

文章编号 1004-924X(2023)03-0404-13

## 融合注意力机制的金属锅圆柱表面缺陷检测

乔健<sup>1,2</sup>, 陈能达<sup>1</sup>, 伍雁雄<sup>3</sup>, 吴阳<sup>1</sup>, 杨景卫<sup>1\*</sup>

(1. 佛山科学技术学院 机电工程与自动化学院, 广东 佛山 528000;

2. 季华实验室, 广东 佛山 528200;

3. 佛山科学技术学院 物理与光电工程学院, 广东 佛山 528000;

**摘要:**为实现高亮反射金属圆柱形锅的自动快速检测及分拣, 破解目前金属锅表面缺陷检测速度慢、效率低的技术难题, 在YOLOX网络基础上引入双向特征融合网络, 提出基于注意力机制的轻量化特征融合网络模型, 实现计算模型的轻量化设计; 同时, 通过注意力机制模块对特征信息进行通道与空间的学习, 有效缓解多尺度特征的语义鸿沟问题, 提高了模型的检测精度; 考虑网络对难易分类样本学习权重分配不平衡, 设计基于衰减因子的分类损失函数; 利用金属锅圆柱表面缺陷数据集完成了特征融合网络对比实验、分类损失函数对比实验和注意力机制模块位置消融实验。实验结果表明, 融合注意力机制模型可有效识别6种不同形态的缺陷, 测试集的平均检测精度 $mAP_{0.5}$ 达到90.92%, 检测帧率达到30.84 frame/s, 实现了金属锅圆柱表面缺陷的高精度快速识别与定位。

**关键词:** 机器视觉; 特征金字塔; 注意力机制; 金属表面缺陷

中图分类号: TP391.41 文献标识码: A doi: 10.37188/OPE.20233103.0404

## Defect detection of cylindrical surface of metal pot combining attention mechanism

QIAO Jian<sup>1,2</sup>, CHEN Nengda<sup>1</sup>, WU Yanxiong<sup>3</sup>, WU Yang<sup>1</sup>, YANG Jingwei<sup>1\*</sup>

(1. School of Electrical and Mechanical Engineering and Automation, Foshan University,  
Foshan 528000, China;

2. Ji Hua Laboratory, Foshan 528200, China;

3. School of Physics and Optoelectronic Engineering, Foshan University, Foshan 528000, China)

\*Corresponding author, E-mail: mejwyang@fosu.edu.cn

**Abstract:** To achieve the automatic and rapid detection and sorting of high-brightness reflection metal cylindrical pots, as well as break through the technical problems of slow speed and low efficiency of metal pot surface defect detection, a bi-directional feature pyramid network (BiFPN) was introduced in this study based on the YOLOX network. In addition, a lightweight feature fusion network model was devised on the basis of the attention mechanism, and the lightweight design of the computing model was realized. Meanwhile, the attention mechanism module was employed to learn the channel and space of feature information, effectively alleviating the semantic gap of multi-scale features and improving the detection precision of the model. Considering the unbalanced distribution of the learning weight of the network for diffi-

收稿日期: 2022-06-01; 修订日期: 2022-07-08.

基金项目: 广东省科技计划资助项目 (No. X220391TH220)

cult and easy classification samples, the classification loss function regarding the attenuation factor was determined. Comparisons of the feature fusion network, classification loss function, and attention mechanism module position ablation were conducted using the metal pot cylindrical surface defect dataset. The experimental results show that the fusion attention mechanism model can effectively identify six types of defects with different shapes, the average detection precision  $mAP_{0.5}$  of the test set realized 90.92%, and the detection frame rate was 30.84 FPS. Thus, cylindrical surface defects of metal pots can be identified and located, rapidly as well as with high precision, by using the proposed model.

**Key words:** machine vision; feature pyramid; attention mechanism; metal surface defects

## 1 引言

随着人们生活质量的不断提高,对日常生活用品的外观及品质提出了更高要求,所以以外观漂亮、坚固耐用及清洁便捷的不锈钢餐厨具备受人们青睐。其中,以圆柱锅为代表的锅类不锈钢产品占整个餐厨具一半以上的份额。圆柱金属锅在抛光加工过程中由于随机可变因素导致擦伤、纹理不均、色差、变形等表面缺陷,产品质量和外观得不到有效保障。同时,圆柱金属锅作为典型多品种、小批量产品,规格种类多、批量小、外观尺寸较大,在产品表面质量检测过程中一直以人工目视检测为主,漏检、误检现象时有发生,效率低、速度慢的目视检测方法严重制约着餐厨具生产加工行业数字化转型的发展进程。

金属圆柱形工件的表面缺陷检测是光学自动检测中的难点,国内外学者开展了大量机器视觉检测装置和算法的研究。其中,圆柱形工件的图像采集装置基本采用旋转机构搭配工业线阵相机<sup>[1]</sup>或工业面阵相机<sup>[2]</sup>的设计。近年来,随着计算机视觉技术的快速发展,人们提出了一系列基于深度学习的金属表面缺陷检测方法。Lin等<sup>[3]</sup>提出基于注意力机制和特征图映射的鲁棒检测方法,解决了X射线检测中铸件缺陷的误检和漏检问题。Sun等<sup>[4]</sup>在Faster R-CNN主干网络中引入RPN结构,在300毫秒内实现了对单张轮毂缺陷图像的快速检测。由于单阶段目标检测算法中,YOLO系列将目标分类和定位问题转化为回归问题,从首次新增特征融合层的YOLOv3发布为起点,特征融合层成为后续发布的YOLO算法必不可少的组成部分。王宸等<sup>[5]</sup>通过替换YOLOv3主干网络的激活函数以及对损失函数进行修改,验证集的mAP值为98.94%,检测速

度达到76.59 frame/s,基本满足轮毂焊缝的实时在线检测需求。针对金属表面检测中微小缺陷易漏检的状况,程婧怡等<sup>[6]</sup>利用YOLOv3主干网络Darknet-53结构中的浅层特征图,提取出更多小目标信息,降低了小缺陷的漏检概率,但无法实现实时的快速检测。2021年,旷视科技的Ge<sup>[7]</sup>公布了YOLOX算法,在动态正样本数量分配以及预测框回归方面展开研究,实现了检测精度的提高,但计算机硬件算力的要求也更高。

在深度学习模型轻量化的研究中,通常采用替换主干特征提取网络<sup>[8]</sup>、设计轻量网络基础模块<sup>[9]</sup>、替换轻量化卷积模块<sup>[10]</sup>、设计运算量较小的卷积运算方式<sup>[11]</sup>以及通过缩小模型宽度和深度的方式来达到轻量化模型的目的。但上述方法均可能导致模型检测精度的损失,因此研究人员又设法在轻量化模型的基础上融合新的模块,以提高精度。为降低深度学习模型在工业生产现场中的应用难度,本文结合BiFPN<sup>[12]</sup>跨尺度特征融合思想,通过替换YOLOX特征融合层(Neck)进行模型轻量化,在Neck层融合注意力机制模块,提高模型检测精度。针对目标检测算法在训练过程中正负样本比例严重失衡的问题,提出了基于衰减因子的分类损失函数(Sinusoidal Attenuation focal Loss, SFL),加强难分类正样本的学习,减少易分类负样本对梯度反向传播的影响。结合上述改进策略开展实验,以验证基于融合注意力机制的跨尺度特征融合网络的有效性,实现了金属锅圆柱表面缺陷的高精度快速检测。

## 2 融合注意力机制模型

基于深度学习的目标检测算法结构通常分为主干网络、特征融合层、预测层以及损失函数

部分。本文以 YOLOX 为基础框架,提出融合注意力机制的跨尺度特征融合网络(Bi-directional YOLOX, BiYOLOX),改进的网络结构如图 1 所示。本文主要通过设计基于注意力机制的特征融合方式,对 BiFPN 的各分支特征进行注意力学习,改善对包含目标区域的关注度,提高模型对目标的检测精度。通过引入衰减因子来设计分类损失函数,重新分配正负样本学习权重,由 SFL 完成正负样本与目标类别的分类,由 CIoU Loss<sup>[13]</sup>完成目标位置的回归。

如图 1 所示,维度大小为  $960 \times 960 \times 3$  的待检测图像首先经过 Focus 结构将原始 RGB 图像的通道数从 3 扩展为 12,并将原来宽高为  $w, h$  的特征缩小为  $w/2, h/2$ ,对输入图像进行  $2 \times 2$  的网格化采样,再输入 CSPDarknet 主干网络进行特征提取;用于特征融合的 Neck 层为在分支引入

注意力机制的 BiFPN 网络;网络头部检测器 YoloHead 对 Neck 层得到的 3 个有效特征层结果进行堆叠,获得包含预测框数量、类别、置信度与坐标位置的特征信息;网络训练过程中,通过计算模型 Loss 获得预测框与真实框的偏差,并利用模型 Loss 更新网络参数。图中,Cls 部分为包含物体的种类判断特征点,Reg 部分为回归参数判断的特征点,Obj 部分为包含物体判断的特征点。其中,Cls 和 Obj 部分通过分类损失函数进行判断,Reg 部分则通过回归损失函数进行判断。模型 Loss 为:

$$Loss = a_1 Loss_{cls} + a_2 Loss_{obj} + a_3 Loss_{IoU}, \quad (1)$$

其中: $Loss_{cls}$  为每一个特征点所包含物体种类损失, $Loss_{obj}$  为每一个特征点是否包含物体损失, $Loss_{IoU}$  为预测框回归损失, $a_i$  为各损失权重系数。

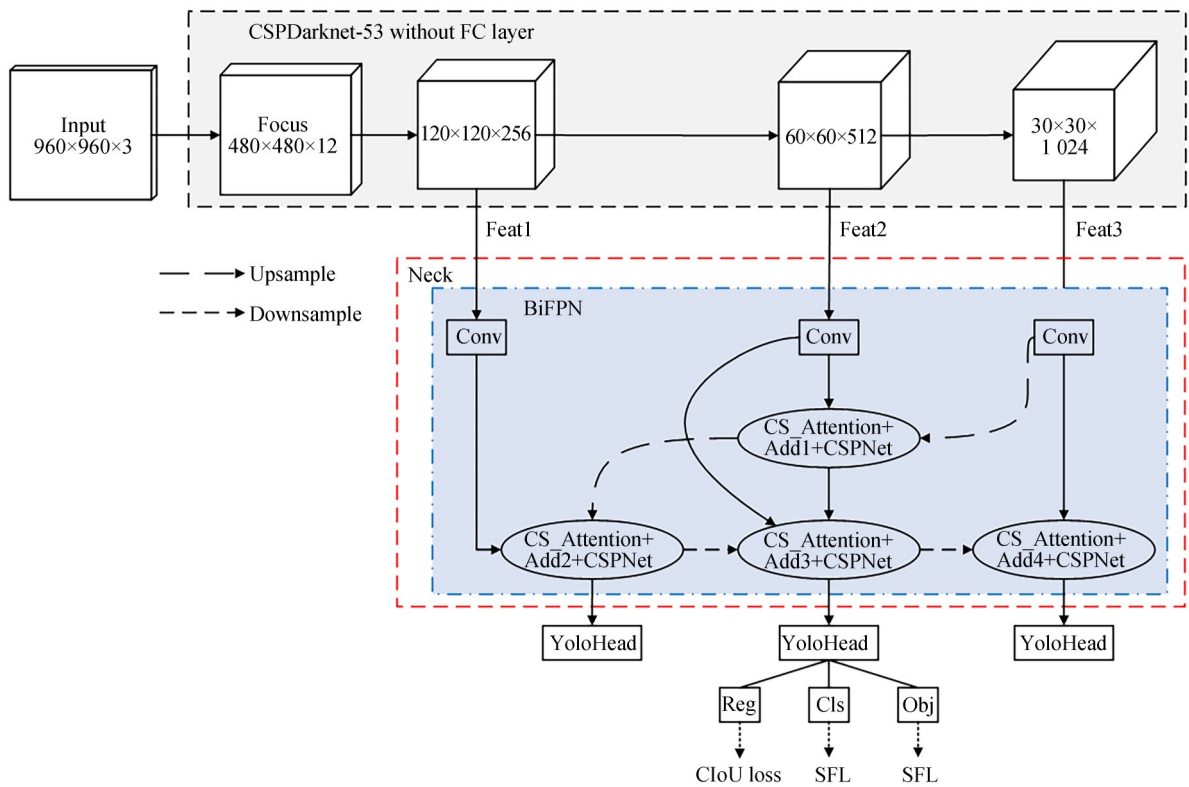


图 1 BiYOLOX 网络结构

Fig. 1 Structure of BiYOLOX network

## 2.1 跨尺度特征融合网络

BiFPN 是通过改进 FPN 来实现多尺度的特征融合网络,并对主干网络提取的三个特征层 Feat1, Feat2 和 Feat3 进行通道数维度的统一,然

后自顶向下和自底向上的双向特征融合。最后,把融合的特征传送到网络头部检测器 YoloHead。BiFPN 采用快速归一化(Fast Normalized Fusion, FNF)计算方式对每个输入添加获得额

外权重,为不同特征层信息提供单独的学习权重分配,从而缓解由高低分辨率特征信息融合带来的语义鸿沟<sup>[14]</sup>,即有:

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i, \quad (2)$$

其中: $w_i$ 为可学习的权重, $I_i$ 为输入特征; $\epsilon$ 取值<sup>[12]</sup>为 0.000 1,以确保分母不为零并获得较好的训练效果;为保证获取的权重大于 0,更新权重前采用 ReLU 激活函数进行映射,使得小于 0 的权值以 0 输出,大于或等于 0 的权值按原值输出。

相对于 YOLOX 使用的路径聚合网络(Path Aggregation Network, PANet)<sup>[15]</sup>,BiFPN 在使用规则和高效的连接下,完成信息自上而下的流动,通过削减只有一个入度的节点来去除冗余计算,达到轻量化网络的效果。同时,为避免特征信息丢失,使用 CSPNet<sup>[16]</sup>替换深度来分离卷积进行特征提取。

### 2.2 注意力机制的特征融合

为在轻量化改进模型中尽量提高模型检测精度,本文在模型 Neck 层引入注意力机制。深度学习模型的大小通常由主干部分决定,但受制于计算机硬件的算力,需对神经网络深度、宽度和输入分辨率进行平衡,才可使模型表现出更好的性能<sup>[17]</sup>。文献[18]针对目标背景复杂的问题,

在 SSD 算法中增加通道注意力模块,将算法注意力集中在目标区域的通道,增大目标区域通道的特征响应,减小背景干扰,提高了模型的检测精度。文献[19]为从图像中提取更有效的特征用于后续特征点的提取和描述,将通道注意力模块集成到主干提取网络中,在不增加模型复杂度前提下获得了优良的特征。文献[20]在 SENet 基础上设计了更轻量的通道注意力模块并融合到主干网络中的 CSP 模块,该方法增强了主干网络对复杂背景下特征的筛选能力。上述研究表明,网络融合注意力机制模块同样能够影响深度学习模型的性能。根据人类在识别物体时产生的注意力集中现象,注意力机制描述如下:

$$Attention(x) = f(g(x), x), \quad (3)$$

其中: $g(x)$ 表示产生注意力,对应各类注意力机制算法在关注判别性区域的过程, $f(g(x), x)$ 表示使用注意力  $g(x)$  对输入  $x$  进行处理。

为使网络更好地聚焦在感兴趣区域,本文提出在特征信息进行融合前,将来自不同特征层的信息通道拼接,接着进行特征块的注意力学习,使网络学习到每个特征通道和空间的重要性,并重新进行加权,经 Split 操作恢复为初始维度大小的新特征信息。图 2 为基于注意力机制的多尺度特征学习方式示意图。

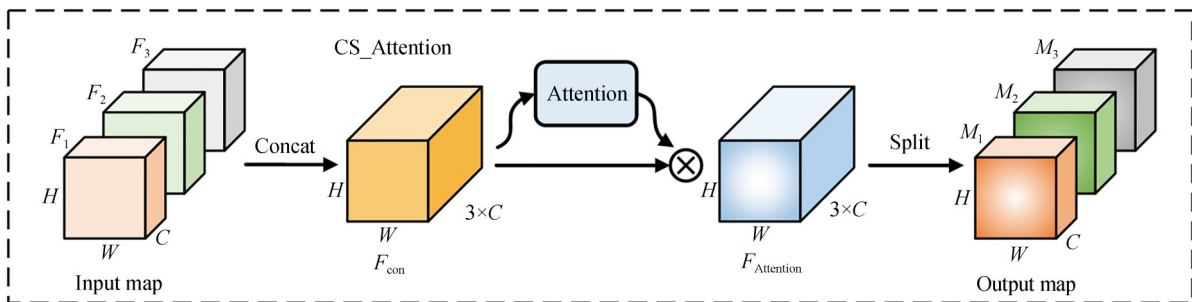


图 2 基于注意力机制的多尺度特征学习

Fig. 2 Multiscale feature learning based on attention mechanism

基于注意力机制的多尺度特征学习计算公式如下:

$$F_{con} = \text{Concat}(F_1, F_2, F_3), \quad (4)$$

$$(M_1, M_2, M_3) = \text{Split}(F_{con} \otimes \text{Attention}(F_{con})), \quad (5)$$

其中: $F_i$ 为输入特征图, $M_i$ 为输出特征图,维度为

$(H, W, C)$ ,  $\otimes$ 为基于元素的乘法运算,Concat为拼接操作,Split为切片操作,Attention为注意力机制模块。

通道注意力模块(Channel Attention Module, CAM)、空间注意力模块(Spatial Attention Module, SAM)、模块级联构成的注意力模块(Convolutional Block Attention Module, CBAM)

具有优越的性能并得到了广泛应用<sup>[21]</sup>,但其计算量较大。为降低整体注意力机制模块的计算量,增大输出特征感受野,本文利用3个卷积核大小为 $3\times 3$ 的卷积代替SAM中 $7\times 7$ 卷积操作,把拼接后通道数为2的特征信息降维为单通道。但由于小卷积核的普通卷积操作对提升特征信息感受野的能力有限,引入空洞卷积替换最后一个 $3\times 3$ 卷积操作,可增大特征信息感受野。改进的SAM在保留网络对特征空间学习能力的同时,增大特征图的感受野,使输出特征尺度更加丰富,对应结构如图3所示。改进后的CBAM命名为空洞卷积注意力模块(Dilated Convolutional Block Attention Module, Di\_CBAM)。

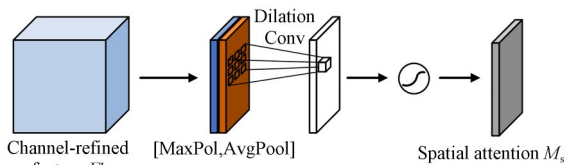


图3 基于空洞卷积的SAM模块

Fig. 3 SAM module based on dilated convolution

基于 $3\times 3$ 卷积的SAM计算公式如下:

$$M_s(F) = \sigma\left(f_3^{3\times 3}\left([\text{AvgPool}(F)], [\text{MaxPool}(F)]\right)\right), \quad (6)$$

其中: $\sigma$ 为Sigmoid激活操作, $f_3^{3\times 3}$ 表示3个 $3\times 3$ 的卷积操作, $\text{AvgPool}(F)$ 和 $\text{MaxPool}(F)$ 分别表示对特征图 $F$ 进行平均池化操作和最大池化操作。

### 2.3 分类损失函数

针对单阶段目标检测算法训练过程中正负样本比例严重失衡的问题,Lin等<sup>[22]</sup>提出了由标准交叉熵损失函数(Cross Entropy Loss, CEL)改进的聚焦损失函数(Focal Loss, FL)。在CEL中引入权重因子 $\alpha_i$ ,即可解决正负样本不平衡的问题。大多数负样本属于易分类样本,正样本通常属于难分类样本。FL的引入可控制难分类样本学习权重的调制因子,降低了大量易分类负样本在训练中所占的权重。具体表达式如下:

$$p_i = \begin{cases} p & , y = 1 \\ 1 - p & , y = 0 \end{cases}, \quad (7)$$

$$\alpha_i = \begin{cases} \alpha & , y = 1 \\ 1 - \alpha & , y = 0 \end{cases}, \quad (8)$$

$$\text{CEL}(p_i) = -\log(p_i), \quad (9)$$

$$\text{FL}(p_i) = -\alpha_i(1 - p_i)^\gamma \log(p_i), \quad (10)$$

其中: $y$ 表示预测样本,当 $y=0$ 时,预测样本为负样本,当 $y=1$ 时,预测样本为正样本; $p$ 为预测样本的预测结果对应真实类别的概率; $\alpha \in [0, 1]$ ,是用于平衡正负样本权重的系数,参数 $\gamma \in [0, \infty)$ 用于控制难分类和易分类样本的权重, $(1 - p_i)^\gamma$ 为调制因子。

通常预测样本为真的概率越大,易分类样本的学习权重会越小,且衰减较缓慢,而难分类样本的学习权重通常较大。因此,FL会导致网络过多关注难分类样本的学习,使得模型Loss收敛缓慢,优化结果偏离预期。针对难分类正样本的学习,减少易分类负样本对梯度反向传播的影响,根据正弦函数的曲线变化趋势,在FL中引入 $(0, 1]$ 区间内具有非线性衰减趋势的因子 $s$ 。把改进的损失函数命名为正弦衰减聚焦损失函数(Sinusoidal attenuation Focal Loss, SFL),其表达式为:

$$s(p_i) = \frac{\sin(\theta \cdot p_i)}{p_i}, \quad (11)$$

$$\text{SFL}(p_i) = -\alpha_i(1 - p_i)^\gamma s \log(p_i), \quad (12)$$

其中: $\alpha_i$ , $p_i$ 和 $\gamma$ 的定义与FL中的参数相同, $s$ 为正弦衰减因子,且 $\theta \in (0, \pi]$ 。图4展示了正弦衰减因子 $s$ 随 $p_i$ 的变化情况,图5展示了本文提到的分类损失函数曲线示例。

由图4可知,随着 $p_i$ 的增大,正弦衰减因子 $s$ 呈下降趋势,并随着参数 $\theta$ 的增大, $s$ 的变化范围

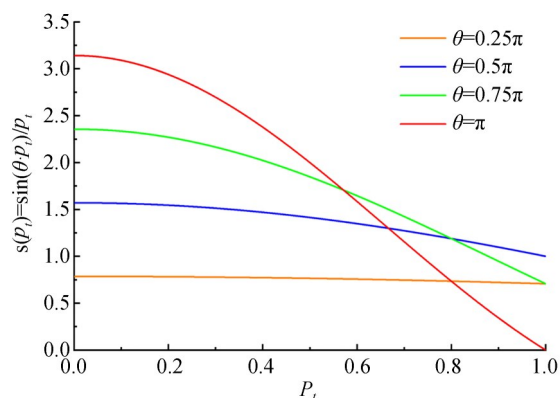


图4 正弦衰减因子变化曲线

Fig. 4 Variation curves of sinusoidal attenuation factor

逐渐变大,模型会增大对难分类样本的学习权重分配。

由图 5 可知,SFL 通过引入具有非线性衰减趋势的因子  $s$ ,可使模型在侧重学习难分类样本的同时,增大易分类样本的学习权重,并随着样本预测概率的增大,易分类样本学习权重逐渐减小。因此,通过引入参数  $\theta$  来改变 FL 调制因子,可加快模型损失的收敛速度,提高整体学习效果。

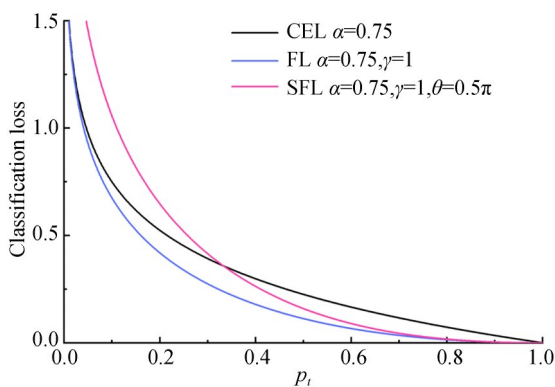


图 5 分类损失函数曲线示例

Fig. 5 Example of classification loss function curves

### 3 实 验

#### 3.1 实验数据集与训练策略

考虑金属锅具生产环境中存在光照变化,以及车间灰尘对镜头的污染等影响,分别采集不同清晰度、亮度以及干扰光源下的图像,图像采集装置如图 6 所示。

根据物理属性和后续修复操作,将高亮反射

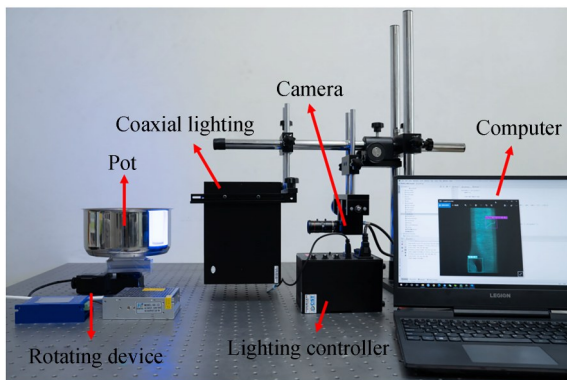


图 6 图像采集与检测系统

Fig. 6 Image acquisition and detection system

金属圆柱锅产品常见的柱面缺陷定义为凸起 (Bulges, Bu)、指纹油污 (Finger print, FP)、黑斑块 (Patches, Pa)、凹坑 (Pit, Pi)、划痕 (Scratches, Sc) 和擦伤 (Surface scratch, SS)。CMOS 工业面阵相机的分辨率为  $2\ 048 \times 1\ 536$  像素,实验共采集有效图像样本 1 000 幅,保证数据集的丰富程度<sup>[23]</sup>。为降低深度学习模型的训练时间,去除图像中的冗余信息,提高算法在生产现场中的实时性,裁剪部分非光斑与无缺陷区域,统一图像尺寸为  $960 \times 960$  像素。同时,定义标记框面积  $S < 32^2$  个像素为小型目标,  $32^2 \leq S \leq 96^2$  为中型目标,  $S > 96^2$  为大型目标。

使用 Vott 软件对图像进行处理,数据集以 VOC2007 格式进行保存,各类别缺陷标注样例如图 7 所示(彩图见期刊电子版)。其中,红色标注框为凸起、紫色标注框为指纹油污、绿色标注框为黑斑块、青色标注框为凹坑、蓝色标注框为划痕、黄色标注框为擦伤。

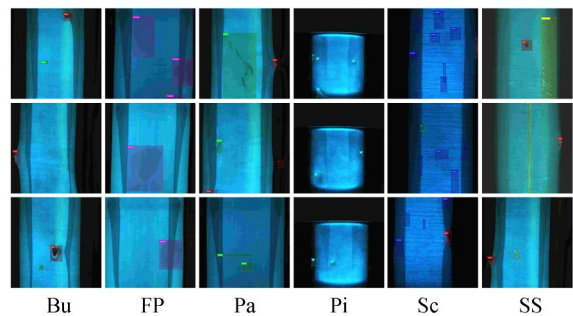


图 7 金属锅圆柱表面各类缺陷样本

Fig. 7 Samples of various defects on cylindrical surface of metal pot

随机抽取训练集和验证集,它们与测试集图像的比例为 9:1,其中训练集与验证集抽取比例同为 9:1,即训练集 810 张,验证集 90 张,测试集 100 张。数据集命名为金属锅圆柱表面缺陷数据集,图 8 展示了该数据集标记框大中小目标数量的分布情况。结合图 7 展示的样本例可知,数据具有类别不平衡、同一图像存在多个实例、目标尺寸各异特性。

实验配置为: Intel(R) Core(TM) i5-9300H 处理器,运行内存 16 GB,图形处理单元 NVIDIA GeForce GTX 1660 Ti,显存 6 GB,安装 CUDA 11.0, cuDNN 8.0.5,深度学习框架 Py-

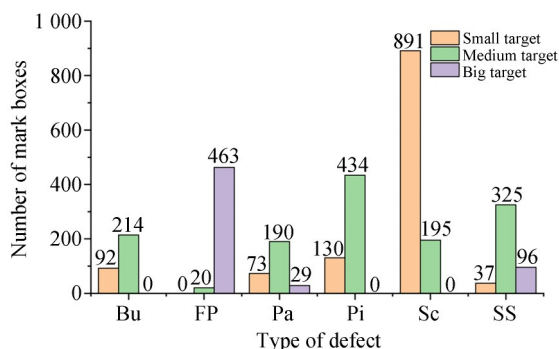


图8 各类标签的目标数量分布

Fig. 8 Target quantity distribution of various labels

torch1.4。模型输入图像尺寸为 $960 \times 960$ 像素,模型超参数设置如表1所示。

表1 模型超参数设置

Tab. 1 Model super parameter setting

Super parameter	Value
Batch size	4
Cosine scheduler	0.000 1
Weight decay	0.000 01
NMS	0.05
Epoch	300

正式训练前,在金属锅圆柱表面缺陷数据集上进行100个Epoch预训练,使用预训练得到的权重文件进行正式训练。训练过程中,通过AdamW对训练过程进行优化,利用Cosine scheduler学习率调整策略进行训练,避免网络陷入局部最优的情况,并通过Mosaic算法对网络输入数据进行增强。

### 3.2 评价指标

为较客观地评价所设计模型的检测效果,实验通过召回率(Recall)、平均精度(Average Precision, AP)以及均值平均精度(mean Average Precision, mAP)来评估模型的性能。AP用于评价模型对某一类别目标的检测效果,AP值越高,模型检测效果越佳。其计算公式如下:

$$AP = \int_0^1 P(R) dR. \quad (13)$$

对于目标检测来说,当预测框与真实框的矩形框交并比IoU值大于某个阈值时,才会将该预测框认定为真样例,因此IoU阈值的选择对于目标检测的效果极为重要。mAP是衡量目标检测

算法整体性能的最为全面的评价指标,其公式如下:

$$mAP = \frac{\sum_{i=1}^n AP_{(i)}}{n}, \quad (14)$$

其中: $n$ 为类别数, $AP_{(i)}$ 表示第 $i$ 个类别的AP值。

经测试,模型的精确率曲线拐点出现在IoU=0.5附近,为观察模型在高IoU阈值设定下的性能表现,实验环节将IoU阈值设为0.5和0.75,记均值平均精度为 $mAP_{0.5}$ 和 $mAP_{0.75}$ 。

对于模型计算效率评价,本文采用模型参数量(Params)、浮点运算数(FLOPs)和检测帧率(FPS)来衡量。其中,FPS为计算模型对单张图像的平均检测时间,即:

$$FPS = \frac{1}{t}, \quad (15)$$

其中: $t$ 为单张 $960 \times 960$ 像素图像的检测时间。

### 3.3 实验结果与分析

本文进行了特征融合网络对比实验、损失函数对比实验、注意力机制模块位置消融实验以及与其他轻量化改进模型的性能对比实验。根据已搭建的实验条件,使用金属锅圆柱表面缺陷数据集集中的测试集进行性能测试,且实验均采用size为 $s$ 的模型,即YOLOX $_s$ 和BiYOLOX $_s$ 。

PANet和BiFPN特征融合网络对比实验中,模型的性能表现如表2所示。本组实验采用IoU Loss<sup>[24]</sup>作为回归损失函数,以CEL作为分类损失函数。

表2 特征融合网络性能对比

Tab. 2 Performance comparison of feature fusion network

Neck	Params /M	FLOPs /G	FPS /(frame·s <sup>-1</sup> )	mAP <sub>0.5</sub> /%
PANet	9.0	29.977	28.00	86.52
BiFPN	6.4	25.924	32.20	90.06

模型Neck层融合BiFPN后,性能得到明显提升。相比于原YOLOX网络使用的PANet结构,Params减少2.6M,FLOPs下降4.053G,网络复杂度明显降低。随着Params的减少,FPS得到4.2 frame/s的提升,单张 $960 \times 960$ 像素图片的检测时间从YOLOX的35.7 ms降低至31.1 ms,并且 $mAP_{0.5}$ 提升3.54%,达到90.06%。

损失函数对比实验中,为了比较IoU计算方式对BiYOLOX算法检测效果的影响,分别使用IoU Loss, GIoU Loss<sup>[25]</sup>, DIoU Loss<sup>[26]</sup>和CIoU Loss作为算法的回归损失函数,测试结果如表3所示。其中,模型分类损失函数为FL,且 $\alpha=0.75, \gamma=2$ 。

表3 回归损失函数的检测效果对比

Tab. 3 Comparison of detection effect of regression loss functions

Reg_Loss	FPS/(frame·s <sup>-1</sup> )	mAP <sub>0.5</sub> /%	mAP <sub>0.75</sub> /%
IoU	32.10	85.20	36.20
GIoU	32.00	86.11	38.47
DIoU	32.00	88.94	35.29
CIoU	32.00	89.30	39.63

从表3可以看出,使用GIoU Loss, DIoU Loss和CIoU Loss作为BiYOLOX的回归损失函数可进一步提升模型的检测精度,并且保持较高的检测速度。相比于原始的IoU Loss,采用GIoU Loss的mAP<sub>0.5</sub>和mAP<sub>0.75</sub>指标分别提升了0.91%和2.27%;采用DIoU Loss的mAP<sub>0.5</sub>提升3.74%,mAP<sub>0.75</sub>下降0.91%。本组实验中,采用CIoU Loss作为回归损失函数的模型在检测精度方面得到了最大提升,mAP<sub>0.5</sub>和mAP<sub>0.75</sub>指标分别提升了4.10%和3.43%。

为验证设计的SFL对BiYOLOX算法检测效果的影响,在CIoU Loss作为回归损失函数的情况下进行分类损失函数对比实验,采用4分法划分 $\theta$ 的取值范围,实验结果如表4所示。

表4 分类损失函数的检测效果对比

Tab. 4 Comparison of detection effect of classification loss functions

Obj_Loss	$\alpha$	$\gamma$	$\theta$	mAP <sub>0.5</sub> /%	mAP <sub>0.75</sub> /%
FL	0.75	1	—	90.02	34.99
FL	0.75	2	—	89.30	33.63
SFL	0.75	1	0.25 $\pi$	90.14	38.95
SFL	0.75	1	0.5 $\pi$	91.37	36.56
SFL	0.75	1	0.75 $\pi$	90.91	35.31
SFL	0.75	1	$\pi$	90.58	35.01

由表4可知,所设计的SFL在参数 $\theta=0.5\pi$ 的设置下,模型获得的mAP<sub>0.5</sub>最高,相比于相同参数设置下的FL提升了1.35%。当参数 $\theta \geq 0.75\pi$ 时,模型的mAP<sub>0.5</sub>与mAP<sub>0.75</sub>均会下降。由此可知,在FL中引入正弦衰减因子 $s$ 能够有效提高模型的检测精度,但当正弦衰减因子 $s$ 的变化幅度较大时,模型对难分类样本的关注度较高,模型的检测精度会下降。

在分类损失函数对比实验中,根据表4中不同 $\theta$ 设置下的SFL分组实验,对金属锅圆柱表面缺陷召回率数据进行统计,结果如图9所示。

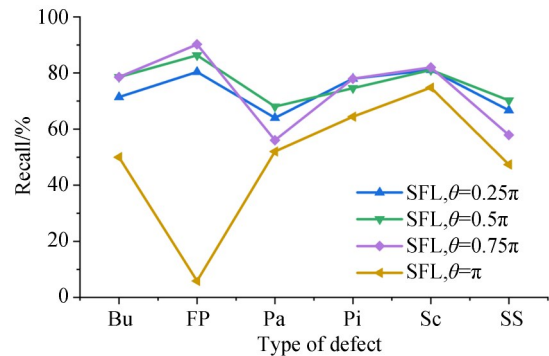


图9 不同 $\theta$ 设置下各类别缺陷的召回率统计

Fig. 9 Statistics of defect recall rate of each category under different  $\theta$  settings

由图9可知,当SFL的参数 $\theta=0.25\pi$ 时,各类别缺陷的召回率在60%~80%内;当参数 $\theta=0.5\pi$ 时,各类别缺陷的召回率在70%~90%内,表现最为均衡;当参数 $\theta=0.75\pi$ 时,黑斑块和擦伤缺陷的召回率低于60%,各类别缺陷的召回率在50%~95%内;当参数 $\theta=\pi$ 时,相比于同参数设置下的FL,各类别缺陷召回率均有不同程度的下降,指纹油污缺陷的召回率下降最严重,已低于10%。结合图8中的数据集标记框大中小目标数量的分布情况,指纹、黑斑块与擦伤缺陷均有大目标存在,而指纹缺陷在数据集中主要为大目标,因此设置较高的 $\theta$ 值会降低模型对大目标缺陷的识别效果。此外,当 $\theta$ 大于 $0.75\pi$ 时,模型增加了难分类样本的学习权重,SFL对系统整体性能的抑制效果较为明显,缺陷的整体召回率偏低。上述实验说明,难易分类样本的学习权重分配策略能够影响模型的召回率,SFL的正弦衰

减因子  $s$  的变化范围应适中,因此后续实验使用参数  $\theta = 0.5\pi$  设置下的 SFL。

为更深入研究注意力机制模块的引入对模型检测速度和精度的影响,开展了关于注意力机制模块位置的消融实验,结果如表 5 所示。其中,注意力机制模块的位置参考图 1 中的加权特征融

合位置,Add1 表示 feat2 的浅层特征融合处,Add2 表示 feat1 的深层特征融合处。模型选用 CIoU Loss, SFL 作为回归损失函数和分类损失函数, SFL 参数设置为:  $\alpha = 0.75, \gamma = 1, \theta = 0.5\pi$ 。FNF 表示归一化的加权特征融合, 0 表示未启用相关模块, 1 表示启用相关模块。

表 5 不同位置融合 Di\_CBAM 的特征融合网络性能对比

Tab. 5 Performance comparison of feature fusion networks with Di\_CBAM introduced at different locations

Index	Location				FNF	FPS/ (frame·s <sup>-1</sup> )	mAP <sub>0.5</sub> /%	mAP <sub>0.75</sub> /%
	Add1	Add2	Add3	Add4				
1	0	0	0	0	1	32.22	89.57	33.28
2	1	1	0	0	1	29.98	89.66	35.68
3	0	0	1	1	1	30.84	90.92	38.21
4	0	1	1	1	1	29.71	89.69	34.57
5	1	1	1	1	1	27.70	89.59	35.98
6	1	1	1	1	0	27.78	88.46	33.52

由表 5 可知,引入 Di\_CBAM 模块后模型的检测帧率存在明显变化,并且在特征融合过程中使用 FNF 能够有效提高 mAP。其中,模型检测帧率的下降主要由特征层的多层感知机运算引起,由于拼接后的特征通道数变为原通道数的 3 倍,增加了计算开销,最终导致检测帧率下降。在 feat2 和 feat3 层的底层特征融合处引入注意力机制,即图 1 中的 Add3 和 Add4 位置,模型能达到 30.84 frame/s 的检测速度,并且 mAP<sub>0.5</sub> 达到

90.92%,说明在此位置引入注意力机制既能维持较高的检测精度,又能保证模型以较快速度运行。

为验证替换特征融合网络、引入注意力机制和设计分类损失函数 3 个改进策略对模型的耦合影响,开展了消融实验,结果如表 6 所示。其中,Di\_CBAM 引入位置位于 Add3 和 Add4 处, SFL 参数设置为:  $\alpha = 0.75, \gamma = 1, \theta = 0.5\pi$ , 0 表示未启用相关模块, 1 则表示启用相关模块。

表 6 改进策略对模型性能的影响对比

Tab. 6 Comparison of impact of improvement strategies on model performance

Index	BiFPN	Di_CBAM	SFL	FPS/(frame·s <sup>-1</sup> )	mAP <sub>0.5</sub> /%	mAP <sub>0.75</sub> /%
1	0	0	0	28.00	86.52	39.07
2	1	0	0	32.20	90.06	39.53
3	0	1	0	27.60	88.13	39.54
4	0	0	1	28.00	86.89	38.86
5	1	1	0	30.84	90.38	39.65
6	1	0	1	32.00	91.37	36.56
7	0	1	1	27.60	88.26	38.97
8	1	1	1	30.84	90.92	38.21

对比表 6 中的第 1 组与第 4 组、第 2 组与第 6 组、第 3 组与第 7 组实验结果可知,使用 SFL 不会降低检测速度,能够起到提高  $mAP_{0.5}$  和降低  $mAP_{0.75}$  的作用。第 3 组实验单独使用 Di\_CBAM 的模型相对于 YOLOX 原始模型在检测精度方面提升较明显,  $mAP_{0.5}$  和  $mAP_{0.75}$  分别提高 1.61% 和 0.47%, 但检测速度出现轻微下降, 说明引入 Di\_CBAM 虽然会降低模型的检测速度, 但能够有效提高模型的检测精度。在第 7 组同时使用 BiFPN 和 SFL 的实验中, 检测速度达到 32.00 frame/s,  $mAP_{0.5}$  达到 91.37%, 但  $mAP_{0.75}$

降至 36.56%; 在第 8 组启用 Di\_CBAM 的实验中, 虽然检测速度和  $mAP_{0.5}$  指标出现了轻微下降, 但能将  $mAP_{0.75}$  提升至 38.21%, 说明同时启用 3 个改进策略能够确保模型在检测速度和检测精度之间取得平衡。

最后, 为验证本文设计的 BiYOLOX 算法的性能表现, 将本文算法与替换卷积操作的轻量化 YOLOv3<sup>[11]</sup>、轻量化特征提取网络的 YOLOv5<sup>[10]</sup> 和 YOLOX 算法进行对比实验。模型性能如表 7 所示, 高亮反射金属锅圆柱表面缺陷的部分检测结果如图 10 所示。

表 7 轻量化改进模型的性能对比

Tab. 7 Performance comparison of lightweight improved models

Model	Params/M	FPS/(frames·s <sup>-1</sup> )	$mAP_{0.5}/\%$	$mAP_{0.75}/\%$
Light-YOLOv3	8.6	28.70	86.12	35.92
Light-YOLOv5	7.3	29.06	87.35	37.48
YOLOX	9.0	28.00	86.52	39.07
BiYOLOX(Di_CBAM+SFL)	6.4	30.84	90.92	38.21

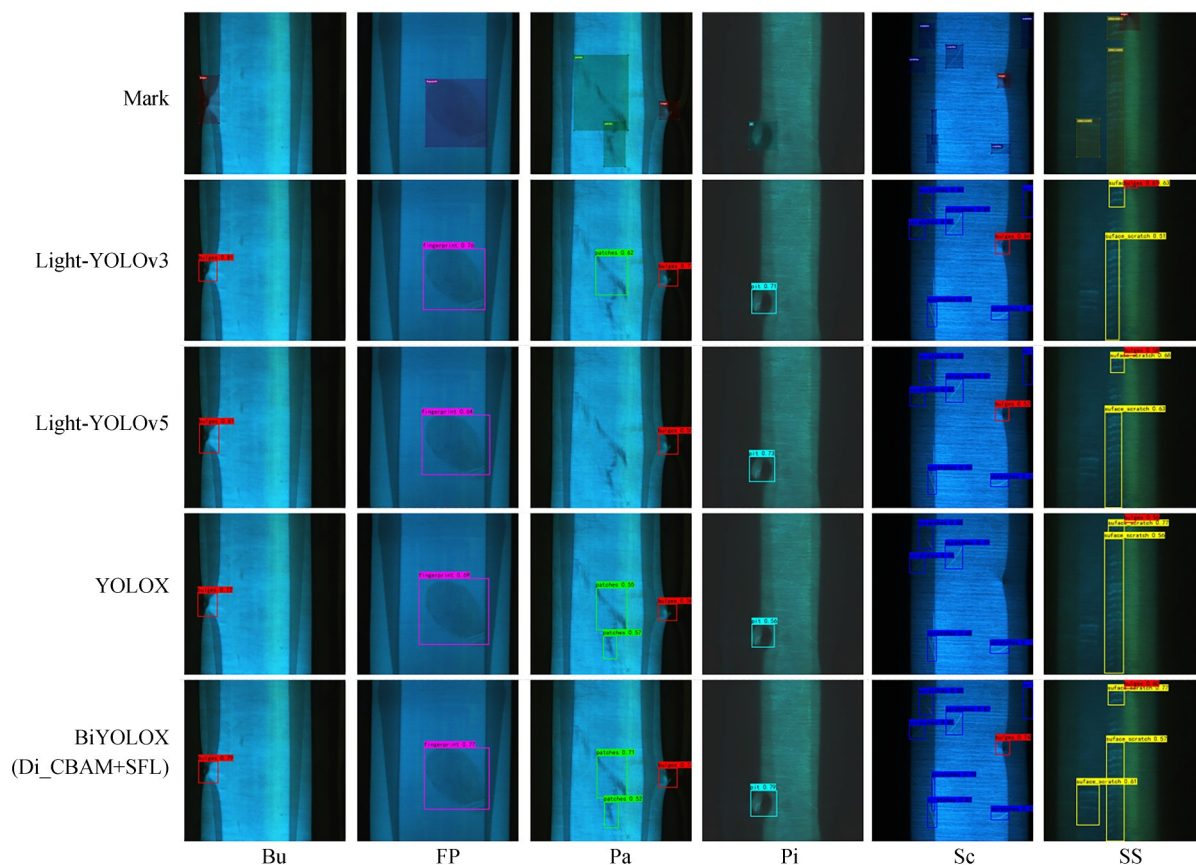


图 10 高亮反射金属锅圆柱表面缺陷检测结果

Fig. 10 Detection results of cylindrical surface defects of high light reflective metal pot

由表 7 可知,本文设计的 BiYOLOX 算法相比于列举的轻量化 YOLO 系列算法,模型的参数量更少、单张图像的检测速度提升明显,并且获得了较高的检测精度。由图 10 可知,在斑块缺陷检测中,Light-YOLOv3, Light-YOLOv5 均存在漏检;在划痕缺陷检测中,图像右侧的凸起和左下方的划痕缺陷在 Light-YOLOv3, Light-YOLOv5 和 YOLOX 算法中被漏检,并且 YOLOX 算法还漏检了一处凸起缺陷;在擦伤缺陷检测中,图像左侧的一处缺陷在 Light-YOLOv3, Light-YOLOv5 和 YOLOX 算法中均出现漏检,而本文提出的融合注意力机制和 SFL 的 BiYOLOX 算法则能够准确定位缺陷目标位置,无错检、漏检情况,并且预测框的置信度普遍高于前者,说明本文设计的 SFL 能够有效提高模型对目标的分类效果。分析斑块、划痕和擦伤的检测数据可知,Light-YOLOv3, Light-YOLOv5 和 YOLOX 漏检目标的尺寸与位置不一,此现象的出现归结于网络的输出特征层信息不够丰富。实验结果表明,模型通过引入注意力机制学习特征的通道和空间信息,以跨尺度的特征融合方式获得了更丰富的特征信息,对提高金属锅

圆柱表面缺陷的检测效果具有积极作用。

## 4 结 论

本文针对金属锅圆柱表面缺陷高精度快速检测的深度学习模型设计需求,提出了基于注意力机制的轻量化特征融合网络,并利用 BiFPN 替换 YOLOX 网络的特征融合层,在降低网络计算资源消耗的情况下提高多尺度特征在通道和空间上的表达。针对训练过程中正负样本不平衡的问题,设计了基于衰减因子的分类损失函数 SFL,对难易分类样本的学习权重进行重新分配。完成了在金属锅圆柱表面缺陷数据集上对 BiYOLOX 与 YOLOX 的特征融合网络、分类损失函数对比实验和注意力机制模块位置消融实验。实验结果显示,所设计的网络能够有效降低模型参数量和浮点计算数,并提高检测速度和精度,较原始 YOLOX 的 Parames 减少 2.6M, FLOPs 下降 4.053 G,模型检测帧率提升至 30.84 frame/s,检测精度  $mAP_{0.5}$  达到 90.92%,实现了金属锅圆柱表面缺陷的高精度快速识别与定位。

### 参考文献:

- [1] 张静,叶玉堂,谢煜,等. 金属圆柱工件缺陷的光电检测[J]. 光学精密工程, 2014, 22(7): 1871-1876.  
ZHANG J, YE Y T, XIE Y, *et al.* Optoelectronic inspection of defects for metal cylindrical workpieces [J]. *Optics and Precision Engineering*, 2014, 22(7): 1871-1876. (in Chinese)
- [2] 石炜,张袁祥,李嘉楠. 列车滚子轴承表面缺陷机器视觉检测方法研究[J]. 机械设计与制造, 2022(4): 183-186.  
SHI W, ZHANG Y X, LI J N. Research on machine vision detection method for surface defects of train roller bearings[J]. *Machinery Design & Manufacture*, 2022(4): 183-186. (in Chinese)
- [3] LIN J H, YAO Y, MA L, *et al.* Detection of a casting defect tracked by deep convolution neural network [J]. *The International Journal of Advanced Manufacturing Technology*, 2018, 97(1): 573-581.
- [4] SUN X H, GU J N, HUANG R, *et al.* Surface defects recognition of wheel hub based on improved faster R-CNN[J]. *Electronics*, 2019, 8(5): 481.
- [5] 王宸,张秀峰,刘超,等. 改进 YOLOv3 的轮毂焊缝缺陷检测[J]. 光学精密工程, 2021, 29(8): 1942-1954.  
WANG CH, ZHANG X F, LIU CH, *et al.* Detection method of wheel hub weld defects based on the improved YOLOv3[J]. *Optics and Precision Engineering*, 2021, 29(8): 1942-1954. (in Chinese)
- [6] 程婧怡,段先华,朱伟. 改进 YOLOv3 的金属表面缺陷检测研究[J]. 计算机工程与应用, 2021, 57(19): 252-258.  
CHENG J Y, DUAN X H, ZHU W. Research on metal surface defect detection by improved YOLOv3 [J]. *Computer Engineering and Applications*, 2021, 57(19): 252-258. (in Chinese)
- [7] GE Z, LIU S T, WANG F, *et al.* YOLOX: exceeding YOLO Series in 2021 [EB/OL]. 2021: arXiv: 2107.08430. <https://arxiv.org/abs/2107.08430>
- [8] 伍济钢,成远,邵俊,等. 基于改进 YOLOv4 算法

- 的PCB缺陷检测研究[J]. 仪器仪表学报, 2021, 42(10): 171-178.
- WU J G, CHENG Y, SHAO J, *et al.* A defect detection method for PCB based on the improved YOLOv4[J]. *Chinese Journal of Scientific Instrument*, 2021, 42(10): 171-178. (in Chinese)
- [9] 罗钧, 曾伟, 龚燕峰, 等. 融入IBN-NET的轻量网络在金属圆柱工件缺陷识别中的应用[J]. 计算机辅助设计与图形学学报, 2020, 32(1): 112-120.
- LUO J, ZENG W, GONG Y F, *et al.* Application of lightweight network with IBN-NET in defect recognition of metal cylindrical workpieces[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2020, 32(1): 112-120. (in Chinese)
- [10] LAN Y, XU W X. Insulator defect detection algorithm based on a lightweight network[J]. *Journal of Physics: Conference Series*, 2022, 2181(1): 012007.
- [11] SHU Y F, LI B, LI X M, *et al.* Deep learning-based fast recognition of commutator surface defects [J]. *Measurement*, 2021, 178: 109324.
- [12] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 10778-10787.
- [13] ZHENG Z H, WANG P, REN D W, *et al.* Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. *IEEE Transactions on Cybernetics*, 2022, 52(8): 8574-8586.
- [14] LIU S T, HUANG D, WANG Y H. *Learning Spatial Fusion for Single-Shot Object Detection* [EB/OL]. 2019: *arXiv*: 1911.09516. <https://arxiv.org/abs/1911.09516>
- [15] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 8759-8768.
- [16] WANG C Y, MARK LIAO H Y, WU Y H, *et al.* CSPNet: a new backbone that can enhance learning capability of CNN[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 14-19, 2020, Seattle, WA, USA. IEEE, 2020: 1571-1580.
- [17] TAN M X, LE Q V. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks* [EB/OL]. 2019: *arXiv*: 1905.11946. <https://arxiv.org/abs/1905.11946>
- [18] 陈欣, 万敏杰, 马超, 等. 采用多尺度特征融合SSD的遥感图像小目标检测[J]. 光学精密工程, 2021, 29(11): 2672-2682.
- CHEN X, WAN M J, MA CH, *et al.* Recognition of small targets in remote sensing image using multi-scale feature fusion-based shot multi-box detector [J]. *Opt. Precision Eng.*, 2021, 29(11): 2672-2682. (in Chinese)
- [19] 梁水波, 刘紫燕, 袁浩, 等. 注意力和多重特征融合的图像局部特征检测及描述[J]. 传感技术学报, 2021, 34(8): 1075-1081.
- LIANG SH B, LIU Z Y, YUAN H, *et al.* Image local feature detection and description based on attention and multi-layer feature fusion [J]. *Chinese Journal of Sensors and Actuators*, 2021, 34(8): 1075-1081. (in Chinese)
- [20] 朱威, 王立凯, 靳作宝, 等. 引入注意力机制的轻量级小目标检测网络[J]. 光学精密工程, 2022, 30(8): 998-1010.
- ZHU W, WANG L K, JIN Z B, *et al.* Lightweight small object detection network with attention mechanism [J]. *Optics and Precision Engineering*, 2022, 30(8): 998-1010. (in Chinese)
- [21] WOO S, PARK J, LEE J Y, *et al.* CBAM: Convolutional Block Attention Module [M]. *Computer Vision-ECCV 2018*. Cham: Springer International Publishing, 2018: 3-19.
- [22] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
- [23] SHAHINFAR S, MEEK P, FALZON G. "How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring [J]. *Ecological Informatics*, 2020, 57: 101085.
- [24] YU J H, JIANG Y N, WANG Z Y, *et al.* Unit-Box: an advanced object detection network [C]. *Proceedings of the 24th ACM international conference on Multimedia*. October 15 - 19, 2016, Amsterdam, The Netherlands. New York: ACM,

2016: 516-520.

- [25] REZATOFIGHI H, TSOI N, GWAK J, *et al.* Generalized intersection over union: a metric and a loss for bounding box regression[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long

Beach, CA, USA. IEEE, 2020: 658-666.

- [26] ZHENG Z H, WANG P, LIU W, *et al.* Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993-13000.

#### 作者简介:



乔 健(1980—),男,山西忻州人,博士,教授,博士生导师,2003年、2008年于吉林大学分别获得学士、博士学位,主要从事光学精密传动机构设计、光学精密检测技术、非晶材料先进加工技术与装备等方面的研究。E-mail: qiaojj99065@163.com

#### 通讯作者:



杨景卫(1983—),男,河南许昌人,博士,副教授,硕士生导师,2007年于吉林大学获得学士学位,2015年于华南理工大学获得博士学位,主要从事机电一体化、焊接自动化等方面的研究。E-mail: mejwyang@fosu.edu.cn