

# 基于改进YOLOX-S网络的金属表面缺陷图像检测

吴超, 叶鑫龙, 蒋劲茂, 徐尚龙\*

电子科技大学机械与电气工程学院, 四川 成都

收稿日期: 2022年7月12日; 录用日期: 2022年8月9日; 发布日期: 2022年8月16日

## 摘要

本文为了实现金属零件制造和加工环节中表面缺陷的图像检测识别, 提高金属零件加工流水线过程中不合格产品的图像检测准确率, 提升金属缺陷检测过程中的设备自动化水平, 解决人工检测中易疲劳、检测效率低、检测精度低、主观性强和无法适应大批量高质量零件的检测问题。提出一种基于改进YOLOX-S的卷积神经网络金属表面缺陷图像检测模型。该模型在原有的YOLOX-S的网络基础上引入注意力机制模块, 优化模型参数, 改进边界框位置回归损失函数与置信度损失函数, 构建和实现对金属表面常见缺陷种类的图像检测模型。改进后的YOLOX-S模型能更快地收敛, 添加边界框位置回归损失函数和注意力机制的模型, mAP值由94.23%提升为96.14%, 准确率提升效果最好, 同时仅增加较少的推理时间。与YOLOX-S模型以及添加不同改进方法模型作对比分析表明基于边界框位置回归损失函数和注意力机制改进的模型综合识别效果最好, 能满足金属表面缺陷图像检测的要求, 减少金属零件生产过程中的缺陷产品流出。

## 关键词

改进YOLOX-S, 注意力机制, 损失函数, 图像检测

# Image Detection of Metal Surface Defects Based on Improved YOLOX-S Network

Chao Wu, Xinlong Ye, Jinmao Jiang, Shanglong Xu\*

School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan

Received: Jul. 12<sup>th</sup>, 2022; accepted: Aug. 9<sup>th</sup>, 2022; published: Aug. 16<sup>th</sup>, 2022

## Abstract

In order to realize the image detection and identification of surface defects in the manufacturing

\*通讯作者。

文章引用: 吴超, 叶鑫龙, 蒋劲茂, 徐尚龙. 基于改进 YOLOX-S 网络的金属表面缺陷图像检测[J]. 机械工程与技术, 2022, 11(4): 384-392. DOI: 10.12677/met.2022.114044

and processing of metal parts, improve the image detection accuracy of unqualified products in the process of metal parts processing assembly line, improve the automation level of equipment in the process of metal defect detection, and solve the problems of easy fatigue, low detection efficiency, low detection accuracy, strong subjectivity and inability to adapt to the detection of large quantities of high-quality parts in manual detection. A convolution neural network model based on improved YOLOX-S is proposed for metal surface defect image detection. The model changed the structure of the unit module and introduced the attention mechanism module based on the original YOLOX-S model, which optimized the model parameters and improved bounding box position regression loss function and confidence loss function, and then the image detection model of common defects on the metal surface was constructed and realized. The results showed that the loss of improved YOLOX-S model could converge faster, the model with bounding box position regression loss function and attention mechanism added, its mAP increased from 94.23% to 96.14%, and the accuracy improvement effect is the best, while only a small amount of reasoning time is increased. Compared with the YOLOX-S model and the model with different improved methods, it is shown that the improved model based on bounding box position regression loss function and attention mechanism has the best comprehensive recognition effect, which can meet the requirements of metal surface defect image detection and reduce the outflow of defective products in the production process of metal parts.

## Keywords

Improved YOLOX-S, Attention Mechanism, Loss Function, Image Detection

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着人类社会的工业革命发展,金属就逐渐成为生产生活以及工业制造中极为重要的原材料。冶炼金属与加工金属的技术不断创新,使得以铁、铝、铜等金属及其合金为主要材料的制品广泛用于汽车、船舶、建筑、航空航天、生产制造等工业领域。在金属零件制造加工环节中,由于人员操作不当或者加工工艺的问题,金属零件会出现如毛刺、条痕、孔洞、裂纹、磨损、尺寸偏差等缺陷[1]。为了保证金属产品的精度,质量和性能,需要对金属产品的表面缺陷进行检测。

对金属产品缺陷检测的方法主要分为人工检测、X射线检测、超声波检测[2]、机器视觉与深度学习检测[3]等等。其中,人工检测、X射线检测、超声波检测等主要依赖于人工肉眼或者物理特性来检测,存在检测效率低,精度低等问题。基于深度学习的图像检测方法克服了人为提取特征的缺点,能够实现快速、准确、自动化的缺陷识别[4]。近年来,关于金属图像检测的深度学习方法不断发展,2020年,李维刚等[5]提出了一种改进的YOLOV3模型用于带钢表面缺陷检测,通过调整网络结构与优化锚框聚类算法等手段改善基准模型性能。2021年,孙连山等人[6]提出了一种改进YOLOV3模型(AM-YOLOV3),改进措施包括添加设计的注意力引导模块、增加预测尺度和使用K-medians聚类生成锚框,在铝材缺陷数据集上mAP达到了0.9905。2022年,Tian和Jia[7]提出了一种基于CenterNet改进的钢表面缺陷目标检测器(DCC-CenterNet),改进措施包括扩展特征增强模块、新的中心权重函数和CIOU等,在NEU-DET数据集上mAP达到了0.7941,速度为71.37 fps。上述大都在YOLO-V3的基准图像检测模型上进行改进,YOLO-V3网络结构的性能存在滞后的问题,逐渐不能满足金属产品表面缺陷新的检测要求。

本文选取最新的单阶段、无锚框的 YOLOX 模型，将其作为基准图像检测模型改进，采用 NEU-DET 数据集[8]与自制金属插头表面缺陷数据集，包含有六种典型表面缺陷，即轧制氧化皮、斑块、开裂、点蚀表面、内含物和划痕，实现金属表面缺陷图像检测。

### 2. YOLOX-S 算法模型

Yolo [9]目标检测系列算法自从 2015 被提出以来，不断地进化发展，成为单阶段目标检测模型中最杰出的物体检测器之一。2021 年，旷世科技提出了以 YOLOV3-SPP 和 ultralytics 公司 YOLOV5 系列为基准模型改进的 YOLOX 系列[10]。YOLOX-S 的组成主要包括三部分，分别是 CSPDarkNet 模型构成的主干特征提取网络、FPN-PAN 结构组成特征提取加强 Neck 和实现结果预测的解耦头，其具体结构见图 1 所示。在主干网络与颈中存在两种跨阶段局部(CSP)结构，分别为 CSP\_block1 与 CSP\_block2。CSP\_block1 的后缀表示残差组件个数，CSP\_block2 的后缀表示卷积组个数，YOLOX 通过设定 CSP 结构大小来控制模型深度。YOLOX-S 由 YOLOV5-S 改进得来，其在网络结构方面的主要区别在于：第一，将基本卷积模块中的激活函数由 Leaky ReLU 换为 SiLU；第二，将预测头部分由传统的耦合的 YOLO Head 改为解耦头。使用解耦头能提高模型精度，还能加快网络的收敛速度。在整体上，YOLOX 使用强大的数据增强、无锚框、多正样本、SimOTA 标签分配策略等技巧来优化模型性能。

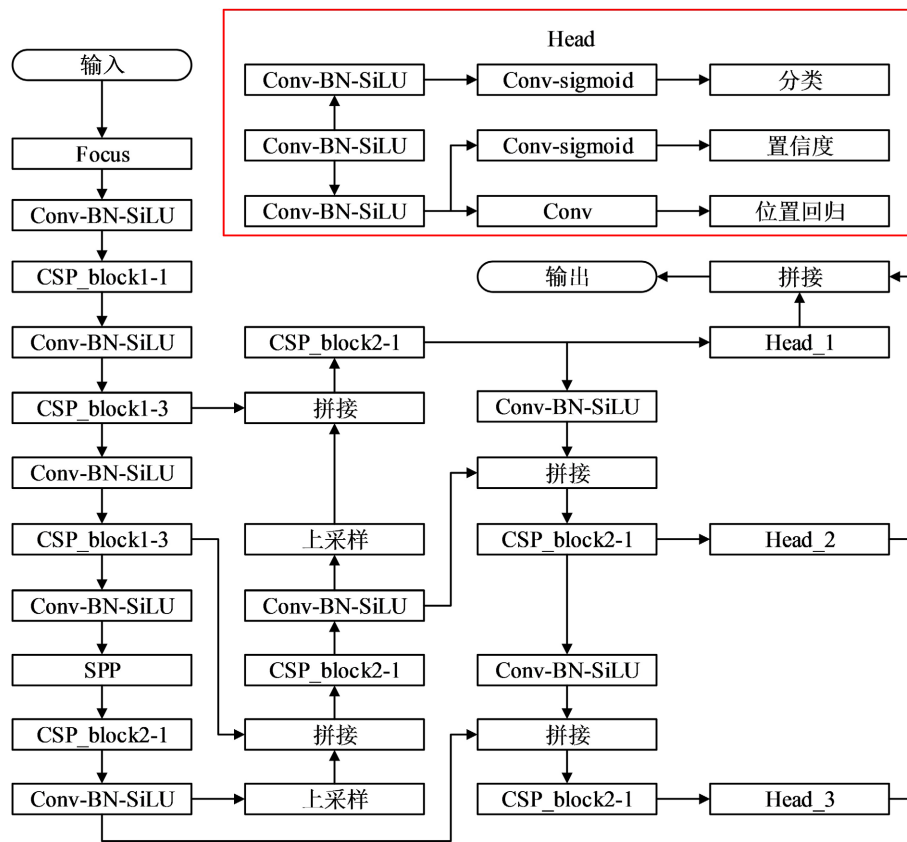


Figure 1. Diagram of the network structure of YOLOX-S model  
图 1. YOLOX-S 模型的网络结构图

### 3. YOLOX-S 模型改进

本文 YOLOX-S 模型的改进方向主要包括两部分，第一，在图像检测模型的基本网络不同位置添加

注意力机制，第二，图像检测模型使用改进的损失函数。通过以上改进，本文的检测方法优势在于通过注意力机制可以提高对某些形状发散，边界模糊以及位置信息难以判断的缺陷分类的准确率，同时改进损失函数考虑了预测框回归的重叠面积，中心点距离、长宽边长真实差，基于 CIOU 解决了纵横比的模糊定义，并添加 Focal Loss 解决预测框回归中的样本不平衡问题，提高了金属缺陷检测定位的精度。

### 3.1. 添加注意力机制

注意力机制是机器学习中一种数据处理方法，广泛应用在自然语言处理、图像识别及语音识别等各种不同类型的机器学习任务中。针对金属检测中某些形状发散、缺陷核心微小、边界模糊的缺陷检测问题，可以对 YOLOX-S 检测模型进行相应的改进，在主干网络与特征加强颈之间的三个特征图中添加注意力机制模块 CBAM，提升模型在该问题上的性能表现。CBAM 是通过通道注意力模块 CAM 和空间注意力模块 SAM 串联组成的，并且通道注意力模块在前，空间注意力模块在后。CBAM 与只关注通道域注意力 SE 模块相比，增加了空间域的注意力机制，强调了空间和通道两个维度上有意义的特征；同时，在实施过程中增加了全局最大池化，两种不同的池化意味着提取更高层次的特征更加丰富，CBAM 的具体实现见图 2。

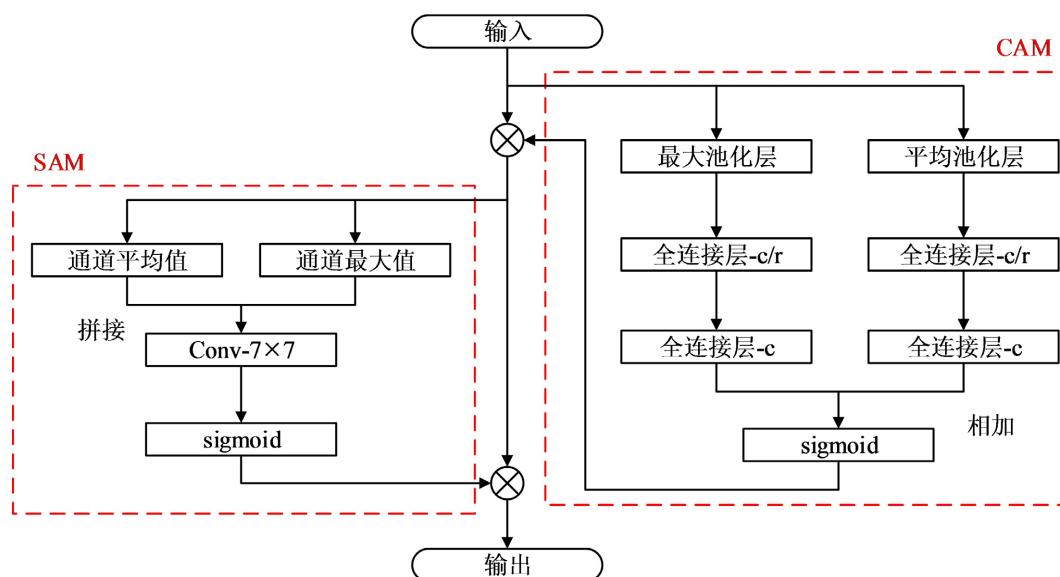


Figure 2. Diagram of attention mechanism CBAM

图 2. 注意力机制 CBAM 图

### 3.2. 改进损失函数

YOLOX 的损失函数的计算公式如(1)所示，由三个部分组成，分别是类别损失  $loss_{cls}$ 、置信度损失  $loss_{obj}$  和边界框位置回归损失  $loss_{reg}$ 。其中类别损失与置信度损失均采用二值交叉熵损失函数计算，而位置损失采用传统的  $IOU\_Loss$  实现。

$$Loss = w_{iou} loss_{reg} + loss_{obj} + loss_{cls} \quad (1)$$

对于 YOLOX-S 图像检测模型原有的  $IOU\_Loss$ ，存在两个不足：一是预测框与目标框不相交时是不能反映两个框的距离远近的，此时  $IOU$  为 0，其损失函数不可导，无法回传梯度进一步学习；二是对于确定大小的预测框与目标框，相同的  $IOU$  值不能反映两个框的相交情况。改进 YOLOX-S 的损失函数主要考虑置信度损失与位置损失，其中置信度损失采用  $Focal\_Loss$  [11] 函数来进行衡量。

*Focal\_Loss* 的计算公式如(2)所示,其主要解决的是在单阶段的目标检测器中正负样本不平衡的问题。*Focal\_Loss* 通过给正样本较大的权重,对于负样本更小的权重,使得正样本在损失函数中的贡献更大,有利于提高网络模型对正样本的识别能力。

$$Focal\_Loss = -\alpha_t (1 - p_t)^\gamma \log(p_t) \tag{2}$$

式中,  $(1 - p_t)^\gamma$  为调制因子;  $\gamma$  是可调的专注参数, 范围为[0, 5];  $p_t$  表示的模型预测属于前景目标的概率  $p$  的函数,  $\alpha_t$  是用于调节样本权重参数  $\alpha$  的函数。

*GIOU\_Loss* 解决了 *IOU\_loss* 未考虑预测框与目标框如何相交的问题,但是存在无法区分预测框在目标框内部时相对位置的问题。与 *IOU\_Loss* 相比,引入了两框最小外接矩形与两框并集的差集形成的惩罚性项,计算公式如(3)所示。

$$GIOU\_Loss = 1 - IOU + \frac{|C - (A \cup B)|}{|C|} \tag{3}$$

式中,  $C$  是预测框与目标框的最小外接矩形。

*EIOU\_Loss* [12]在 *CIOU\_Loss* 的基础上将长宽比的影响因子拆开,分别计算目标框与预测框的长和宽。*EIOU\_Loss* 主要组成包括三部分,分别是重叠损失、中心距离损失和宽高损失,计算公式如(4)所示。

$$EIOU\_Loss = loss_{iou} + loss_{dis} + loss_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \tag{4}$$

式中,  $C_w$  和  $C_h$  两框的最小外接矩形的宽和高。

### 3.3. 实验与分析

本文实验采用 NEU-DET 数据集与自制金属插头表面缺陷数据集, NEU-DET 数据集是东北大学的 NEU 数据集中目标检测部分的数据集,其图像数据部分是 NEU-CLS 数据集构成。数据集的缺陷边界框的标注信息采用 xml 格式文件保存,标注的边界框信息共 4186 个,其中夹杂种类缺陷标注数最多,达到了 1011 个,麻点种类缺陷标注数最少,为 432 个。对于自制金属插头表面缺陷数据集,使用图像标注工具 Labellmg 对含有缺陷的图像进行目标的边界框标注,形成标签数据集。

将数据集按照 7:3 的比例划分为训练集与测试集,训练集中 10%的数据用于验证。采用的实验配置环境见表 1,相关的实验主要分为两个主要部分,一是在主干网络与特征加强颈之间的三个特征图中添加注意力机制模块 CBAM;二是改变网络的置信度损失函数为 *Focal\_Loss* 和预测框位置回归损失函数为 *EIOU\_Loss*。

**Table 1.** Description of experimental environment configuration

**表 1.** 实验环境配置说明

| 环境配置名称 | 配置说明                     |
|--------|--------------------------|
| CPU    | AMD R5 2600 6 核心 3.6 GHz |
| GPU    | Nvidia GTX1070 8 GB      |
| 内存     | 16 GB 2666 频率            |
| 操作系统   | Windows 10               |
| 机器学习框架 | Pytorch1.11              |
| 编程语言   | Python3.9                |

为了观察数据增强的效果,通过设置 COCO 数据集的预训练权重,设置批次大小为 6,训练的 epoch 数设置为 100,在最后的 15 个 epoch 中关闭 MixUp 与 Mosaic 数据增强策略。由图 3 可以看出,在训练的最后的 15 个 epoch 中,目标检测模型的 mAP 值增长明显。这说明,在训练末尾去除数据增强能够有效地改善 YOLOX-S 模型训练指标情况。

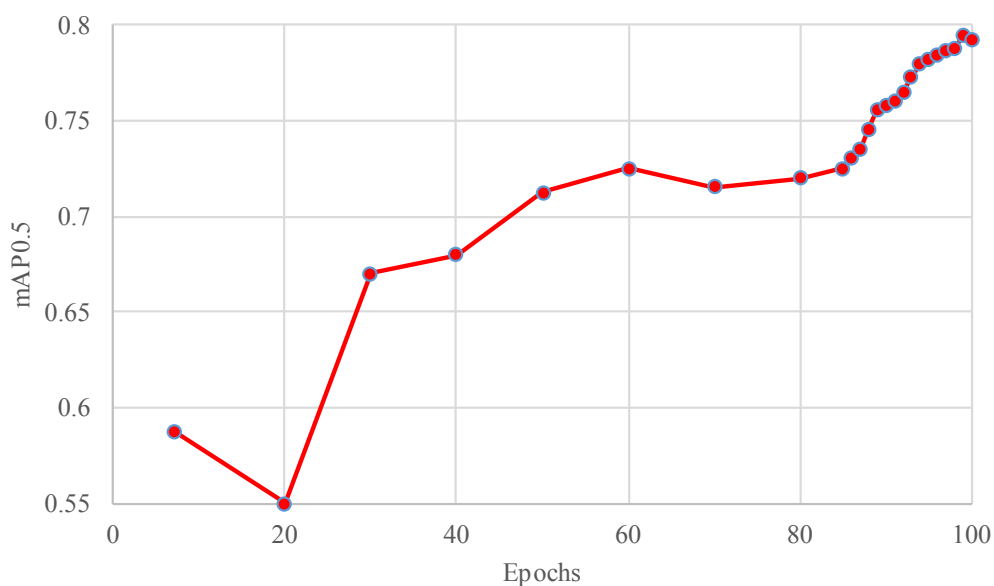


Figure 3. Diagram of mAP changes in model training

图 3. 模型训练 mAP 变化示意图

通过多次改进损失函数和添加注意力机制训练, YOLOX-S 网络改进实验结果见表 2, 首先考虑应用单个改进的 YOLOX-S 的目标检测性能指标情况。其中, 添加注意力机制 CBAM 与采用预测框的位置回归损失  $EIOU\_Loss$  对基准模型的识别性能都略有提升, 分别提高了 0.21% 和 1.23%, 注意力机制表明在获取特征时关注不同特征的情况下, 对某些缺陷检测的提升, 会影响模型对其它种类缺陷的特征提取, 整体识别性能提高不明显。但是, 采用置信度损失  $Focal\_Loss$  使 mAP 下降了 4.38%, 可能的原因是受到噪音的干扰, 存在正负样本标注错了的情况, 就会一直针对训练标错的样本, 同时在推理时间方面有较大增加, 另外单独使用该技巧在模型训练的验证评估环节的时间大幅增加。就单个改进措施而言,  $EIOU\_Loss$  考虑了预测框的宽高差异值, 解决样本中预测框回归不平衡的问题, 对模型整体性能提升较好。

接下来主要考虑添加多个改进措施时的情况。由于 CBAM 对模型提升微小与  $Focal\_Loss$  对模型的负优化, 因此主要讨论  $EIOU\_Loss$  与其他两种技巧的组合情况。从表 2 中可以看出, 采用  $EIOU\_Loss$  与  $Focal\_Loss$  作为改进损失函数时, 模型的 mAP 指标为 89.87%, 相比基准模型减少了 4.36%, 推理时间大幅增加。 $EIOU\_Loss$  与 CBAM 组合对基准模型的性能最好, 解决了缺陷检测中预测框的回归问题, 以及通过空间和通道上两个维度提取更加丰富的特征, 提高了整体缺陷类别识别准确率, 其 mAP 值提高了 1.91, 推理时间仅增加 0.45 ms。

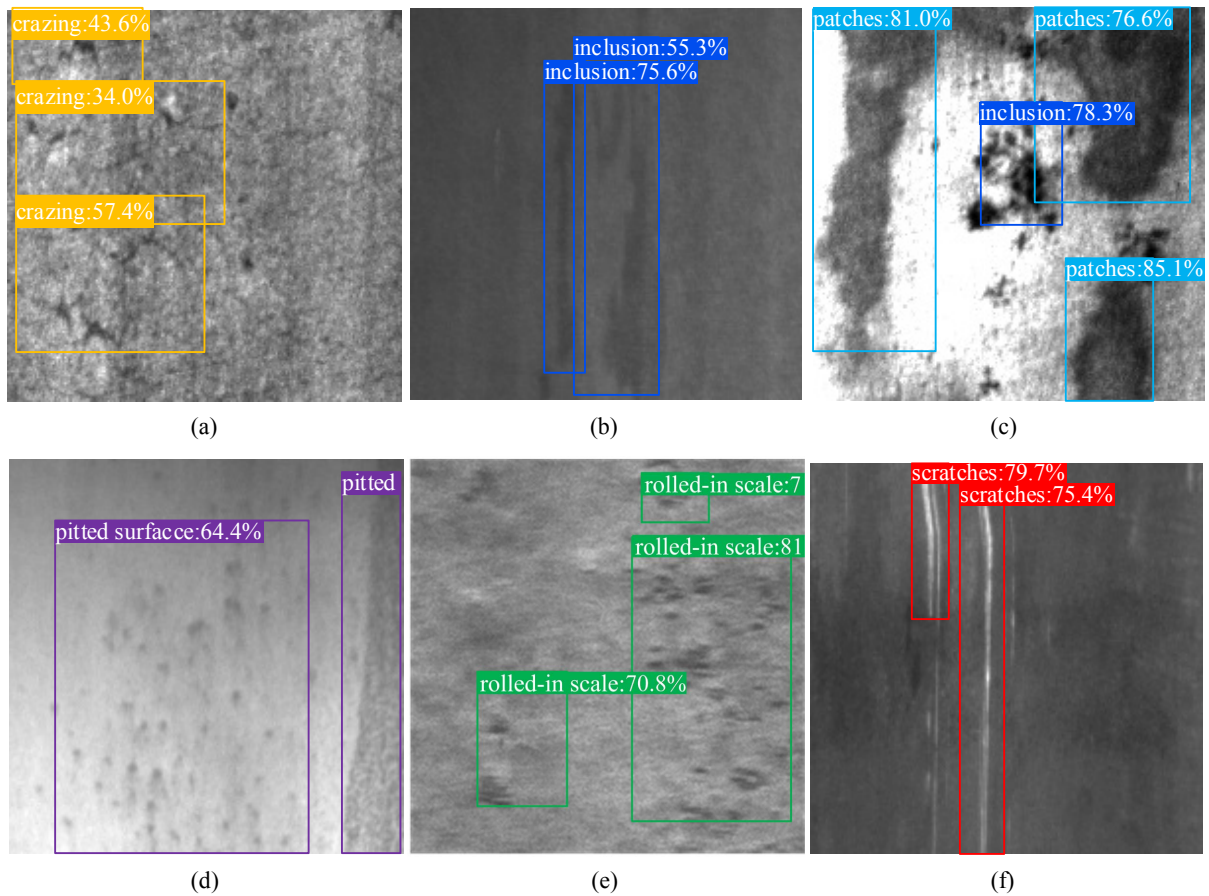
最后, 添加了  $GIOW\_Loss$  与 CBAM 组合的补充对比实验, 结果表明其组合能略微改善单一采用 CBAM 的模型性能。

综上所述, 采用  $EIOU\_Loss$  加注意力机制作为模型预测框位置回归损失对 YOLOX-S 性能提升最为明显, 同时仅增加较少的推理时间。

**Table 2.** Improved YOLOX-S index comparison results  
**表 2.** 改进 YOLOX-S 指标对比结果

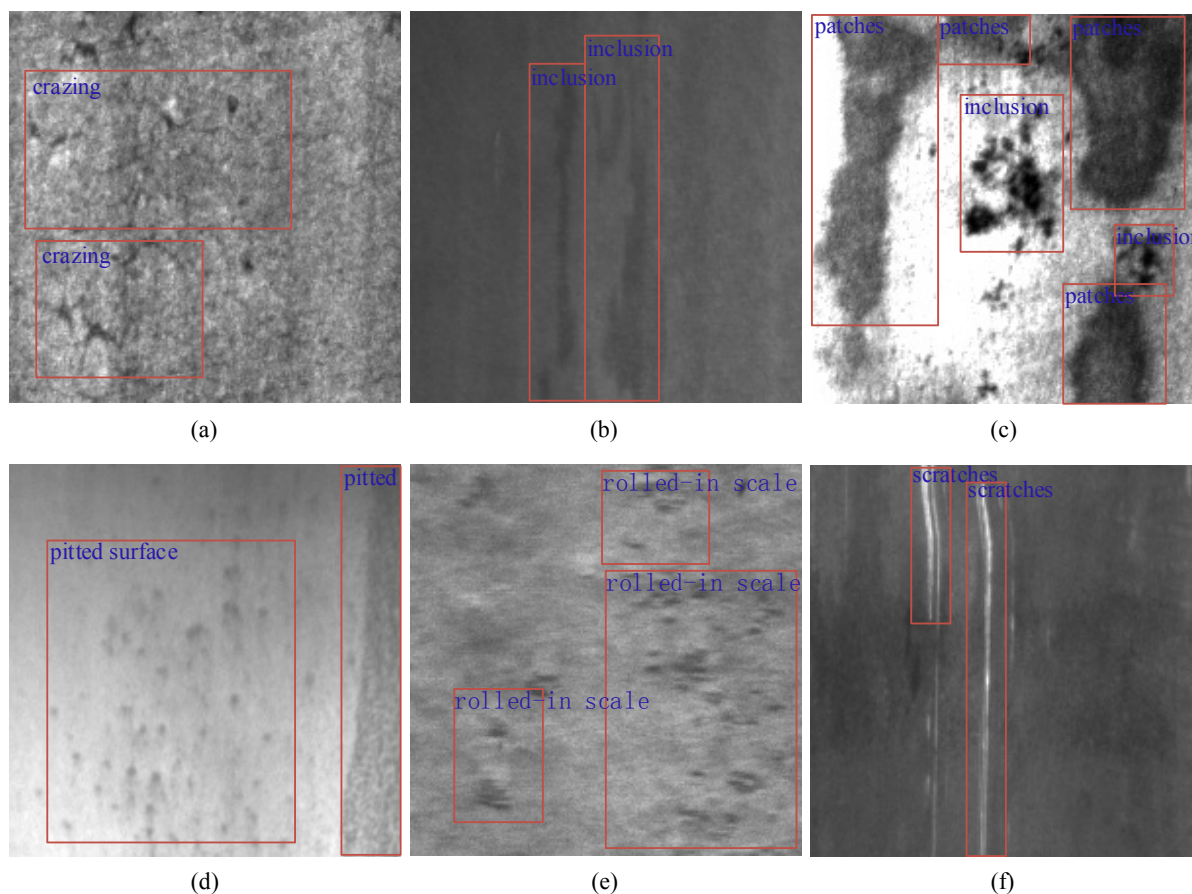
| 方法                      | mAP/% | Cr/%  | In/%  | Pa/%  | PS/%  | RS/%  | Sc/%  | T/ms  |
|-------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| YOLOX-S                 | 94.23 | 86.78 | 96.74 | 96.17 | 99.68 | 88.98 | 97.34 | 13.79 |
| +Focal_Loss             | 89.85 | 82.98 | 89.94 | 96.90 | 93.69 | 83.48 | 92.12 | 15.01 |
| +EIOU_Loss              | 95.46 | 89.90 | 95.15 | 96.85 | 97.61 | 95.69 | 97.55 | 14.18 |
| +CBAM                   | 94.54 | 88.23 | 97.55 | 96.15 | 96.55 | 93.58 | 95.17 | 14.80 |
| +Focal_Loss + EIOU_Loss | 89.87 | 83.09 | 89.99 | 96.84 | 93.88 | 83.35 | 92.08 | 19.76 |
| +CBAM + EIOU_Loss       | 96.14 | 94.12 | 97.71 | 96.48 | 96.77 | 94.40 | 97.35 | 14.24 |
| +CBAM + GIOU_Loss       | 94.84 | 88.29 | 94.01 | 97.53 | 98.69 | 93.66 | 96.86 | 14.85 |

以改进实验中 mAP 指标最高的模型，即采用 *EIOU\_Loss* 加注意力价值改进的 YOLOX-S 检测模型，多种类别金属表面缺陷的平均检测准确率达到 96% 以上，其中单张检测时间为 14.24 ms，实验中 T 用于评估图像检测模型的识别速度。其中 T 表示的是单张图像推理时间，单位为毫秒，推理时间越小，图像检测模型识别速度越快。实际中，考虑金属工件在流水线过程中的图像动态获取时间一般低于 0.3 s，再



**Figure 4.** Diagram of prediction result samples; (a) crazing; (b) inclusion; (c) patches; (d) pitted surface; (e) rolled-in scale; (f) scratches

**图 4.** 预测结果样本图; (a) 裂纹; (b) 夹杂; (c) 斑块; (d) 麻点; (e) 氧化铁皮压入; (f) 划痕



**Figure 5.** Diagram of real label samples; (a) crazing; (b) inclusion; (c) patches; (d) pitted surface; (e) rolled-in scale; (f) scratches

**图 5.** 真实标签样本图; (a) 裂纹; (b) 夹杂; (c) 斑块; (d) 麻点; (e) 氧化铁皮压入; (f) 划痕

加上单张推理时间, 为检测每个金属工件的总时间。由式(5)计算可知检测速率约为每分钟 200 个以上。

$$\text{Number} = 60/T_{total} \quad (5)$$

式中,  $T_{total}$  为检测每个金属工件的总时间。

图像预测样本如图 4 所示, 包括六类表面缺陷, 其对应的缺陷的真实标签图像如图 5 所示。对照预测结果样本与真实标签样本可以看出, 在预测框的类别检测方面未出现错误, 整体识别效果较好。在图 4 中, 目标检测模型成功识别了斑块和夹杂缺陷, 存在的识别误差主要来自于缺陷的位置判断偏差与缺陷检出偏差上, 图像算法模型的检测效果达到预期要求。

#### 4. 结论

本文提出了一种改进的 YOLOX-S 网络的金属表面缺陷图像检测方法, 首先引入了金属表面缺陷目标检测数据集 NEU-DET, 分析了新一代的 YOLO 模型 YOLOX 的网络结构与算法特点。针对 YOLOX-S 模型在金属表面缺陷数据集上检测存在的问题, 提出了添加注意力机制模块与改进损失函数的优化措施。在 NEU-DET 数据集上通过空间和通道多维度提取缺陷特征, 以及  $EIOU\_Loss$  损失函数利用预测框和真实框的宽高差异值取代纵横比, 并添加  $Focal\_Loss$  解决预测框回归中的样本不平衡问题。实验结果表明, 在不影响其它缺陷种类的检测情况下, 重点提升了裂纹和氧化铁皮压入的缺陷分类准确率和定位精度, 其模型整体 mAP 值达到了 96.14%, FPS 仅为 14.24 ms, 即金属表面多种类别缺陷平均检测准确率 96%



以上以及检测金属工件速率每分钟 200 个以上。同时由于实验中的缺陷数据集存在分布不均的问题, 后续可以通过扩展更加丰富的自制数据集, 提升网络结构的稳定性, 进一步提高模型的泛化能力。

## 基金项目

四川省重点研发计划项目: 高速、高精度机器仿生视觉外观检测方法研究与系统研制(2020YFG0113)。

## 参考文献

- [1] 李燕. 铁基粉末冶金零件常见生产缺陷的金相分析[J]. 山西冶金, 2017, 40(5): 87-89.
- [2] 方德伟. 超声波检测技术在无缝钢管质量缺陷检测中的应用研究[J]. 科技视界, 2019(12): 37-38.
- [3] 周心悦. 基于机器视觉的金属片表面缺陷实时检测与分类研究[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2021: 11-23.
- [4] 李少波, 杨静, 王铮, 等. 缺陷检测技术的发展与应用研究综述[J]. 自动化学报, 2020, 46(11): 2319-2336.
- [5] 李维刚, 叶欣, 赵云涛, 等. 基于改进 YOLOv3 算法的带钢表面缺陷检测[J]. 电子学报, 2020, 48(7): 1284-1292.
- [6] 孙连山, 魏婧雪, 朱登明, 等. 基于 AM-YOLOv3 模型的铝型材表面缺陷检测算法[J]. 激光与光电子学进展, 2021, 58(24): 360-370.
- [7] Tian, R. and Jia, M. (2022) DCC-CenterNet: A Rapid Detection Method for Steel Surface Defects. *Measurement*, **187**, Article ID: 110211. <https://doi.org/10.1016/j.measurement.2021.110211>
- [8] He, Y., Song, K., Meng, Q., *et al.* (2020) An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features. *IEEE Transactions on Instrumentation and Measurement*, **69**, 1493-1504. <https://doi.org/10.1109/TIM.2019.2915404>
- [9] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. CoRR. arXiv:1804.02767.
- [10] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [11] Lin, T.Y., Goyal, P., Girshick, R., *et al.* (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2999-3007. <https://doi.org/10.1109/ICCV.2017.324>
- [12] Zhang, Y.F., Ren, W., Zhang, Z., *et al.* (2021) Focal and Efficient IOU Loss for Accurate Bounding Box Regression. CoRR. arXiv:2101.08158.