出的最终模型实现了 0.690 的缺陷检测精度和 0.662 的缺陷分割精度, 虽然算法运行速度有所降低, 但 FPS 仍能达到 28.9, 满足算法实时性要求。

在目标检测领域, 依据图像中物体的像素大小将目标分为小目标、中级目标和大目标。为了验证所提方法对各种尺寸的钢铁表面缺陷识别的泛化能力, 将数据集 中的缺陷分为小缺陷(缺陷尺寸小于 1 024 像素)、中 级缺陷(缺陷尺寸在 1 024 像素与 9 216 像素之间)和大 缺陷(缺陷尺寸大于 9 216 像素), 测试集中 3 种等级 的缺陷分别有 635 个、2 255 个和 994 个, 可以看出, 所用钢铁表面缺陷数据集中主要为中级缺陷, 其占比超 过了一半, 而小缺陷的占比也有近 16%。图 8 为不同 对比方法对 3 种等级缺陷识别的召回率, 由于 SOLOv2 方法没有检测过程, 故仅对比其分割的召回率。由图 8 可以看出所提方法对小缺陷、中级缺陷和大缺陷的识别 召回率几乎高于其他的所有对比方法, 尤其是小缺陷和 中级缺陷的识别召回率提升最为明显, 对小缺陷的检测 和分割召回率均超过了 45%, 对中级缺陷的检测和分 割召回率都超过了 80%, 虽然其对大缺陷的检测召回 率稍低于 Mask R-CNN 方法, 但依然达到 90%。总之, 所提方法对尺寸等级不同的各缺陷均有较好的识别效 果, 证明该算法模型具有良好的泛化能力。

图 9 为不同算法模型对钢铁表面缺陷的检测和分割 结果示例。从麻点、开裂、划痕和斑块缺陷的检测与分

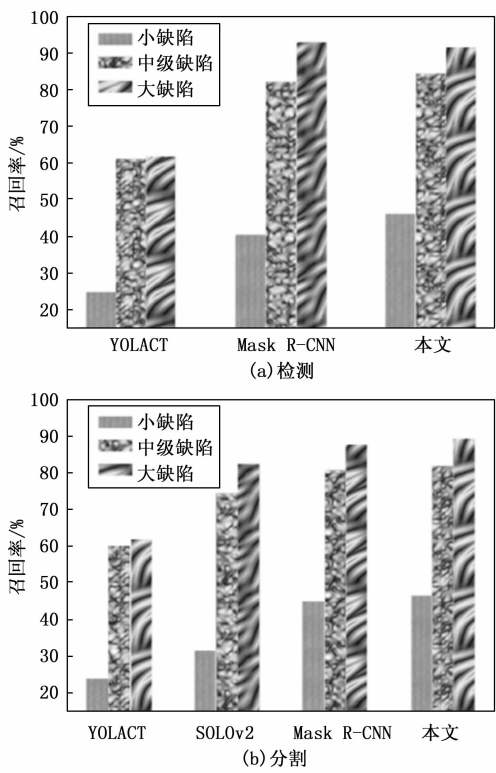


图 8 各方法对不同级别缺陷识别的召回率

割结果可以看出, YOLACT 方法、SOLOv2 方法及 Mask R-CNN 方法都有明显的不足, 而提出的方法与缺陷标签较为接近, 更符合真实缺陷的存在情况。纵观 4 种缺陷, 提出的基于改进 Mask R-CNN 的缺陷检测与分割方法具有有效性和先进性。

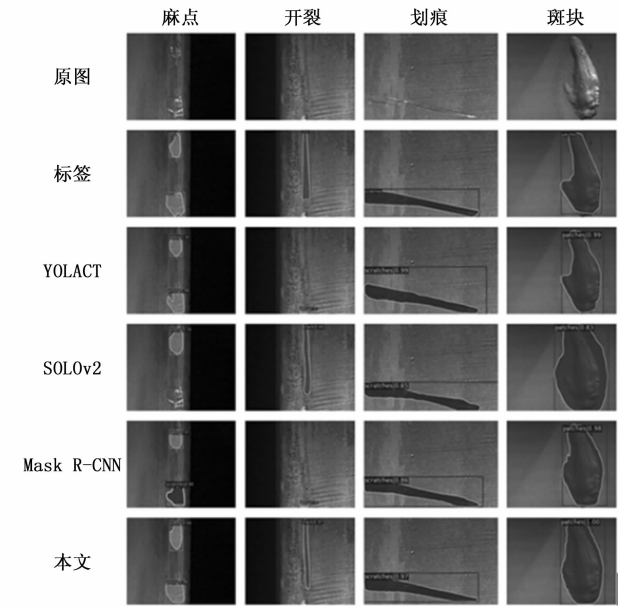


图 9 不同算法模型的缺陷检测与分割结果示例

4 结束语

提出一种基于改进 Mask R-CNN 的表面缺陷检测

与分割方法, 本方法以两阶段实例分割网络 Mask R-CNN 作为基本模型, 通过使用先进的 ConvNeXt-T 替换 ResNet-50 以改进用于特征提取的骨干网络, 使网络提取的特征能更准确的描述钢铁表面缺陷; 在特征金字塔部分添加交错稀疏自注意力模块以增加模型的长距离依赖, 增强模型的全局建模能力和表示能力; 同时通过多级区域特征融合以加强模型的上下文信息表达能力, 实现先进和有效的钢铁表面缺陷检测与分割。在钢铁表面缺陷数据集上的实验验证了所提方法的先进性及其对缺陷识别的高精度, 通过各改进点不同组合方式的消融实验验证了所提各改进点的有效性, 其中由骨干网络改进的效果提升最明显, 其指标和指标分别提升了 8.2% 和 6.3%, 相较于几种对比方法, 所提方法对钢铁表面缺陷的检测和分割精度最高, 指标和指标分别达到了 0.690 和 0.662。

参考文献:

[1] CAO J G, YANG G T, YANG X Y. A pixel-Level segmentation convolutional neural network based on deep feature fusion for surface defect detection [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1 - 12.

[2] YANG J, LI S B, WANG Z, et al. Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges [J]. Materials, 2020, 13 (24): 5755.

[3] CAO J J, ZHANG J, WEN Z J, et al. Fabric defect inspection using prior knowledge guided least squares regression [J]. Multimedia Tools and Applications, 2015, 76 (3): 4141 - 4157.

[4] MASOUMEH A, THOMAS K. Automatic thresholding for defect detection by background histogram mode extents [J]. Journal of Manufacturing Systems, 2015, 37: 83 - 92.

[5] ZHOU X N, WANG Y N, XIAO C Y, et al. Automated visual inspection of glass bottle bottom with saliency detection and template matching [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68 (11): 4253 - 4267.

[6] XI J Q, SHENTU L F, HU J K, et al. Automated surface inspection for steel products using computer vision approach [J]. Applied Optics, 2017, 56 (2): 184 - 192.

[7] SONG K, HU S, YAN Y, et al. Surface defect detection method using saliency linear scanning morphology for silicon steel strip under oil pollution interference [J]. ISIJ International, 2014, 54 (11): 2598 - 2607.

(下转第 75 页)