

基于改进 YOLOx 网络的金属齿轮表面缺陷检测方法

张曙文^{1,2}, 钟振宇^{1,2}, 朱大虎^{1,2*}¹武汉理工大学汽车工程学院现代汽车零部件技术湖北省重点实验室, 湖北 武汉 430070;²武汉理工大学汽车工程学院汽车零部件技术湖北省协同创新中心, 湖北 武汉 430070

摘要 针对工业干扰环境下金属齿轮表面缺陷自动化检测容易出现误检和漏检的问题, 提出一种改进的 YOLOx 算法。首先, 通过 adaptively spatial feature fusion (ASFF) 充分利用不同尺度下缺陷和干扰项的特征之间的差异, 提高模型的抗干扰能力; 接着, 通过 efficient channel attention (ECA) 模块, 增加网络的特征提取能力; 最后, 修改置信度损失函数为 Varifocal 损失函数, 减少困难样本对网络的干扰。实验结果表明, 改进的 YOLOx 网络与原版相比在召回率、准确率和平均精度均值上分别提升 6.1 个百分点、4.6 个百分点和 9.4 个百分点。

关键词 机器视觉; 齿轮缺陷检测; YOLOx 算法; 注意力机制; 特征融合

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP230469

Gear Surface Defect Detection Method Based on Improved YOLOx Network

Zhang Shuwen^{1,2}, Zhong Zhenyu^{1,2}, Zhu Dahu^{1,2*}¹Hubei Key Laboratory of Advanced Technology for Automotive Components, School of Automotive Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, China;²Hubei Collaborative Innovation Center for Automotive Components Technology, School of Automotive Engineering, Wuhan University of Technology, Wuhan 430070, Hubei, China

Abstract Herein, an improved YOLOx algorithm is proposed to address the challenges concerning false and missing detection of metal gear surface defects in an industrial interference environment. First, by utilizing the adaptive spatial feature fusion (ASFF) to fully utilize the differences between the features of defects and interference items at different scales, the model's anti-interference ability is improved. Second, through the effective channel attention (ECA) module, the network's feature extraction capability is increased. Finally, the confidence loss function is modified to the Varifocal loss function, which reduces the interference of complex samples in the network. Experimental results indicate that the improved YOLOx network outperforms the original network. Particularly, the recall rate, accuracy, and mean average precision indexes of the improved YOLOx network are improved by 6.1, 4.6, and 9.4 percentage points, respectively, as compared with the original network.

Key words machine vision; gear defect detection; YOLOx algorithm; attention mechanism; feature fusion

1 引言

金属齿轮是机械设备及汽车中的重要基础零件之一, 在其加工生产过程中不可避免地会出现崩角、划痕、碰伤等表面缺陷, 进而影响运动部件机构的使用寿命、运动精度等。目前齿轮生产制造企业主要依靠人工目视检测齿轮缺陷, 劳动强度大, 并存在误检、漏检,

影响机械设备后期运行安全。因此, 迫切需要一种具有检测效率高、检测准确率高的齿轮表面缺陷自动化检测方法。现有的齿轮表面缺陷检测方法主要可以分为基于信号的方法^[1-4]和基于图像的方法^[5], 分别从应变测量、波形分析以及颜色特征等方面进行检测, 但是这些方法对于静态下的齿轮表面缺陷的效果并不好。目标检测类的深度学习算法在近些年的检测任务中表

收稿日期: 2023-01-09; 修回日期: 2023-02-20; 录用日期: 2023-03-06; 网络首发日期: 2023-03-10

基金项目: 国家自然科学基金(51975443)

通信作者: *dhzhu@whut.edu.cn

现较为出色,将深度学习算法引入缺陷检测是近些年的研究趋势之一。

目标检测类的深度学习算法主要可以分为两阶段和一阶段算法。两阶段的目标检测算法以 R-CNN^[6-8] 系列为代表,主要包括 R-CNN^[6]、Fast R-CNN^[7] 以及 Faster R-CNN^[8]。其主要特点是需要先在图像上分割得到一些候选区域,然后在这些候选区域的基础上利用卷积网络提取特征,最后送入分类器中进行分类与回归。R-CNN 在分割时利用随机选择搜索算法筛选出候选区域,然后将这些候选区域送入卷积神经网络(CNN)提取特征,最后利用支持向量机(SVM)分类器进行分类和回归。R-CNN 的速度较慢而且训练需要的空间较大,为了解决这个问题在 R-CNN 的基础上提出 Fast R-CNN^[7] 算法,添加的感兴趣区域(ROI)池化方法将特征图矩阵缩放到目标尺度,然后对特征图进行分类和回归。而且 Faster R-CNN^[8] 在筛选候选框的时候采用区域候选网络(RPN),加快了候选框的筛选。

一阶段的算法主要包括 YOLO^[9-13] 系列和 SSD^[14-17] 系列。SSD 系列结合 YOLO 中的回归思想和两阶段网络中的锚框机制,这样既保持了 YOLO 系列检测速度快的特点又保证了预测的精准性^[14]。SSD 系列的核心思想是在不同尺度的特征图上采用卷积核来预测一系列边界框的类别和坐标偏移,但是其网络结构较为简单,对于复杂特征的检测效果并不理想^[14]。YOLO 系列从最初的 YOLOv1^[9] 发展到最新的 YOLOx^[13],其主要的思想是将输入的图片划分为 $s \times s$ 个框,然后在每个框内进行目标检测,YOLOv2~YOLOv5 均采用先验框的方法,通过对标注的目标框进行聚类先设定几组不同尺度的先验框,预测的时候在先验框的基础上进行回归,这样可加快预测的速度,但是针对不同的训练集需要设置不同的先验框较为麻烦,因此 YOLOx 回归了无先验框的方法。

鉴于目标检测类的深度学习算法取得的巨大成功,越来越多的研究人员将目标检测类的深度学习算法运用到各种检测中。董乙杉等^[18] 利用改进后的 YOLOv5 算法检测 X 光图片中的违禁品,解决了图片中的遮挡和复杂背景干扰问题。郎磊等^[19] 在 YOLOx 算法中添加协同注意力模块提升特征的关注度并优化损失函数,在平衡精度和速度的情况下实现了轻量级遥感图像中的目标检测。郭岚等^[20] 利用注意力机制 squeeze-and-excitation(SE) 模块和 ghost module 模块对 YOLOv5 网络进行改进以实现光伏组件中缺陷的精准定位和检测。袁田等^[21] 利用 YOLOv4 算法识别交通场景中的道路使用者,并以此分析驾驶人在不同运动类型和交通密度下反应特性的差异。刘子龙等^[22] 提出一种改进的 YOLOv5 算法用以检测路面上的小目标。李亚东等^[23] 将 YOLOx 算法用于检测检测轴承缺陷小目标。上述算法证明了目标检测类算法在实际

检测中的有效性。

目标检测算法中的 R-CNN 系列的检测精度虽然很高,但是在训练检测的过程中需要大量的存储空间划分候选框,对于实际的应用来说需要的条件更为严格,并且 R-CNN 系列的检测速度较慢,不能满足实际情况下的检测需求。从上述参考文献中可以看出,YOLO 系列已经被广泛应用在实际应用场景下的缺陷检测任务之中,YOLOx 算法作为 YOLO 系列的最新一代,不仅保留了 YOLO 系列检测速度快的特点,而且检测精度相比之前还有提升。在实际的生产过程中,齿轮表面会有油污、防锈粉等干扰物,这些干扰项在图像中的特征和缺陷的特征十分相似,会对检测造成一定的干扰。而直接利用 YOLOx 算法进行检测并不能避免这些干扰项的影响,在训练的时候会存在误学到非目标特征的情况,导致网络的检测效果有所下降,不能满足实际的检测需求。

为了满足实际的检测需求,需要对网络的结构进行一定的改进。注意力机制作为近些年深度学习领域新兴的方法,能够使网络在训练的时候更加专注在目标特征上,因此添加注意力机制能够在一定程度上减小干扰项的影响。而干扰项的特征虽然和缺陷有些接近,但是在不同尺度上具有一定差异,因此融合不同尺度的特征能够使网络的权重更多地分配在缺陷的特征上,确保网络能够学习到足够多的缺陷特征。

2 检测方法

2.1 YOLOx 算法

YOLOx 算法的网络架构如图 1 所示,主要由 4 个部分组成,分别是网络的输入端、进行目标特征提取的骨干网络、进行特征融合的颈部网络以及检测头^[13]。

YOLOx 的输入端主要对输入的图片进行处理,通过 mosaic 方法和 mixup 方法丰富数据集的多样性,增加小目标的数量,从而增强检测效果。骨干网络的主要作用是通过卷积操作提取不同尺度下的目标特征并将这些特征图送入颈部网络中进行特征融合以提升特征的多样性和鲁棒性。图 1 展示了颈部网络的特征金字塔网络(FPN)^[24] 和路径聚合网络(PAN)^[25] 两种融合特征的方法。其中,自顶向下的特征融合的方法为 FPN 方法,通过上采样的方式进行特征融合,而自下向上的特征融合的方法则为 PAN 方法,这两种方法的结合能够从不同的主干层对不同的检测层进行特征聚合。特征融合完成之后的特征图为 $20 \times 20 \times 512$ 、 $40 \times 40 \times 256$ 、 $80 \times 80 \times 128$ 等 3 个尺度。特征图越小代表每个网格中的图像的面积越大,能识别的目标的尺寸也越大。

骨干网络中最主要的是 CSPDarknet 结构。CSPDarknet 结构是在 Darknet53 网络结构的基础上借鉴了 CSPNet 的思路改进而来的,有效解决了大型卷积网络在 backbone 中网络优化的梯度信息重复问题,

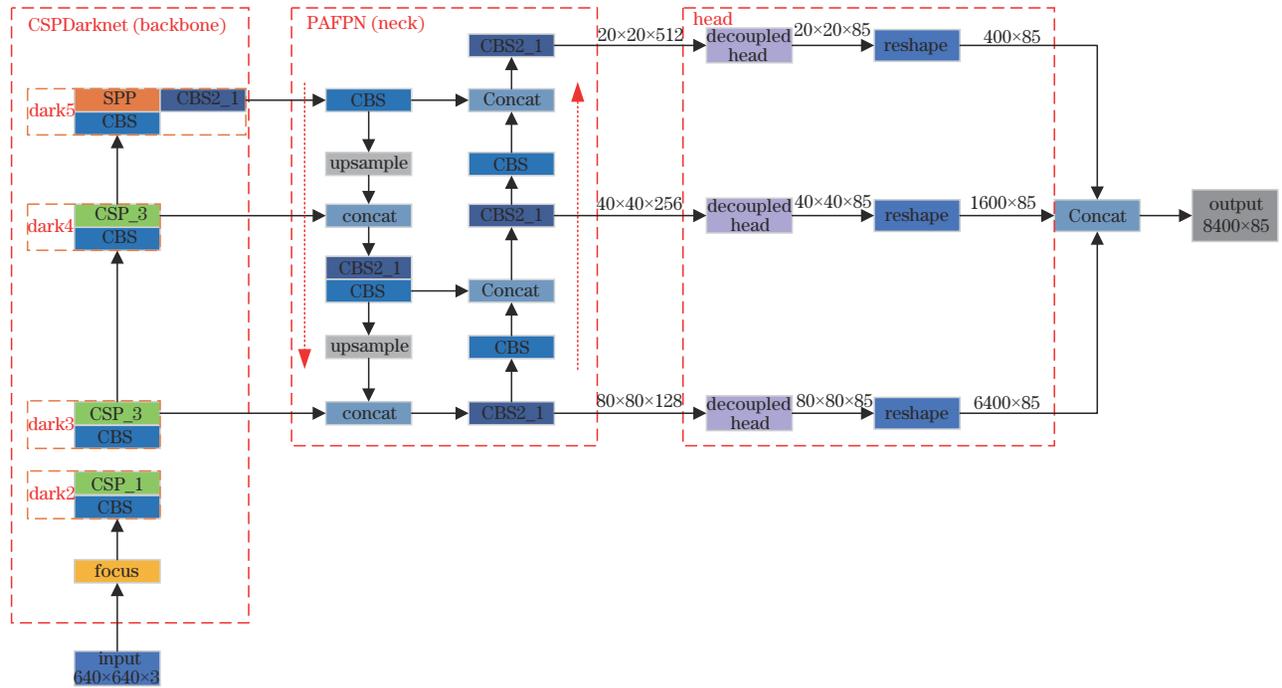


图 1 YOLOx 算法的主体结构
Fig. 1 The main structure of the YOLOx algorithm

将梯度的变化从头到尾集成在特征图中^[12],其中主要包含了CBS组件和CSP_X组件。CBS的结构包括卷积、归一化层和SiLU激活函数。而CSP_X组件的结构如图2所示,其中,X代表残差块的数量。检测头部分与以往的YOLO系列不同,采用了解耦头的方式进行检测,如图3所示。在进行Concat操作之前,一共有3个分支,分别代表分类输出、目标输出和回归输出。分类输出部分主要对应的是目标框的类别和预测分数,类别输出部分主要判断目标框是背景还是前景,而回归输出分支对目标框的坐标信息(x,y,w,h)进行预测。在检测头部分一共有3个尺度的解耦头,分别对

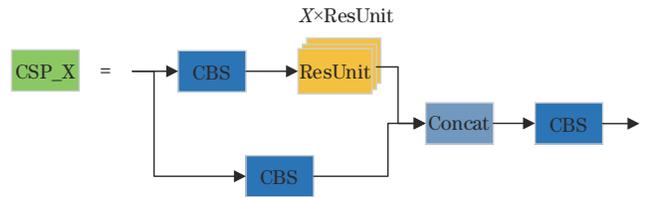


图 2 CSP_X 结构图
Fig. 2 Structure diagram of CSP_X

应不同大小的目标,在解耦头的输出部分进行Concat操作得到8500x85维的特征向量,最终在这个特征向量中检测目标。

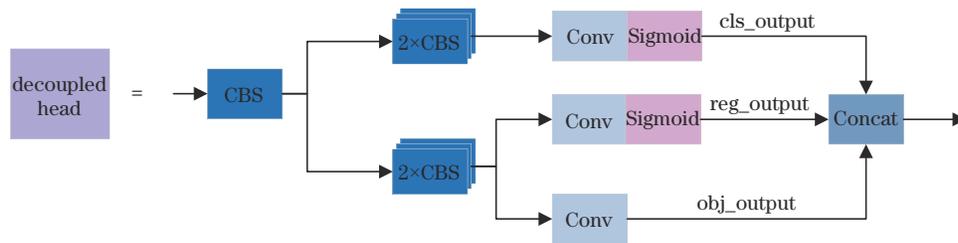


图 3 解耦头结构
Fig. 3 Structure of decoupled head

2.2 ASFF 方法

adaptively spatial feature fusion(ASFF)作为一种自适应特征融合方法,在进行特征融合时能够自动学习,过滤掉其他层无用的信息,保留有用的信息以达到高效的特征融合^[26]。具体步骤如下:在特征融合时对于某一层的特征,在融合其他层特征之前先将其他层的特征尺寸调整到与当前层相同,然后通过加权的方式将其他层的特征与本层的特征融合。ASFF的结构

如图4所示,左边是一个经典的FPN结构,图中3个特征图的下采样倍数分别是32、16和8。以ASFF-3的特征融合为例,ASFF-3的输出由3个尺度的特征图共同决定,融合方式^[26]可描述为

$$Y_{ASFF-3} = X^{1 \rightarrow 3} \otimes \alpha^3 + X^{2 \rightarrow 3} \otimes \beta^3 + X^{3 \rightarrow 3} \otimes \gamma^3, \quad (1)$$

式中: α, β, γ 是可学习的参数; \otimes 代表卷积操作; $X^{i \rightarrow j}$ 代表将第*i*层的特征调整到第*j*层特征的尺度后的特

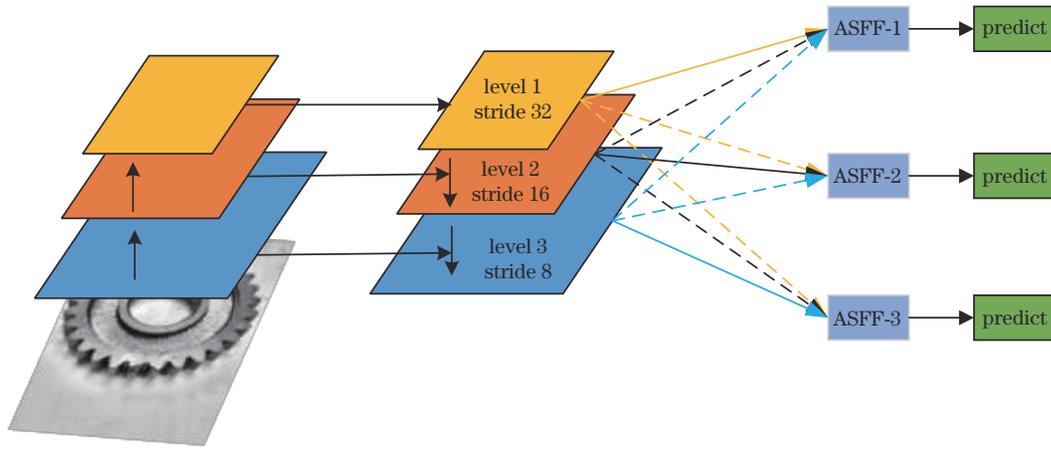


图 4 经典 ASFF 结构示意图

Fig. 4 Schematic diagram of the structure of classic ASFF

征图。将式(1)推广到更一般的形式,令 $\mathbf{x}_{ij}^{n \rightarrow l}$ 表示第 n 层特征图上的特征向量在第 l 层上对应的特征向量,则第 l 层特征图的方式^[26]为

$$y_{ij}^l = \alpha_{ij}^l \cdot \mathbf{x}_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot \mathbf{x}_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot \mathbf{x}_{ij}^{3 \rightarrow l}, \quad (2)$$

$$\begin{cases} \alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \\ \alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \\ \beta_{ij}^l = \frac{e^{\lambda_{\beta_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \\ \gamma_{ij}^l = \frac{e^{\lambda_{\gamma_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \end{cases}, \quad (3)$$

式中: y_{ij}^l 代表输出特征图上对应位置的特征向量; α_{ij}^l 、 β_{ij}^l 和 γ_{ij}^l 是在第 l 层特征图学习到的权重^[26]; λ^l 由 1×1 卷积分别作用在 $\mathbf{x}^{n \rightarrow l}$ 上得到,可以通过反向传播更新。

YOLOx 颈部网络中的特征融合采用的是 PAFPN 结构,FPN 是自顶向下的,将高层的特征通过

上采样和底层特征进行融合得到预测的特征图,从而增强多个尺度上的语义表达。PANet 的结构是自底向上的,将低层的特征传递到高层特征之中,增强多个尺度上的定位能力。FPN 会导致不同尺度特征的不一致性,具体地,在 FPN 形式的网络中启发式地选择特征,高层语义信息中检测大目标、低层语义信息检测小目标。当某个目标在某一层上被当作正类时,相应的目标区域在其他层被当作负类。而一张图上既有大目标也有小目标,这样不同层间的特征的不一致性会影响最后的检测结果。

YOLOx 中 backbone 输出的特征图经过颈部网络的 PAFPN 进行特征融合后会输出 3 个尺度的特征图,分别为 $20 \times 20 \times 512$ 、 $40 \times 40 \times 256$ 和 $80 \times 80 \times 128$,将这 3 个尺度的特征图利用 ASFF 模块融合,在颈部网络之后添加 ASFF 模块,修改后的结构示意图如图 5 所示。ASFF 模块的每个输出都是融合了 3 个尺度的特征,将输出和检测头部分连接,这样能够更好地避免干扰项的影响。

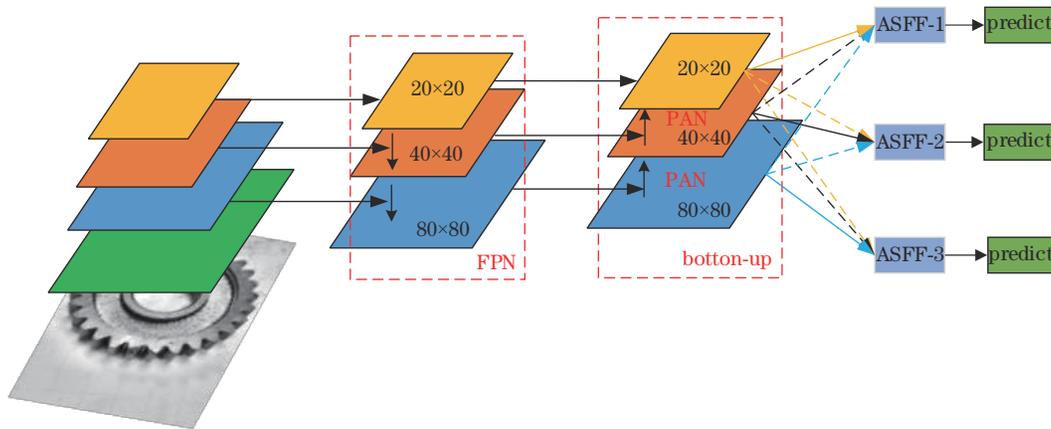


图 5 修改后的结构示意图

Fig. 5 Revised structural diagram

2.3 注意力机制

efficient channel attention(ECA)注意力机制是在

SE-Layer 的基础上修改得到的^[27],SE-Layer 模块的示意图如图 6 所示,具体流程图如图 7 所示^[28]。SE-Layer

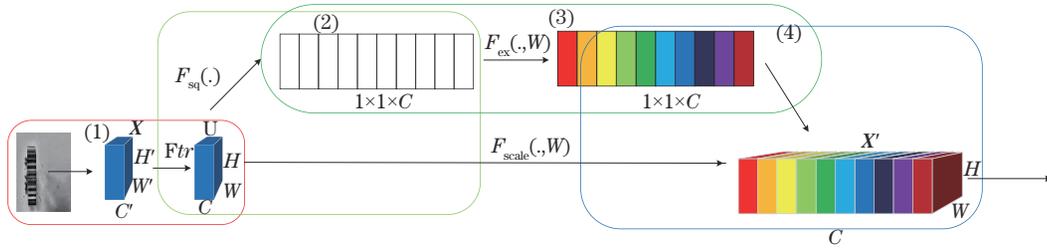


图 6 SE-Layer 示意图

Fig. 6 Schematic diagram of SE-Layer

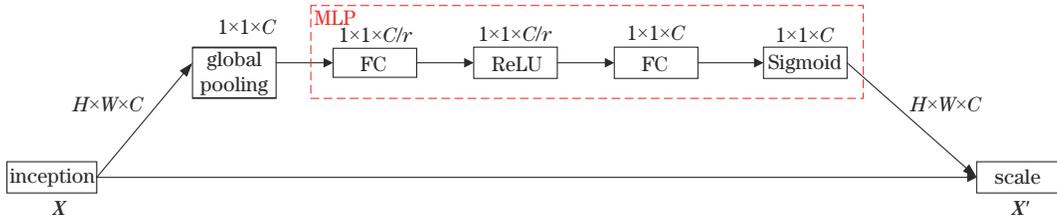


图 7 SE-Layer 流程图

Fig. 7 Flowchart of the SE-Layer

从输入的特征图 X 中提取特征,当前的特征层 U 的维度为 $[C, H, W]$ 。其中, C 代表通道数, H 代表特征图高度, W 代表特征图宽度。然后对特征图的 $[H, W]$ 维度进行平均池化或者最大池化,经过池化后的特征图的大小从 $[C, H, W]$ 变为 $[C, 1, 1]$,即为图 6 中操作 (2)。 $[C, 1, 1]$ 代表对于每一个通道 C 都有一个权重参数与之对应,而 $[C, 1, 1]$ 的特征可以理解为由每个通道提取出来的权重,权重则表示了每个通道对特征提取的影响力。全局池化后的向量通过多层感知器 (MLP) 网络后,可得到每个通道的权重,即图 6 中操作 (3)。然后将权重 $[C, 1, 1]$ 作用于特征图 $U [C, H, W]$,即每个通道各自乘以各自的权重,即图 6 中操作 (4)。可以理解为:当权重较大时,该通道特征图的数值相应增大,对最终输出的影响也会变大;当权重小时,该通道特征图的数值就会更小,对最终输出的影响也会变小。而 SE-Layer 之所以有用是因为在特征提取的过程中不可避免地会出现有些特征图作用大而另一些特征图作用小的情况,而由通道本身提取出的

权重施加在特征图上,保证了在特征图提取特征的基础上,自适应地给定通道权重,让作用大的特征图对结果的影响更大。因此,SE-Layer 可比普通的卷积层更有效地提取特征。

而 SE-Layer 在有效提取特征的前提下增加了模型的复杂性,这会不可避免地会对模型的性能造成影响。为了克服性能和复杂性之间的矛盾,ECA 模块在 SE 的基础上被提出,其结构如图 8 所示。将 SE-Layer 步骤 (3) 中的 MLP 模块转变为一维卷积的形式,给定通过全局平均池化 (GAP) 获得的聚合特征 $[C, 1, 1]$,ECA 模块通过执行卷积核大小为 k 的一维卷积来生成通道权重,其中, k 通过通道维度 C 的映射自适应确定。自适应确定卷积核大小^[27]的公式为

$$k = \left\lfloor \frac{\log_2 C + b}{\gamma} \right\rfloor_{\text{odd}}, \quad (4)$$

式中: k 表示卷积核的大小; C 表示通道数; $\lfloor \cdot \rfloor_{\text{odd}}$ 代表 k 只能取奇数; γ 和 b 一般设置为 2 和 1,用于改变通道数 C 和卷积核大小之间的比例。

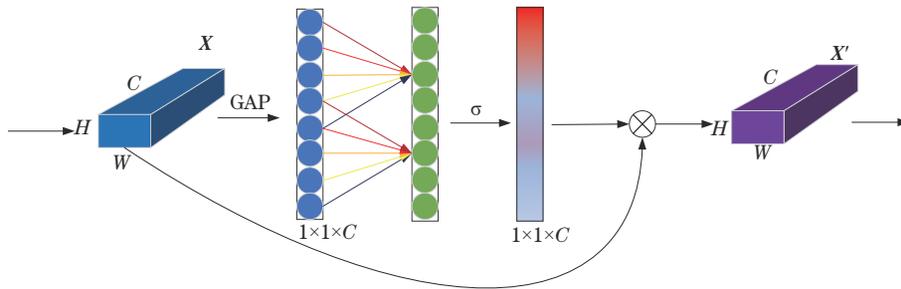


图 8 ECA 示意图

Fig. 8 Schematic diagram of ECA

在 YOLOx 网络中插入 ECA 模块改进后的网络结构如图 9 所示。ECA 模块在 backbone 之后,从

backbone 提取的特征中再次提取目标特征,在训练的过程中自适应地确定卷积核 k 的值,然后将权重更多

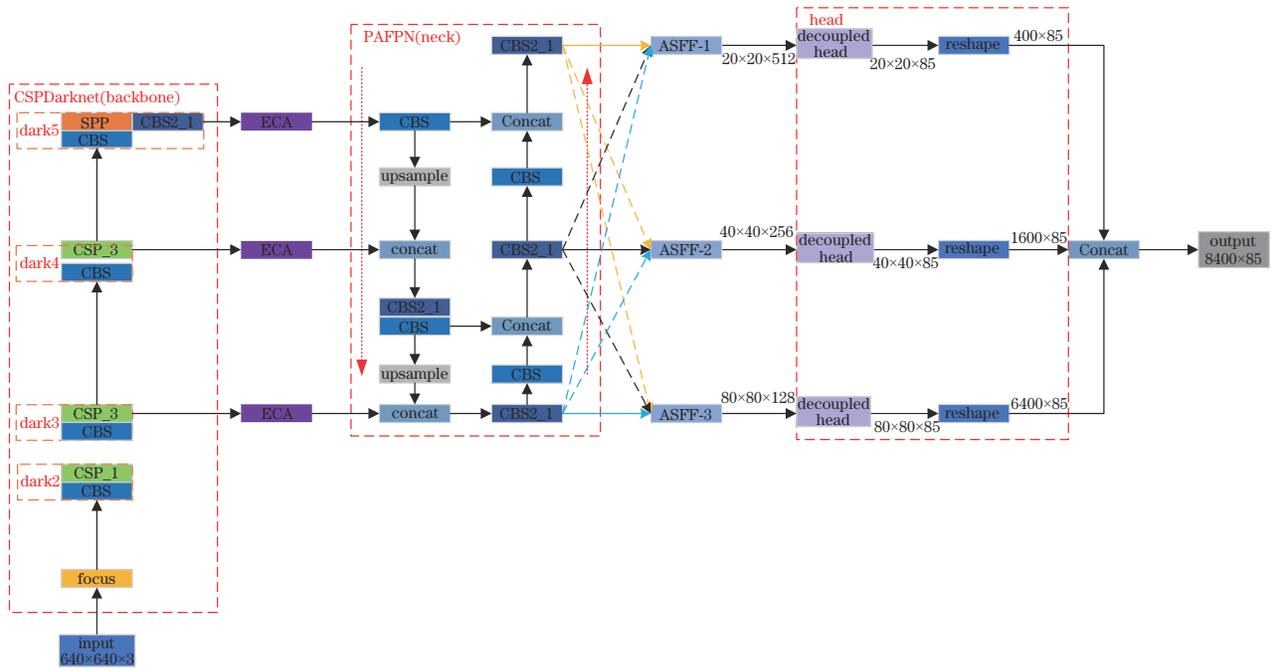


图9 改进后的YOLOx结构
Fig. 9 Structure of improved YOLOx

地分配在缺陷的特征上而不是背景或者干扰项上,这样能有效地减小干扰项的影响。

2.4 修改预测置信度的损失函数

解耦头结构,在原始网络中 Obj 分支采用的是 binary cross entropy(BCE)损失,其公式如下:

$$L = \frac{1}{N} \sum_{i=1}^N L_i = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^M y_{ic} \log p_{ic}, \quad (5)$$

式中: M 代表样本的类别数量; y_{ic} 代表符号函数(样本 i 的真实类别等于类别 c 则取1,否则取0); p_{ic} 表示待检测的样本属于类别 c 的概率。BCE损失主要是为了解决梯度消失的问题,但是对于难易样本不平衡的情况来说效果并不好,而困难样本正是造成网络训练效果不佳的因素之一。

Varifocal损失是在 focal 损失的基础上修改得到的,focal 损失^[29]的公式如下:

$$L_{FL}(p, y) = \begin{cases} -\alpha(1-p)^{\gamma} \log p, & y = 1 \\ -(1-\alpha)p^{\gamma} \log(1-p), & \text{otherwise} \end{cases}, \quad (6)$$

式中: $y \in \{\pm 1\}$,为 ground truth 类; $p \in [0, 1]$,为前景类的预测概率。如式(6)所示,调制因子 $[(1-p)$ 为前景类和 p 为背景类]可以减少简单样例的损失贡献,相对增加误分类样例的重要性。因此,focal 损失可以防止训练过程中大量的简单样本淹没检测器,并将检测器聚焦在困难的例子上。Varifocal 损失^[30]借鉴了 focal 损失和 BCE 损失,其定义为

$$L_{VFL}(p, q) = \begin{cases} -q[q \log p + (1-q) \log(1-p)], & q > 0 \\ -\alpha p^{\gamma} \log(1-p), & q = 0 \end{cases}, \quad (7)$$

式中: p 为预测的交并比感知分类分数(IACS); q 为目标分数。对于前景点将其 ground truth 类 q 设为生成的边界框和它的 ground truth(gt_IoU)之间的 IoU,否则为0;而对于背景点,所有类的目标 q 为0。正样本相对于负样本来说是罕见的,所以 Varifocal 损失通过使用 p 的因子缩放损失,仅减少了负例($q=0$)的损失贡献,而以不同的方式降低正例($q>0$)的权重从而保留正样本的学习信息。

3 数据集及实验结果

3.1 数据集

数据集原始图像为在工厂实地采用型号为 MV-CE060-10UM 的工业相机,以面光源打光的方式拍摄得到的。经过数据集扩增的方式一共得到 1250 张缺陷齿轮的图片。由于缺陷的特征在常规的打光方案下和齿轮的背景部分的差异性不大,因此在进行数据集的标注之前先对图片进行对比度增强处理。利用 LabelImg 对图片进行标注,缺陷的名称(gear collision)和位置($x_{min}, y_{min}, x_{max}, y_{max}$)等信息保存在对应的 xml 文件中。为了在训练过程中实时地检验模型的训练效果,将数据集按照 9:1 的比例划分。

3.2 实验设置

本实验采用的 YOLOx 代码来自 Github, 采用 Python 语言编写。操作系统为 Windows 10, CPU 为 Intel (R) Core (TM) i5-10400F, GPU 选用 NVIDIA GeForce RTX 2080Ti, 各种环境包版本分别为 CUDA11.2、PyTorch1.7.1、Python3.8。网络的初始 5 个 epoch 采用 warmup 方法修改网络的学习率, batch_size 为 14, 训练轮数为 300, 网络的输入为 640×640 。

3.3 衡量指标

选择的衡量指标是精确率 (P)、召回率 (R)、平均精确率 (mAP)。精确率是模型在检测时的准确性, 召回率是模型在检测时的全面性, mAP 根据召回率和精确率在不同的 IoU 值下计算得到, 本研究选择 IoU 为 0.5 时的值。各个评价指标的计算公式如下:

$$R_{\text{IoU}} = \frac{A \cap B}{A \cup B}, \quad (8)$$

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (9)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (10)$$

$$R_{\text{AP}} = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{\text{inter}}(r_i + 1), \quad (11)$$

$$R_{\text{mAP}} = \frac{\sum_{i=1}^k R_{\text{AP}_i}}{k}, \quad (12)$$

式中: A 代表真实框; B 代表预测框; $A \cap B$ 代表 A 和 B 交集的面积; $A \cup B$ 代表 A 和 B 并集的面积; N_{TP} 代表预测为正类的正类数量; N_{FP} 代表将负类判定为正类的数量; N_{FN} 代表将正类判定为负类的数量; r_1, r_2, \dots, r_n 是按照升序排列的精确率插值段第一个插值出对应的召回率的值; k 代表类别数, 取 1。

3.4 训练结果

原版网络和改进后的 YOLOx 在相同的条件下分别训练 300 epoch。图 10 和图 11 分别为原版 YOLOx 和改进 YOLOx 训练过程中损失函数值的变化曲线。其中, 横坐标为训练的步数, 其计算方法如下:

$$N_{\text{total_step}} = \frac{N_{\text{training_dataset}}}{N_{\text{batch_size}}} \times N_{\text{epoch}}, \quad (13)$$

式中: $N_{\text{training_dataset}}$ 代表训练集中图片的数量。可以看到, 改进后的总的损失函数值相比未改进的有明显下降。置信度损失函数值在修改之后的数值已经远低于 1, 而修改之前的数值在 1 左右, 说明改进的函数是有效的。图 12 为改进前后 YOLOx 的 mAP@0.5 值的对比曲线。可以看到, 在 100~300 epoch 两个模型逐渐趋于收敛并且没有出现拟合的现象, 虽然在 100 epoch 前改进后网络的值较低, 但是最后改进后网络的值已经远超未改进的, 证明修改的模型是有效的。

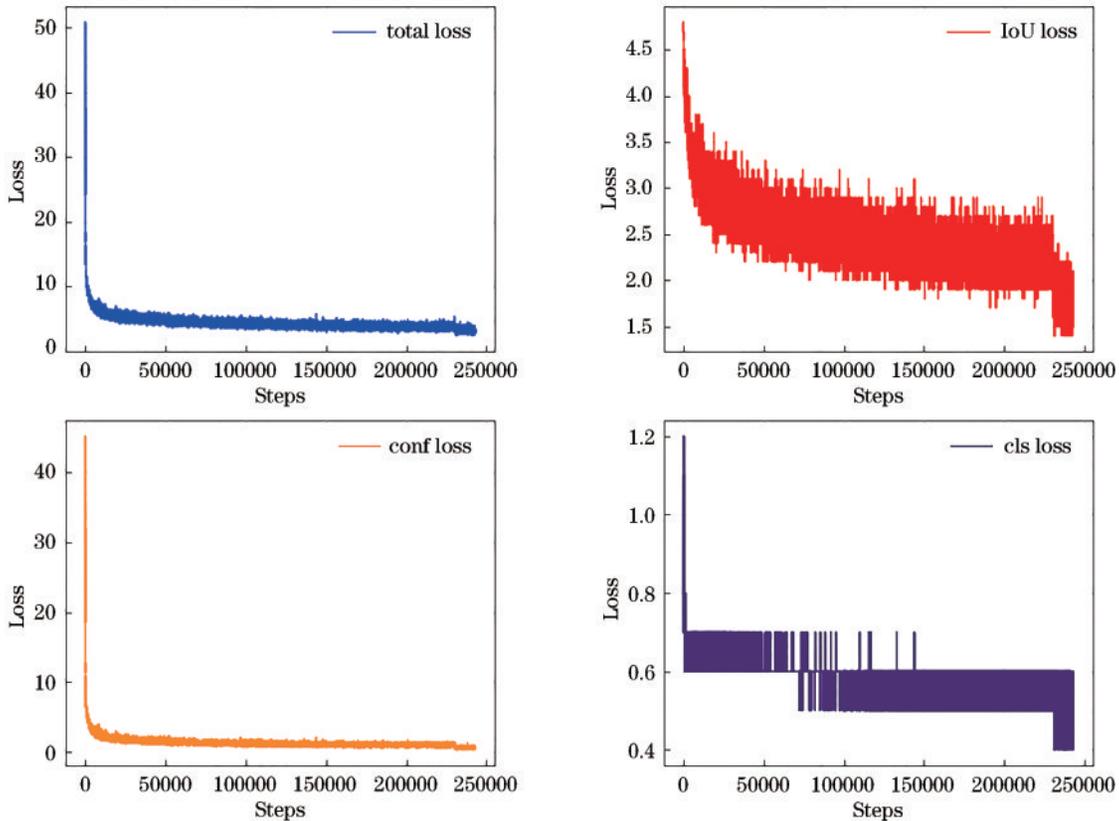


图 10 原版 YOLOx 训练过程中损失函数值变化曲线

Fig. 10 Change curves of loss function value during original YOLOx training

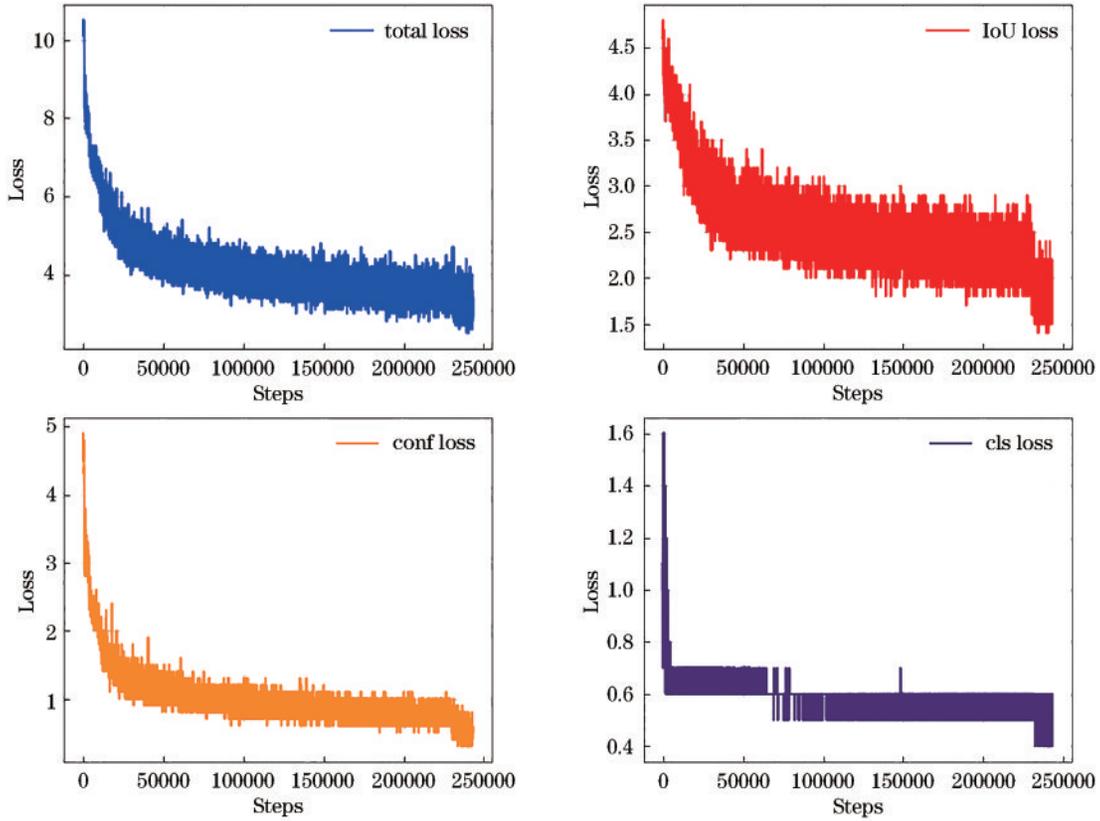


图 11 改进后的 YOLOx 训练过程中损失函数值变化曲线

Fig. 11 Change curves of the loss function value during the training of the improved YOLOx

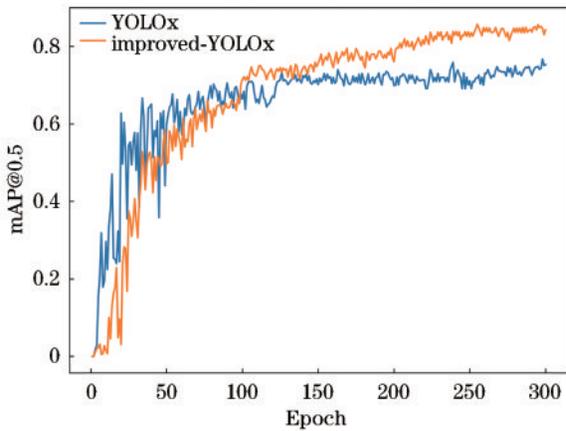


图 12 原版网络和改进后的 YOLOx 网络的 mAP@0.5 曲线对比
Fig. 12 Comparison of the mAP@0.5 curves by the improved and original YOLOx network

3.5 消融实验及对比试验

设置消融实验以验证在 YOLOx 中添加某个模块对于网络整体性能的影响。具体实验结果见表 1。从表中可以看到,未修改网络的召回率为 75.1%,精确率为 78.3%,mAP@0.5 的值为 74.2%,而改进后的网络的各项指标相比原版来说均有明显增加,其中, YOLOx+ASFF+Varifocal Loss+ECA 模型的召回率提高了 6.1 个百分点,精确率提高了 4.6 个百分点, mAP@0.5 提高了 9.4 个百分点。而精确率数值虽然

表 1 消融实验对比表

Module	Recall	Precision	mAP@0.5
YOLOx	75.1	78.3	74.2
YOLOx+ASFF	70.1	87.3	76.1
YOLOx+ECA	78.7	84.2	77.7
YOLOx+Varifocal loss	76.7	85.8	77.4
YOLOx+ASFF+ECA	78.1	86.6	80.7
YOLOx+ECA+Varifocal loss	75.6	86.1	78.0
YOLOx+ASFF+Varifocal loss	75.6	85.6	76.6
YOLOx+ASFF+Varifocal loss+ECA	81.2	82.9	83.6

不是所有模型里最高的一个,但是召回率数值是所有里最高的,说明修改后的模型检测的全面性是最好的,能够把一些不明显的缺陷也检测出来,而 mAP@0.5 的数值是最高的,说明最终的网络的综合性能是最好的。因此,改进后的 YOLOx 网络相比原版网络以及其他的改进模型来说是更有效的。

进一步将改进 YOLOx 算法和原版 YOLOx、YOLOv5s、YOLOv5m、SSD 以及 Faster R-CNN 算法进行比较,在相同的条件下利用本研究制作的数据集训练以上的算法,结果如表 2 所示。

从表 2 可以看出:改进 YOLOx 算法的召回率和 mAP@0.5 分别为 81.2% 和 83.6%,相比其他算法均

表 2 与相关算法的对比结果
Table 2 Comparison of different algorithms

Algorithm	Recall / %	Precision / %	mAP@0.5 / %	FPS
YOLOx	75.1	78.3	74.2	45
YOLOv5s	63.6	93.3	74.3	33
YOLOv5m	72.7	88.9	76.7	19
SSD	70.2	91.6	75.8	13
Faster R-CNN	77.5	71.7	77.1	15
Improved YOLOx	81.2	82.9	83.6	34

有明显差异,说明改进算法的性能优于其他算法。而精确率相比其他算法有所下降是因为召回率高更容易检测出困难样本,但是困难样本的置信度值小于 IoU 阈值而被认为是非样本,因此根据式(9)来看精确率的数值有所下降,但是在实际的过程中仍被认为是缺陷的。此外,还添加了FPS作为新的评价指标,用以反映网络的检测速度。可以看到,虽然改进 YOLOx 算

法的FPS值比未改进的下降了11,但仍能满足实际的检测需求。

3.6 检测结果

图 13 为部分齿轮检测效果对比图,其中,图 13(a)是原版网络的检测结果,图 13(b)是改进网络的检测结果。从图 13(b1)、(b2)可以看出,改进后的网络解决了原版网络中所出现的漏检问题。相比原版网络在环境干扰项的影响下所出现的误检,如图 13(a3)、(a4)所示,改进 YOLOx 网络基本解决了这个问题,如图 13(b3)、(b4),证明了改进后的网络相比原版网络是有效的。另外,所提算法已在实际齿轮生产制造企业中进行部署应用,现场检测系统如图 14 所示,其中,图 14(a)为一号检测工位的布置图,图 14(b)为二号检测工位的布置图,目前该系统运行良好,表 3 为企业生产过程中一周的实际检测情况。表中的检测准确率为检测正确数量与检测总数之比。从一周的检测数据来看,检测算法的准确率维持在 95% 左右,能够符合企业实际的检测需求。

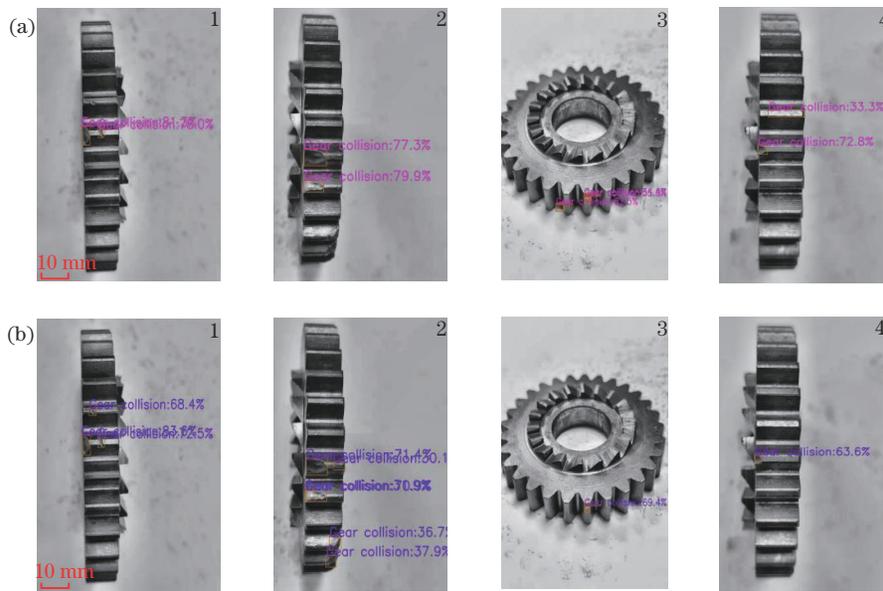


图 13 原版和改进 YOLOx 网络的齿轮检测效果对比图。(a)原版 YOLOx; (b) 改进 YOLOx

Fig. 13 Comparison of the metallic gear defect detection results by the original and improved YOLOx networks. (a) Original YOLOx; (b) improved YOLOx

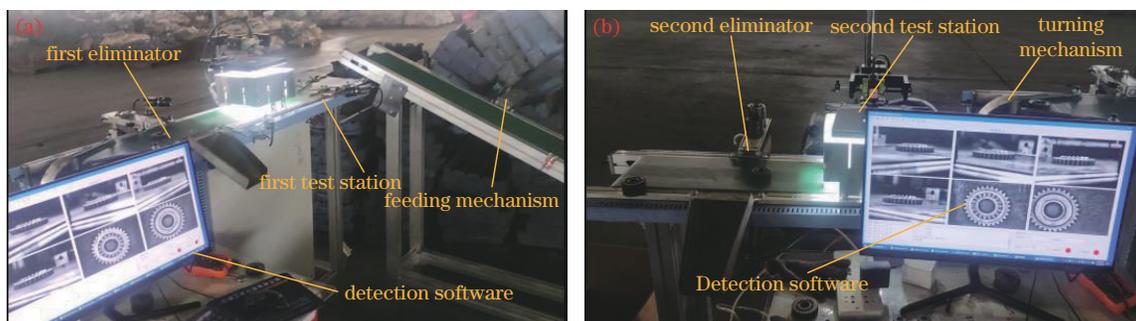


图 14 齿轮表面缺陷检测现场。(a)一号检测工位; (b)二号检测工位

Fig. 14 Detection site of gear surface defects. (a) The first detection station; (b) the second detection station

表 3 检测系统运行情况表

Table 3 Operating conditions of the detection system

Data	Total number	Missing quantity	Returned quantity	Misjudgment	Accuracy /%
2022-06-28	4960	12	265	16	95.0
2022-06-29	5700	15	235	13	95.5
2022-06-30	4000	10	130	8	95.3
2022-07-01	5000	7	210	15	94.6
2022-07-02	6000	12	183	15	95.3
2022-07-03	5500	18	272	18	94.4
2022-07-04	7000	26	346	19	95.6

4 结 论

为了解决工业干扰环境下金属齿轮表面缺陷自动化检测所面临的误检和漏检问题,提出一种改进的 YOLOx 算法,并将其与其他经典算法的检测效果进行对比。1)提出一种结合特征融合和注意力机制的 YOLOx 网络结构,提高了网络提取特征的能力,解决了有干扰情况下齿轮表面缺陷检测难题;2)与原始的 YOLOx 网络相比,改进后的 YOLOx 网络在召回率、精确率和 mAP@0.5 上均有提升,相比原版分别提升 6.1 个百分点,4.6 个百分点和 9.4 个百分点;3)与经典的目标检测网络相比,改进后的 YOLOx 网络能够解决由于油污、阴影等干扰项的影响而产生的误检和漏检,可以用于实际检测。

参 考 文 献

- [1] Wu J T, Yang Y, Wang P, et al. A novel method for gear crack fault diagnosis using improved analytical-FE and strain measurement[J]. *Measurement*, 2020, 163: 107936.
- [2] Li B, Zhang P L, Wang Z J, et al. Gear fault detection using multi-scale morphological filters[J]. *Measurement*, 2011, 44(10): 2078-2089.
- [3] Li R, Cui X L, Zhao Y, et al. A detection method of gear micro-cracks based on tuned resonance[J]. *Measurement*, 2021, 186: 110112.
- [4] Amarnath M, Praveen Krishna I R. Local fault detection in helical gears via vibration and acoustic signals using EMD based statistical parameter analysis[J]. *Measurement*, 2014, 58: 154-164.
- [5] Gadelmawla E S. Computer vision algorithms for measurement and inspection of spur gears[J]. *Measurement*, 2011, 44(9): 1669-1678.
- [6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [7] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 1440-1448.
- [8] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [10] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [11] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2022-08-09]. <https://arxiv.org/abs/1804.02767>.
- [12] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2022-11-09]. <https://arxiv.org/abs/2004.10934>.
- [13] Wang G H, Zheng H, Zhang X. A robust checkerboard corner detection method for camera calibration based on improved YOLOX[J]. *Frontiers in Physics*, 2022, 9: 819019.
- [14] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[EB/OL]. (2015-12-08)[2022-11-09]. <https://arxiv.org/abs/1512.02325>.
- [15] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23)[2022-11-09]. <https://arxiv.org/abs/1701.06659>.
- [16] Zheng L W, Fu C M, Zhao Y. Extend the shallow part of single shot multibox detector via convolutional neural network[J]. *Proceedings of SPIE*, 2018, 10806: 1080613.
- [17] Cui L S, Ma R, Lü P, et al. MDSSD: multi-scale deconvolutional single shot detector for small objects[J]. *Science China Information Sciences*, 2020, 63(2): 120113.
- [18] 董乙杉, 李兆鑫, 郭靖圆, 等. 一种改进 YOLOv5 的 X 光违禁品检测模型[J]. *激光与光电子学进展*, 2023, 60(4): 0415005.
Dong Y S, Li Z X, Guo J Y, et al. An improved YOLOv5 model for X-ray prohibited items detection[J]. *Laser & Optoelectronics Progress*, 2023, 60(4): 0415005.
- [19] 郎磊, 刘宽, 王东. 基于 YOLOX-Tiny 的轻量级遥感图像目标检测模型[J]. *激光与光电子学进展*, 2023, 60(2): 0228004.
Lang L, Liu K, Wang D. Lightweight remote sensing object detector based on YOLOX-Tiny[J]. *Laser & Optoelectronics Progress*, 2023, 60(2): 0228004.
- [20] 郭岚, 刘正新. 基于改进 YOLOv5 的光伏组件缺陷检测[J]. *激光与光电子学进展*, 2023, 60(20): 2015005.
Guo L, Liu Z X. Defect detection of photovoltaic modules based on improved YOLOv5[J]. *Laser &*

- Optoelectronics Progress, 2023, 60(20): 2015005.
- [21] 袁田, 赵轩, 刘瑞, 等. 基于自然驾驶数据的城市交叉口纵向驾驶特征分析[J]. 汽车工程, 2022, 44(6): 821-830.
Yuan T, Zhao X, Liu R, et al. An analysis on longitudinal driving characteristics in urban intersection based on natural driving data[J]. Automotive Engineering, 2022, 44(6): 821-830.
- [22] 刘子龙, 沈祥飞. 融合 Lite-HRNet 的 Yolo v5 双模态自动驾驶小目标检测方法[J]. 汽车工程, 2022, 44(10): 1511-1520, 1536.
Liu Z L, Shen X F. Small target detection method for dual-modal autonomous driving with Yolo v5 and lite-HRNet fusion[J]. Automotive Engineering, 2022, 44(10): 1511-1520, 1536.
- [23] 李亚东, 马行, 穆春阳, 等. 改进 YOLOX 网络的轴承缺陷小目标检测方法[J]. 计算机工程与应用, 2023, 59(1): 100-107.
Li Y D, Ma X, Mu C Y, et al. Improved small target detection method of bearing defects in YOLOX network [J]. Computer Engineering and Applications, 2023, 59(1): 100-107.
- [24] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [25] Wang W H, Xie E Z, Song X G, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 8439-8448.
- [26] Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-21) [2022-11-09]. <https://arxiv.org/abs/1911.09516>.
- [27] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [28] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [29] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [30] Zhang H Y, Wang Y, Dayoub F, et al. VarifocalNet: an IoU-aware dense object detector[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 8510-8519.