

# 基于改进级联 R-CNN 的钢材带状碳化物检测与分级

郝亮, 周诗洋, 莫允扬, 陈勇勇, 徐勇, 苏敬勇

(哈尔滨工业大学(深圳)计算机科学与技术学院, 深圳 518055)

**摘要:** 在钢铁行业中, 碳化物是钢材中一种非常重要的组成成分, 其在钢材中的分布对于评估钢材的质量具有很高的参考价值。然而, 目前棒材碳化物的检测手段主要为人工检测, 成本高昂且缺乏稳定性。引入人工智能领域的深度学习技术, 收集并标注了 3 192 张高质量钢铁棒材带状碳化物图像与 11 个完整样品数据, 创建了工业场景下的棒材带状碳化物目标检测数据集 (Banded carbide dataset on object detection for steel bar, BCDOD)。使用深度学习领域中常见的目标检测方法对数据集进行了实验分析, 针对应用场景与数据的特点, 引入旋转数据增强、Focal Loss 函数与负样本微调对级联 R-CNN 模型进行改进, 提升了模型的性能, 平均精度达到 96%。同时, 在完整样品数据取得了 100% 的识别准确率, 取得了较为理想的效果, 弥补了人工智能技术在碳化物金相检测领域的空缺。

**关键词:** 碳化物; 金相组织; 缺陷检测; 目标检测; 级联 R-CNN

中图分类号: TP391 文献标志码: A

## Detection and Classification of Banded Carbide in Steel Based on Improved Cascade R-CNN

HAO Liang, ZHOU Shiyang, MO Yunyang, CHEN Yongyong, XU Yong, SU Jingyong

(School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China)

**Abstract:** In the steel industry, carbide is a vital constituent, whose distribution in steel materials holds significant reference value for evaluating steel quality. However, the current detection methods for carbide in steel bars primarily rely on manual inspection, which is costly and lacks stability. This study introduces advanced deep learning techniques from the domain of artificial intelligence, which collects and annotates 3 192 high quality images of banded carbides on steel bars, alongside 11 complete samples to create a banded carbide dataset on object detection for steel bars (BCDOD). Common deep learning methods for object detection are applied to the dataset through experimental analysis. With a focus on the specific characteristics of the application scenario and data, the cascade R-CNN model is enhanced with rotation data augmentation, improvement to the Focal Loss function and negative sample fine-tuning, resulting in performance improvement. The achieved average precision reaches 96%, with 100% recognition accuracy on complete sample data, showcasing promising outcomes that address the existing gap in artificial intelligence technology within the field of carbide metallographic detection.

**Key words:** carbide; metallographic structure; defect detection; object detection; cascade R-CNN

## 引言

钢铁行业中,碳化物作为钢材的一种组成成分,可以有效提高钢材的硬度和强度。然而,在钢材的冷却过程中,浓度过高的碳可能以碳化物的形式结晶析出,形成网状、带状等碳化物富集区域,造成钢材的缺陷,这类缺陷会使得钢材具有很高的脆性,并且韧性大大降低,导致产品的耐磨性和疲劳强度显著变差,严重影响钢材的质量<sup>[1]</sup>。因此,对于钢材碳化物的不均匀度检测具有重要意义。然而,现阶段钢铁行业中的碳化物不均匀度金相检测手段主要为人工检测,需要专业性较强的技术人员,成本高昂,效率较低,无法进行高效、大规模且稳定的应用。

近年来,人工智能领域的深度学习技术得到了快速的发展,其在分类、检测、分割等视觉任务上均取得了大量的成果<sup>[2]</sup>。随着卷积神经网络、Transformer等结构的提出,现有的深度学习目标检测算法可以自主学习如何从图像中提取特征,进而应对复杂的任务场景。因此,在碳化物检测领域引入人工智能技术可以消除人员主观性误差,降低人力成本,提高碳化物不均匀度分析的效率和准确性。同时,它还能推动钢铁行业金相检测信息化,降低门槛,减少对检测人员状态、能力和经验的依赖,使检验判定结果更可靠。

在工业领域,深度学习方法带来了出色的检测精度、稳定性以及效率,被广泛应用于缺陷检测、质检等领域<sup>[3-4]</sup>。Xu等<sup>[5]</sup>在钢辊细微缺陷检测中使用标签膨胀的方式增加缺少类别样本,解决样本不平衡问题,接着通过半监督学习方法,通过分析特征图的响应强度获取含缺陷的图像,扩充原有数据集,提升了检测精度。叶卓勋等<sup>[6]</sup>提出一种改进的YOLOv4模型,使用GhostNet替代YOLOv4主干网络,并通过改进式PANet结构增加YOLO预测头中高维特征图,显著提升工业环境小目标缺陷的预测速度和性能。李刚等<sup>[7]</sup>提出一种基于注意力机制的轻量级工业缺陷检测网络,通过通道注意力和坐标注意力机制获取到更丰富的特征信息,在满足低硬件配置的同时,增强了模型针对复杂多样缺陷的检测性能,并在NRSD-MN、NEU-DET等多个公开数据集上取得了理想的效果。He等<sup>[8]</sup>提出了一种基于深度学习的钢板缺陷检测系统,通过多级特征融合和区域建议网络实现精确定位和分类,同时建立了NEU-DET数据集进行训练和评估,最终模型以20 ft/s的速度进行实时检测,达到较高的精度。

以上方法针对不同的工业场景,在不同的缺陷检测数据集显示出了优秀的性能,然而在轴承钢棒材碳化物检测领域,人工智能技术仍处于空白,尚不存在相应的数据集。基于上述情况,本文通过在工厂实地采集、标注数据,自建了一个具有多个等级的工业场景下的棒材带状碳化物目标检测数据集(Banded carbide dataset on object detection for steel bar, BCDOD)。在自建数据集的基础上,使用多种经典的目标检测模型对数据集进行测试,根据实验结果,对级联R-CNN进行改进,针对数据集的不足进行了数据的旋转增强,针对模型的不足提出了Focal Loss函数的改进,并针对应用场景提出了负样本微调,最终本文提出的改进级联R-CNN在BCDOD上进行了多项测试,取得了较为理想的结果。

## 1 带状碳化物数据集的构建

根据数据的细粒度不同,BCDOD数据集中包括3 192张含有带状碳化物的图像级数据,以及11个用于实际场景测试的样品级数据。图像级数据中每张图像都包含至少一个带状碳化物的检测框标签。样品级数据中每个样品给出一个最高等级标签。下面将介绍BCDOD数据集的构建过程,包括数据采集、数据标注和数据集制作3个阶段。

### 1.1 分级标准

根据国家标准GB\_T 18254—2016,钢材碳化物不均匀度检测是基于试样检验面上最严重的视场和区域进行评级。对于带状碳化物,参照标准与行业专家经验,根据碳化物的宽度、跨视场程度和密集

程度进行判断和分级,将带状碳化物分为1级至4级,每隔0.5级进行划分,共7个类别。在后续以带状+等级的形式表示,如带状1代表1级带状碳化物。

## 1.2 数据采集

数据采集阶段,在钢厂的实际生产场景中对样本进行了筛选和拍摄,总共采集了133个轴承钢样品的数据,拍摄了75 195张图片。

轴承钢碳化物带状检验样品的制成方式分为3个步骤,第一步通过对棒材轴承钢进行横切,第二步对表面进行抛光,最后一步用4%硝酸酒精溶液浸蚀,得到碳化物带状检验样品。根据样品的类型不同,样品存在不同的制成形状与制样方式,最终碳化物的走向主要可分为纵向和横向两种。不同样品的形状和制样方式如图1所示。目前每张图像根据标准要求均裁剪为1 463像素×1 463像素,对应710 μm×710 μm的视场。部分棒材带状碳化物采集图像如图2所示。

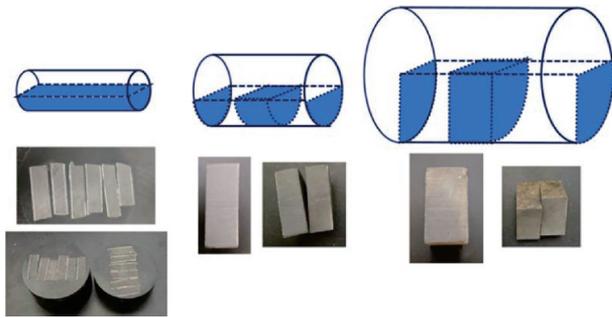


图1 各式样品及其制样方式

Fig.1 Various sample types and their preparation methods

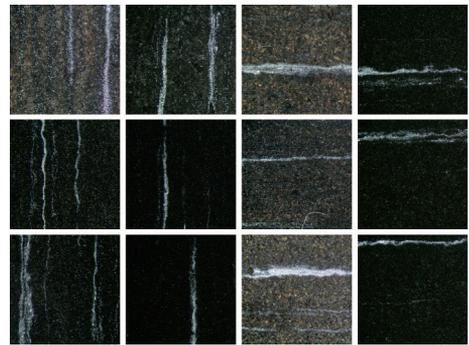


图2 棒材带状碳化物数据集部分图像

Fig.2 Partial images of the banded carbide dataset for steel bars

## 1.3 图像级数据的标注

图像级数据标注的过程主要分为3个阶段。首先对数据进行分析,并做初步的筛选;然后对筛选出的有效样本进行带状碳化物预标注;最后与专家一同完成数据的精确标注。具体过程如下。

### (1) 数据筛选

本项目拍摄的样品图片超过7万张,其中部分样品由于存储过程不当,或是制样过程不够严谨,存在一定的污染现象。对这类污染的图像通过人工进行了初筛。

### (2) 预标注

在经过上述的数据筛选后,得到了数千张污染少、高质量图片并对其中带状碳化物进行后续的位置和等级标定。在这一过程中,本文使用了前期标注的200余张图像在YOLOv5模型上训练<sup>[9]</sup>,得到了一个简单的碳化物目标检测模型,并使用该模型对剩余图像进行自动预标注。

### (3) 专家标注

最终的标注工作在预标注基础上与工厂内的专家合作完成,符合实际生产的标准,有效标注数据样本统计如表1所示。该数据集为目标检测数据集,存在单个视场(图片)及多条带状碳化物、完成多个标注的情况。

表1 图像级样本标注数据统计

Table 1 Statistics of image-level annotation data

带状碳化物等级	图片数量	标注数量
带状1	937	1 581
带状1.5	1 348	1 543
带状2	679	729
带状2.5	200	202
带状3	21	21
带状3.5	7	7
带状4	0	0
总数	3 192	4 083

#### 1.4 样品级数据的标注

依据国家标准,钢材碳化物不均匀度这一检测项目是在试样检验面上以碳化物析出最严重的视场区域作为评级依据。因此,相较于图像级数据,样品级数据是按照实际生产场景中的要求,对整个钢材样品(包含多张图像)给出最终的评级结果作为标签,旨在模拟真实的工业场景,并测试模型在实际应用的效果。目前由于生产条件限制,共采集了11个样品,包含10 985张图像。

#### 1.5 数据集划分与统计

根据细粒度的不同,数据分为图像级以及样品级。其中,图像级数据主要用于训练和验证模型的性能,样品级数据用于测试实际工况场景下模型的效果。对于图像级数据,将每个类的数据以8:1:1的比例进行训练、验证和测试划分。对于样品级数据,将11个样品全部用于测试。具体数据集划分统计如表2所示,其中Test代表图像级数据的测试划分,Test-S代表样本级数据。

表2 数据集统计

Table 2 Statistics of dataset

数据集	带状0	带状1	带状1.5	带状2	带状2.5	带状3	带状3.5	带状4	总量
总数据	0	937	1 348	679	200	21	7	0	3 192
Train	0	751	1 080	545	160	17	7	0	2 560
Val	0	93	134	67	20	2	0	0	316
Test	0	93	134	67	20	2	0	0	316
Test-S	1	0	0	1	6	3	0	0	11

注:该表格按照单个样本的最高带状碳化物等级标签作为展示等级。

## 2 带状碳化物目标检测

### 2.1 基线模型及测试

#### 2.1.1 基线模型选择

本节主要测试了数据集在不同基础模型上的效果,并分析结果,决定后续算法改进的方向。现在目标检测领域有多种基于深度学习的算法,如两阶段的Faster R-CNN、Cascade R-CNN,一阶段的RetinaNet、YOLOv5、YOLOv8,以及基于Transformer的DETR(Detection Transformer)等。两阶段模型中,Faster R-CNN<sup>[10]</sup>通过区域生成网络(Region proposal network, RPN)生成候选区域,然后对候选区域进行特征提取,并进行分类和回归,得到检测结果。Cascade R-CNN<sup>[11]</sup>引入级联检测策略,逐级提升检测器性能,加强检测器回归精度。单阶段模型中,RetinaNet<sup>[12]</sup>引入Focal Loss解决难例学习问题,同时用多尺度特征金字塔(Feature pyramid network, FPN)获取更多维度的特征信息。YOLOv5<sup>[13]</sup>利用CSPNet骨干网络增强特征提取,采用空间金字塔池化提取多尺度特征,使用Mish损失函数和Mosaic增强策略,能够快速精准地检测目标。YOLOv8相较于YOLOv5进一步在骨干网络、检测头等进行了升级。DETR<sup>[14]</sup>基于Transformer,利用全局上下文信息进行目标检测,采用了可以同时考虑分类和定位误差的损失函数,同时使用匈牙利算法来匹配预测和真实目标之间的对象,实现了端到端检测。上述算法在多种目标检测任务中表现良好,作为基线模型在BCDOD数据集进行测试。

#### 2.1.2 评价指标

在图像级数据上,选取平均精度(Average precision, AP)和平均召回率(Average recall, AR)作为评价指标<sup>[15]</sup>。在样品级数据上,选择准确率(Accuracy)作为评价指标。接下来将依次介绍评价指标的相关概念。

精度(Precision): TP占TP与FP和的比例,即

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

式中:TP代表候选框中与真值交并比大于阈值、且类别相同或相邻的检测框数量,这是由于基于课题要求,在实际生产中允许机器判定的类别与人工相差0.5个等级;FP代表预测的候选框中,与真值候选框的交并比小于阈值或类别不为相同或相邻的检测框,或者是检测到同一个真值候选框的多余检测框的数量。精度也被称为查准率,衡量模型预测结果的精准程度。

召回率(Recall):检测到的正例TP占实际总正例(TP + FN)的比例,即

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

式中FN代表未被识别的真值。召回率也即查全率,衡量模型是否能够检测到所有的真值。

平均精度(Average precision):精度通过对类别符合要求的预测框根据置信度排序,在不同置信度阈值下求精度与召回率的乘积求平均,即

$$\text{AP} = \frac{1}{n} \sum_{k=1}^n P(k) \cdot R(k) \quad (3)$$

式中: $k$ 表示预测框的数量, $P(k)$ 表示前 $k$ 个预测框的精度, $R(k)$ 表示前 $k$ 个检索的召回率。例如:AP@0.5指IoU阈值为0.5时计算的AP;AP@0.5:0.95指AP在交并比阈值分别为0.5、0.55、0.6、...、0.95时的精度平均值。

平均召回率(Average recall):评估在不同的交并比阈值下的召回率表现。在计算平均召回率时,将样品的召回率在不同的交并比阈值下进行平均,即

$$\text{AR} = \frac{1}{n} \sum_{i=1}^n \text{Recall}_i \quad (4)$$

式中: $i$ 表示第 $i$ 个交并比阈值, $n$ 表示选取的交并比阈值的数量。另外在计算AR时会标注出计算时每张图片选取候选框的最大数量,例如:AR10表示对于IoU阈值分别为0.5、0.55、0.6、...、0.95时,对每个IoU阈值,将所有预测框按照得分从高到低排序,选取置信度最高的10个候选框,计算这些候选框召回率,最后对不同阈值的召回率取平均值作为结果。

准确率(Accuracy):评估在样品级数据上模型的表现效果,即

$$\text{Acc} = \frac{N_{\text{True}}}{N} \quad (5)$$

式中: $N_{\text{True}}$ 表示等级判断正确的样品, $N$ 表示样品的总数量。

### 2.1.3 基线测试结果

基线模型使用图像级数据进行测试。测试过程中,Faster R-CNN(FPN)、RetinaNet、Cascade R-CNN(FPN)、DETR均使用在ImageNet数据集<sup>[16]</sup>上预训练的ResNet-50网络作为骨干网络,YOLOv5和YOLOv8模型均使用对应的Large模型,同时分别使用两种在MS-COCO数据集上预训练的CSP-Darknet变体作为骨干网络。所有模型

均使用随机梯度下降(Stochastic gradient descent, SGD)作为优化器,学习率(Learning rate, LR)为0.02,动量(Momentum)为0.9,权重衰减(Weight decay)为0.0001,训练轮次为100,批大小(Batch size, BS)为4。基线模型测试结果如表3所示。

表3 基线模型测试结果

Table 3 Baseline model test results				
Method	AP@0.5	AP@0.75	AP@0.5:0.95	AR10
Faster R-CNN(FPN)	0.949	0.740	0.642	0.780
Cascade R-CNN(FPN)	<b>0.952</b>	0.783	0.677	0.798
RetinaNet	0.910	0.645	0.593	0.799
YOLOv5	0.892	0.660	0.606	<b>0.801</b>
YOLOv8	0.912	0.465	0.544	0.689
DETR	0.871	<b>0.794</b>	<b>0.683</b>	0.800

碳化物检测的关键指标是对于碳化物类别预测的精度,而对于预测框位置的准确度并不严格要求。IoU指的是真值框与预测框的交并比,这一数值在0.5以上即代表二者已有较高的重合度,可以视为模型已经成功检出目标的所在位置。因此,选择AP@0.5作为在图像级数据上主要的评价指标。分析上述结果,两阶段模型在AP@0.5的效果普遍更好,DETR在AP@0.5:0.95效果最好,说明DETR在不同IoU阈值有较稳定的结果。Cascade R-CNN在关键指标AP@0.5指标的精度最高,在其他指标也有着不错的表现,故选择使用Cascade R-CNN作为基础模型,并在后续加以改进。

## 2.2 带状碳化物目标检测模型

基于改进级联R-CNN的钢材碳化物检测模型是一个两阶段目标检测器,主要分为骨干网络、特征金字塔、区域提议网络、级联式检测器几个部分。根据任务目标与数据特点,针对数据增强方式、损失函数、负样本微调等对模型进行了改进,模型整体结构如图3所示。

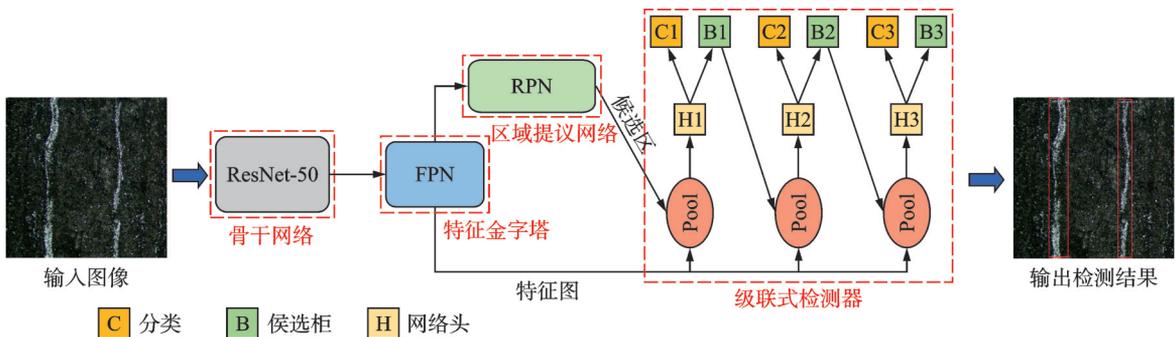


图3 模型结构示意图

Fig.3 Structure diagram of model

### 2.2.1 模型介绍

模型的第一部分是用于提取特征的骨干网络,其主要任务是从原始图像中提取出目标的相关特征,输出提取后的特征图,为后续的模块提供图像的特征信息。模型采用的骨干网络是ResNet-50<sup>[17]</sup>,其核心思想是引入残差连接,避免梯度消失问题的出现,使得网络可以训练更深的层次。

第二部分为特征金字塔<sup>[18]</sup>,是一种用于解决多尺度物体检测问题的卷积神经网络结构。通过骨干网络输入不同尺度的特征图,得到特征融合后的多尺度特征图。传统的卷积神经网络结构只能在固定尺度下提取特征,难以适应不同尺度的目标物体。而特征金字塔通过同时在多个尺度下提取特征,利用不同层级的特征来处理不同尺度的物体,提高了检测的精度。

第三部分为区域提议网络,即RPN。RPN在特征金字塔输出的特征图上通过预设的锚框,使用滑动窗口生成一系列候选框。然后,对每个候选框进行分类和回归,以判断该候选框是否包含物体,并且调整该候选框的位置和大小,使其更加准确地拟合物体。RPN加速了目标检测的速度,提高了目标检测的准确性。在经过RPN网络后,模型会输出多个候选区域用于后续检测。

第四部分为级联式检测器,通过多级检测器级联的方式来逐步提高检测精度。它由多个级联层组成。每个级联层包含1个特征提取器,1个分类器和1个目标框回归器,对于RPN网络输入的候选区,特征提取器使用RoI Align算法进行特征提取,然后经过分类和回归得到该层的输出。前一层目标框回归器的输出将用作后一层的输入,从而使得目标框在每次回归中都获得更好的效果。随着层级的深入,检测器的交并比阈值不断升高,从而过滤掉效果不好的检测框,筛选出高质量的检测目标。通常情况下,前侧的检测器具有较高的召回率和较低的精度,后侧的检测器具有较高的精度和较低的召回率。最终以最后一个检测器的结果作为整个级联式检测器的输出。该结构的优点在于能够逐步提高检测

精度,同时减少误检率和漏检率,进而增加模型的准确率<sup>[19-20]</sup>。

### 2.2.2 旋转数据增强

当前数据集存在数据样本不均衡的问题,高等级样本的数量远小于低等级样本。由于高等级样本数量少,直接使用原始数据进行训练容易导致模型过拟合。使用数据增强方法可以增强数据的多样性,传统的数据增强方法包括但不限于翻转、缩放、标准化等,这些增强方法均可以一定程度增强模型的性能以及泛化性。在此基础上,针对目前数据集的特征,即当前所有的碳化物标注均为横向或纵向,提出了固定角度旋转增强的方式。相较于随机角度的旋转增强,使用 $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ 对数据进行固定角度的增强可以使带状碳化物仍然保持纵向或横向,更加贴近实际,同时减少随机角度产生的无意义数据。可以使模型适应不同方向和形态的样本,有效增加训练数据样本的数量,减少过拟合的风险,提高模型的泛化能力和鲁棒性。

碳化物主要分为横向和纵向两种走向,固定角度的旋转增强可以模拟这一特点,生成符合实际的样本。同时,相较于其他数据增强手段,旋转增强的计算量小,只需要通过矩阵乘法进行坐标变换,计算成本较低。此外,旋转数据增强对检测目标的信息没有影响,不会破坏或丢失图像中的关键信息。旋转数据增强公式如下

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (6)$$

式中: $R$ 表示旋转矩阵; $\theta$ 表示旋转的角度, $\theta$ 取 $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ 。

$$I_{\text{rot}}(x, y) = I \left( R \begin{bmatrix} x \\ y \end{bmatrix} \right) \quad (7)$$

式中: $I_{\text{rot}}$ 表示旋转后的图像, $I$ 表示原始图像, $x$ 、 $y$ 表示图像的像素坐标。旋转数据增强样例如图4所示。

### 2.2.3 损失函数优化

当前数据集中带状碳化物1级和1.5级数据较多,且这两个等级的图像特征相似,因此数据集中的部分数据特征易于混淆,从而产生了部分的难样本。另外,当前数据集存在数据样本不均衡的问题,高等级样本数量更少,对于样本数量少的类别,模型可以学习的类别特征有限,增加了模型判断这些类别的难度,同时模型训练时会偏向样本数量多的类别,对于样本少的类别特征提取效果较差,同样产生部分难样本。

对于原版的Cascade R-CNN,其在级联检测器中的分类头使用的损失为交叉熵(Cross entropy, CE),它通常用于衡量两个概率分布之间的相似度或差异性,在目标检测任务中,常用于分类头,以衡量模型预测结果与实际标签之间的差异性。交叉熵损失公式如下

$$CE(p, y) = - \sum_{i=1}^C y_i \log p_i \quad (8)$$

式中:CE代表交叉熵, $p$ 代表预测概率, $y$ 代表标签, $C$ 代表类别数量, $y_i$ 代表第 $i$ 类的标签(一般 $y_i=1$ 代表正样本, $y_i=0$ 代表负样本), $p_i$ 代表第 $i$ 类的预测概率。

然而,标准的交叉熵损失函数会使模型更偏向样本数量较多的类别,忽视样本较少的类别,且并未针对难样本的学习进行优化,对这些样本学习效果较差。针对上述问题,对损失函数进行优化,使用Focal Loss代替原有的交叉熵损失,能够更好地学习数据中的难样本,同时缓解数据样本不均衡的问题。Focal Loss对于易分类的样本给予更小的权重,对于难分类的样本给予更大的权重。通过加权的方式,可以使模型更加关注难分类的样本,提高模型在难样本上的性能,进而一定程度缓解数据样本不

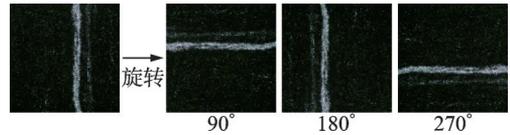


图4 旋转数据增强样例

Fig.4 Example of rotational data augmentation

均衡的问题,有效解决了一般的交叉熵损失函数容易被易分类样本所主导的问题。Focal Loss公式为

$$FL(p, y) = - \sum_{i=1}^C (1 - p_i)^\gamma y_i \log p_i \quad (9)$$

式中:FL代表Focal Loss; $\gamma$ 为调节参数,根据过往研究的经验,取 $\gamma$ 为2。

#### 2.2.4 负样本微调

在目标检测任务中,负样本常指背景或非目标物体的样本,而正样本则指目标物体的样本,因此划痕、气泡等干扰因素均被定义为负样本。然而,当前的训练集主要针对碳化物识别,并未克服干扰因素的影响。在数据集构建过程中,样品易受各类污染的影响。比如,制样和检测时样品互相碰撞可能会产生类似碳化物的白黄色条状划痕;存放时间过久的样品会因锈蚀导致背景发黄,影响对白色碳化物的判断;制样时样品接触部位的气泡也可能造成局部区域的光斑,对碳化物的定位和分级产生干扰。

污染样品现象示意图像如图5所示。目前的训练集中不包含以上干扰因素的图像,因此模型不具备辨别这些干扰因素的能力。引入负样本,可以有效增强模型识别干扰的能力,减少其对检测结果的干扰,降低误判率。具体来说,在模型训练完成后,对原有训练集进行扩充,从原始数据中挑选2 000张左右的污染图片,包含划痕、气泡、样品边缘及生锈等。扩充后的训练集包括原有的2 560张含标注的无污染图像以及2 191张无标注的污染图像。使用扩充后的数据集对训练好的模型进行微调,可以显著增强模型对抗干扰因素的能力,负样本的引入也增加了数据场景的多样性,使数据类型更贴近真实应用场景,提高了模型的适应能力。

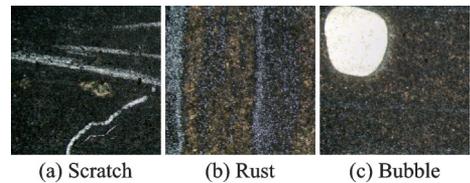


图5 干扰因素示例图像

Fig.5 Example images with interference factors

### 3 实验结果

#### 3.1 图像级检测结果

首先,在图像级数据上测试了模型识别碳化物的能力,并与之前的基线模型进行了比较,各项超参数与2.1.3中基线模型的设置一致,同时,改进模型在训练了100轮次后,使用负样本扩充的数据集进行额外20个轮次的负样本微调。结果如表4所示,其中GFLOPS表示浮点运算数。

在图像级数据中,改良后的算法在关键指标AP@0.5,AP@0.5:0.95,AR10取得最佳,印证了改进措施的有效性。在AP@0.75略低于DETR,说明算法在高IoU阈值的情况下候选框定位仍有提升空间,但整体上各指标都显著优于其他基线模型,表明改进算法在不同阈值下具有良好的检测稳定性。模型在AR10指标上达到最优,说明数据增强、负样本微调等措施扩充了样本的多样性,有效增强模型的泛化性能。

表4 图像级数据实验结果

Table 4 Experimental results on image-level data

Method	AP@0.5	AP@0.75	AP@0.5:0.95	AR10	GFLOPS
Faster R-CNN(FPN)	0.949	0.740	0.642	0.780	134
RetinaNet	0.910	0.645	0.593	0.799	149
YOLOv5	0.892	0.660	0.606	0.801	56
YOLOv8	0.912	0.465	0.544	0.689	83
DETR	0.871	<b>0.794</b>	0.683	0.800	61
Cascade R-CNN(FPN)	0.952	0.783	0.677	0.798	162
Ours	<b>0.960</b>	0.790	<b>0.689</b>	<b>0.810</b>	162

由于碳化物检测更看重分类指标,改进模型在AP@0.5指标上的优秀表现验证了所提出方法的有效性,基本达到了工业场景对带状碳化物检测的要求。如图6所示可视化检测结果,对于识别难度较大的边缘高等级碳化物,改进模型的识别效果相较于基线模型无论在定位框还是分类,均取得了更好的效果。改进模型对于带状碳化物区域的划分也更加精准,减少了区域混淆的情况,如图6右上图所示。同时,对于一些低等级碳化物,改进模型同样表现了更精确的检测结果,如图6下侧两张样例图所示。总体来说,改进模型在保证分类准确率的前提下,提升了模型的检测召回能力,达到了很好的综合效果。

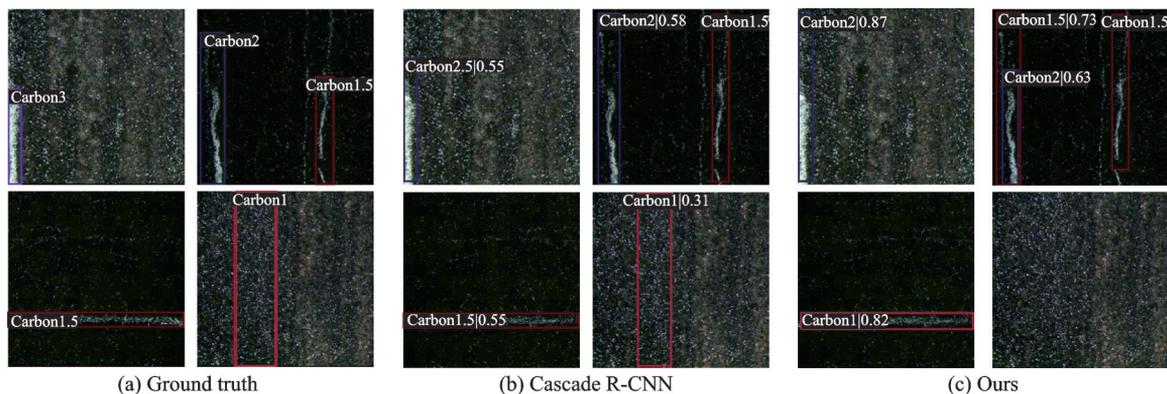


图6 图像级数据上的检测结果

Fig.6 Detection results on image-level data

模型的计算复杂度上测试了浮点运算数,由于当前的模型相较于其他几种方法有着更深的网络结构、更多的参数和更多的层,这使得它能够更好地捕捉复杂的数据模式和关系。与此同时,更复杂的模型带来了更高的计算复杂度,在实际使用算法时,图像采集与图像检测同时进行,由于采集设备的速度限制,目前的时间瓶颈仍在采集阶段,当前的模型复杂度仍然可以接受。

进一步分析图7的混淆矩阵结果,该混淆矩阵的IoU阈值为0.5,置信度阈值为0.3,最下方的一行Background代表本没有标注的位置被模型识别成了碳化物,最右侧的一列Background代表FN,即未被预测到的真实标签。由于允许候选框的类别存在一个类别的误差,因此对角线上的元素,以及对角线元素左右两侧的元素均算作TP。在此基础上,可以看到模型在IoU=0.5的情况下,对于所有类别碳化物的召回率均达到90%以上,对于1.5级碳化物及以上的召回率达到97%以上,对于2级及以上的碳化物召回率达到99%以上。混淆矩阵的最后一行表示本不存在碳化物标注的地方被识别出碳化物,可以看到,这种情况绝大部分的错误结果来自1.5级及以下。在实际场景下,检测高等级的碳化物具有更高的重要性,因此2级及以上的碳化物的准确率,其重要程度远大于1.5级及以下,而当前模型对于1.5级以上的碳化物识别能力已经达到了较高的水平,基本满足工业场景下对于模型算法精确度的要求。

### 3.2 样品级检测结果

除了在图像级数据上检验模型对于碳化物的识别能力,模型还在样品级数据验证实际场景下模型的检

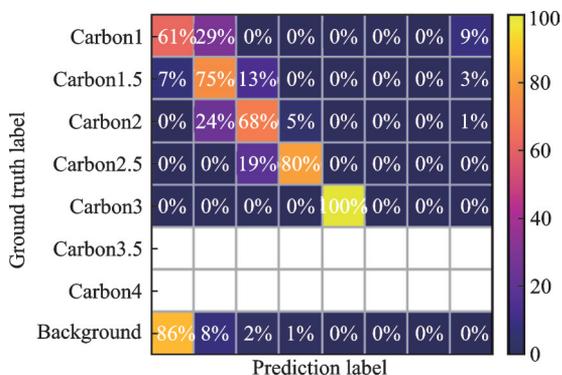


图7 测试集结果的标准化混淆矩阵

Fig.7 Normalized confusion matrix on results of test dataset

测性能,结果如表5所示。每个样品按照模型检测出的最大等级带状碳化物作为结果,并与真值进行对比。

表5中的Label表示真实标签,Baseline指基线模型Cascade R-CNN,Acc.表示准确率。由表5中结果可知,在目前获得的11个完整样品上,相较于基线模型,当前的模型已经取得了100%的准确率。尽管当前样本数量不多,但每个样品包含近千张图像,信息量大,内在复杂度高。同时,100%的准确率验证了模型的效果并非偶然,表明模型对复杂场景有比较好的适应能力,取得了比较理想的效果,证明了模型改进方法的有效性。

表5 样品级数据实验结果

Table 5 Experimental results on sample-level data

编号	1#	2#	3#	4#	5#	6#	7#	8#	9#	10#	11#	Acc/%
Label	2.5	0	2.5	2.5	2.5	3	3	2.5	2.5	3	2	
Baseline	2.5	1.5	2.5	2.5	2.5	3	2.5	2	1.5	2.5	2	82
Ours	2.5	1	2.5	2.5	2.5	3	2.5	2.5	2	2.5	2.5	100

### 3.3 消融实验

消融实验部分,每个实验的超参数进行了统一,与2.1.3中基线模型的设置一致,除改进步骤外,其他实验条件均相同。首先,在图像级数据上,针对数据增强、Focal Loss函数优化以及负样本增强3个改进逐步进行了消融实验,结果如表6所示。样品级数据相较于图像级数据,在数据种类上增加了实际生产中可能遇到的干扰因素,负样本微调正是为了解决这一问题。因此,该实验主要验证负样本微调的有效性,结果如表7所示。

表6中,可以看到仅使用Focal Loss或旋转数据增强,均可以提升模型各指标的性能。而组合使用会使AR10和AP@0.75指标略微下降,但在AP@0.5和AP@0.5:0.95均进一步获得了提升,说明Focal Loss在与旋转数据增强组合使用过程中牺牲了部分易分类样本,降低了召回率,但增强了低IoU阈值下的精度,同时在整体的精度稳定性上表现达到最好。在加入负样本增强后,AP@0.75达到最佳,虽然这一措施略微降低了AP@0.5和AP@0.5:0.95,但这一指标仍保持在96%以上,同时让召回率指标AR10有一定的回升,达到0.81。最重要的是,负样本有效增强了模型抵抗真实场景中干扰因素的能力,体现在表7样品级数据上的准确率。未经负样本增强的模型在样品级数据中存在两处误判,而经过

表6 图像级数据消融实验结果

Table 6 Results of ablation experiment on image-level data

Method	AP@0.5	AP@0.75	AP@0.5:0.95	AR10
Baseline	0.952	0.783	0.677	0.798
Baseline + FL	0.958	0.790	0.687	0.802
Baseline + DA	0.955	0.785	0.686	<b>0.814</b>
Baseline + DA + FL	<b>0.964</b>	0.783	<b>0.692</b>	0.803
Baseline + DA + FL + FA	0.960	<b>0.790</b>	0.689	0.810

注:DA代表数据增强,FL代表Focal Loss,FA代表负样本微调。

表7 样品级数据消融实验结果

Table 7 Results of ablation experiment on sample-level data

Method	1#	2#	3#	4#	5#	6#	7#	8#	9#	10#	11#	Acc/%
Label	2.5	0	2.5	2.5	2.5	3	3	2.5	2.5	3	2	
w/o FA	2.5	1.5	2.5	2.5	2.5	3	2.5	2	1.5	3	2.5	82
ALL	2.5	1	2.5	2.5	2.5	3	2.5	2.5	2	2.5	2.5	100

注:w/o FA表示基线模型经过了数据增强与Focal Loss改进,但未经过负样本微调,ALL表示基线模型经过了数据增强,Focal Loss改进与负样本微调。表头的数字#代表每个样品的编号。

负样本增强的模型对于全部 11 个样品数据均取得正确结果,验证了负样本增强的必要性。分析图 8,也可看到使用了负样本增强的模型可以避免将划痕识别为碳化物。同时,负样本增强后的模型在干扰下的识别能力也有提升,在样品受到划痕干扰的情况下仍能识别出带状碳化物,如图 8 中左下样例。综上所述,最终的模型达到了较为均衡且较为理想的结果。

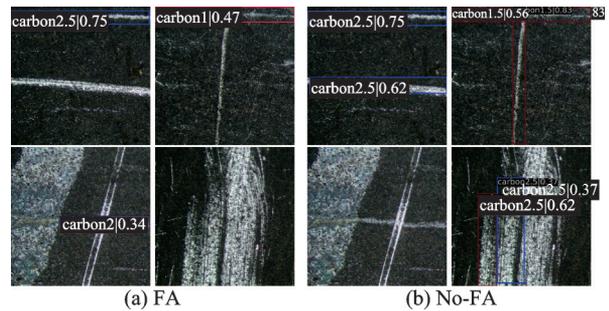


图 8 负样本增强检测结果对比

Fig.8 Comparison of detection results with and w/o FA

#### 4 结束语

在当前工业智能化的时代背景下<sup>[21]</sup>,深度学习在解决工业场景任务展现了巨大的潜力。针对当前碳化物检测的人力成本高、误差大和不稳定等问题,创建了工业场景下的棒材带状碳化物目标检测数据集 BCDOD,对数据集进行了多个模型的基线测试。针对数据和模型的不足,提出了改进级联 R-CNN,采用旋转数据增强与 Focal Loss 函数,并针对实际应用场景提出负样本微调,最终带状碳化物检测分级模型在图像级数据的识别精度达到 96%、在 11 个样品级数据准确率达 100%,基本满足工业生产对于带状碳化物检测的要求。

此外,由于工厂生产条件的限制,目前高等级带状碳化物的样本仍十分稀少,数据集对于 3 级及以上的带状碳化物样本数量不足,难以验证模型对于 3.5 级以上碳化物的有效性。对于完整样品,仍有必要进一步采集样本并对模型在真实场景的有效性进行验证。同时,对于碳化物检测中网状碳化物以及液析的检测,目前的数据集和模型仍未涉及,在未来还需要进一步的扩展完善。

#### 参考文献:

- [1] 薄鑫涛.GB/T 18254—2016《高碳铬轴承钢》标准解读[J].热处理,2019,34(4): 57-60.  
BO Xintao. Interpretation of GB/T 18254—2016 “High Carbon Chromium Bearing Steel” Standard[J]. Heat Treatment, 2019, 34(4):57-60.
- [2] 罗东亮,蔡雨萱,杨子豪,等.工业缺陷检测深度学习综述[J].中国科学:信息科学,2022,52(6): 1002-1039.  
LUO Dongliang, CAI Yuxuan, YANG Zihao, et al. Survey on industrial defect detection with deep learning[J]. Scientia Sinica Informaticis, 2022, 52: 1002-1039.
- [3] 石鑫,化晨冰,张凯,等.基于 SSD 深度神经网络的航拍图像电力目标检测[J].数据采集与处理,2022,37(1): 207-216.  
SHI Xin, HUA Chenbing, ZHANG Kai, et al. Power target detection in aerial images based on SSD deep neural network[J]. Journal of Data Acquisition and Processing, 2022, 37(1): 207-216.
- [4] 唐小煜,熊浩良,黄锐珊,等.基于改进的 U-Net 和 YOLOv5 的绝缘子掩模获取与缺陷检测[J].数据采集与处理,2021,36(5): 1041-1049.  
TANG Xiaoyu, XIONG Haoliang, HUANG Ruishan, et al. Insulator mask acquisition and defect detection based on improved U-Net and YOLOv5[J]. Journal of Data Acquisition and Processing, 2021, 36(5): 1041-1049.
- [5] XU X, ZHENG H, GUO Z, et al. SDD-CNN: Small data-driven convolution neural networks for subtle roller defect inspection [J]. Applied Sciences, 2019, 9(7): 1364.
- [6] 叶卓勋,刘妹琴,张森林.基于轻量化深度学习网络的工业环境小目标缺陷检测[J].控制与决策,2023,38(5): 1231-1238.  
YE Zhuoxun, LIU Meiqin, ZHANG Senlin. Small-scale defect detection in industrial environment based on lightweight deep learning network[J]. Control and Decision, 2023, 38(5): 1231-1238.
- [7] 李刚,邵瑞,周鸣乐,等.基于注意力的轻量级工业产品缺陷检测网络[J/OL].计算机工程: 1-15[2023-06-27]. <https://doi.org/10.19678/j.issn.1000-3428.0066270>.  
LI Gang, SHAO Rui, ZHOU Mingle, et al. Lightweight industrial products defect detection network based on attention[J]. Computer Engineering: 1-15 [2023-06-27]. <https://doi.org/10.19678/j.issn.1000-3428.0066270>.

- [8] HE Y, SONG K, MENG Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1493-1504.
- [9] 许德刚,王露,李凡.深度学习的典型目标检测算法研究综述[J].计算机工程与应用,2021,57(8): 10-25.  
XU Degang, WANG Lu, LI Fan. Review of typical object detection algorithms for deep learning[J]. Computer Engineering and Applications, 2021, 57(8): 10-25.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [11] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2018: 6154-6162.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Columbus: IEEE, 2017: 2980-2988.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2018: 779-788.
- [14] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the European Conference on Computer Vision. Columbus: IEEE, 2020: 213-229.
- [15] 刘瑞昊,于振中,孙强.改进多尺度特征融合的工业现场目标检测算法[J].机械与电子,2022,40(11): 40-45.  
LIU Ruihao, YU Zhenzhong, SUN Qiang. Improved multi-scale feature fusion for industrial field object detection algorithm [J]. Machinery & Electronics, 2022, 40(11): 40-45.
- [16] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2009: 248-255.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2016: 770-778.
- [18] LIN TY, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2017: 2117-2125.
- [19] 胡江宇,贾树林,马双宝.基于改进级联Faster RCNN的PCB表面缺陷检测算法[J].仪表技术与传感器,2022,474(7): 106-110,126.  
HU Jiangyu, JIA Shulin, MA Shuangbao. PCB surface defect detection algorithm based on improved cascade Faster RCNN[J]. Instrument Technique and Sensor, 2022, 474(7): 106-110,126.
- [20] 王子琦,管振玉,朱轶昇,等.基于改进级联RCNN的遥感图像目标检测[J].计算机工程与设计,2023,44(1): 194-202.  
WANG Ziqi, GUAN Zhenyu, ZHU Yisheng, et al. Object detection algorithm of optical remote sensing image based on improved cascade RCNN[J]. Computer Engineering and Design, 2023, 44(1): 194-202.
- [21] 孙郑芬,吴韶波.机器视觉技术在工业智能化生产中的应用[J].物联网技术,2020,10(8): 103-105,108.  
SUN Zhengfen, WU Shaobo. Application of machine vision technology in industrial intelligent production[J]. Internet of Things Technology, 2020, 10(8): 103-105,108.

## 作者简介:



郝亮(1985-),男,博士研究生,研究方向:工业视觉检测,E-mail: 392832849@qq.com。



周诗洋(2001-),男,硕士研究生,研究方向:计算机视觉。



莫允扬(1999-),男,硕士研究生,研究方向:计算机视觉。



陈勇勇(1989-),男,博士,助理教授,硕士生导师,研究方向:模式识别、人工智能。



徐勇(1972-),男,博士,教授,博士生导师,研究方向:计算机视觉。



苏敬勇(1983-),通信作者,男,博士,教授,博士生导师,研究方向:计算机视觉,E-mail: sujingyong@hit.edu.cn。

(编辑:夏道家)