

改进 YOLOv5 的轻量级带钢表面缺陷检测^①



张政超

(上海理工大学 机械工程学院, 上海 200093)

通信作者: 张政超, E-mail: 213341805@st.usst.edu.cn

摘要: 针对工业场景下设备资源有限的情况, 提出一种改进 YOLOv5 的轻量化带钢缺陷检测模型. 首先, 使用 ShuffleNetv2 代替主干特征提取网络, 优化模型参数量和运行速度; 其次, 采用轻量级上采样算子 CARAFE (content-aware reassembly of features), 在增大感受野的同时进一步降低参数和计算量; 同时引入 GSConv 层, 在保证语义信息的同时平衡模型准确性与检测速度; 最后, 设计一种跨层级特征融合机制, 提高网络的检测精度. 实验结果表明, 改进后的模型的平均检测精度为 78.5%, 相较于原始 YOLOv5 算法提升了 1.4%; 模型计算量为 10.9 GFLOPs, 参数量为 5.88×10^6 , 计算量和参数量分别降低 31% 和 15.4%; 检测速度为 49 f/s, 提升了 3.5 f/s. 因此, 改进后的模型提高了检测精度和检测速度, 并且大幅降低了模型的计算量和参数量, 能够满足对带钢表面缺陷进行实时检测.

关键词: YOLOv5; 轻量化; 特征融合; 上采样; 缺陷检测

引用格式: 张政超. 改进 YOLOv5 的轻量级带钢表面缺陷检测. 计算机系统应用, 2023, 32(6): 278-285. <http://www.c-s-a.org.cn/1003-3254/9162.html>

Lightweight Strip Steel Defect Detection Based on Improved YOLOv5

ZHANG Zheng-Chao

(School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: For limited equipment resources in industrial scenarios, a lightweight strip steel defect detection model based on improved YOLOv5 is proposed. First, ShuffleNetv2 is used to replace the backbone feature extraction network to optimize model parameter amount and running speed; secondly, the lightweight up-sampling operator, namely content-aware reassembly of features (CARAFE) is used to further reduce parameters and calculation amount while increasing the receptive field. At the same time, the GSConv layer is introduced to balance the model accuracy and detection speed while ensuring semantic information. Finally, a cross-level feature fusion mechanism is designed to improve the detection accuracy of the network. The experimental results show that the mean average precision of the improved model is 78.5%, which is 1.4% higher than the original YOLOv5 algorithm. The calculation amount of the model is 10.9 GFLOPs; the parameter amount is 5.88×10^6 ; the calculation and parameter amounts are reduced by 31% and 15.4%, respectively; the detection speed is 49 f/s, which is increased by 3.5 f/s. Therefore, the improved model improves the detection accuracy and speed and greatly reduces the calculation and parameter amounts of the model, which can ensure the real-time detection of surface defects of strip steel.

Key words: YOLOv5; lightweight; feature fusion; up-sampling; defect detection

视觉目标检测是计算机视觉的经典任务, 旨在定位图像中存在物体的位置并识别物体的具体类别. 目

标检测是许多计算机视觉任务及相关应用的基础与前提, 直接决定相关视觉任务及应用的性能好坏^[1]. 作为

^① 收稿时间: 2022-12-05; 修改时间: 2023-01-06; 采用时间: 2023-02-13; csa 在线出版时间: 2023-04-20
CNKI 网络首发时间: 2023-04-21

钢铁工业的主要产品形式之一,带钢已成为汽车、机械制造、化工、航空航天和造船等行业不可缺少的原材料,其表面质量的优劣将直接影响其最终产品的性能和质量^[2]。在生产带钢的过程中,受结构及制造工艺等影响,会导致带钢表面出现划痕、裂纹、斑块、夹杂、麻点以及压入氧化铁皮等缺陷,这类缺陷不仅会对产品的外观产生影响,更会降低钢材耐磨性、耐腐蚀性、抗疲劳性和抗氧化性等性能。因此,利用计算机视觉对带钢表面缺陷进行检测对于提高带钢表面质量是十分重要的。

随着计算机视觉与深度学习的快速发展,基于卷积神经网络的目标检测模型在缺陷检测任务中发挥着重要的作用。现阶段的目标检测算法以结构为划分方式主要可以分为两类,一类是以 R-CNN 算法^[3]、Fast R-CNN 算法^[4]、Faster R-CNN 算法^[5]和 Mask R-CNN 算法^[6]为代表的两阶段目标检测算法;另一类是以 SSD 算法^[7]和 YOLO 系列算法^[8-10]为代表的一阶段目标检测算法。对于两阶段的目标检测网络,主要训练两个部分,第 1 步是训练 RPN (region proposal network) 网络,第 2 步是训练目标区域检测的网络。网络的准确度高、速度相对较慢;对于一阶段的目标检测网络,直接通过主干网络给出类别和位置信息,没有使用 RPN 网络。这样的算法速度更快,但是精度相对略低。

近年来,带钢表面缺陷检测算法逐渐被深度学习算法替代。在图像分类领域,Fu 等^[11]提出了一种端到端的卷积神经网络,对带钢表面缺陷进行分类;He 等^[12]利用对抗神经网络生成大量未标注数据,解决了缺陷样本不足的问题。而在目标检测领域,He 等^[13]以 ResNet 和 RPN 为模型,建立了一个带钢缺陷检测系统;Lv 等^[14]等提出了一种新型端到端缺陷检测网络,探讨了改进的钢材表面缺陷检测与分类方法;代小红等^[15]运用图像增强与多层特征融合技术,提出了一种基于改进 Faster R-CNN 的带钢缺陷检测算法;翁玉尚等^[16]使用 K-means II 聚类算法,并调整网络结构,采用改进后的 Mask R-CNN 算法对带钢表面进行检测;李维刚等^[17]利用加权 K-means 聚类算法对 YOLOv3 算法进行改进,提高先验框与特征图层的匹配度;孙泽强等^[18]对 YOLOv5 算法运用频域注意力机制提取目标信息,并采用解耦头将分类任务与回归任务分离,对带钢缺陷目标进行更精确检测。目前针对带钢表面缺陷检测的算法有很多,但大多存在模型体积大、参数量多的问题。

本文为了进一步提高缺陷检测的准确率,并在满足准确率高与模型参数量少的情况下,基于 YOLOv5 算法进行轻量化改进,以此实现在工业化场景下模型的部署能力。首先将 YOLOv5 主干网络改进为 Shuffle-Netv2^[19]网络,在满足检测准确率的同时降低模型参数。然后通改进 YOLOv5 原有网络结构,加入了 CARAFE^[20]上采样算子和更换 GSConv^[21]。最后针对 YOLOv5 骨干网络与颈部网络的特征融合进行修改,设计一种跨层级的多级特征融合机制,提升目标检测性能。本文提出的算法在 NEU-DET 数据集上进行测试,实验结果表明,该方法具有良好的检测性能,满足工业部署需求。

1 YOLOv5 算法

作为一阶段端到端检测算法的代表作,YOLOv5 是在 YOLOv4 基础上进一步优化改进得到的目标检测算法。YOLOv5 算法针对子模块的深度和宽度分别设计了 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 五种不同的模型,这五种模型检测精度与模型大小依次提升。YOLOv5s 很好地将参数量与准确性进行了平衡,所以本文选择以 YOLOv5s 为基础进行带钢缺陷检测。

YOLOv5 的检测流程主要分为 4 个部分,分别为输入端、Backbone 特征提取端、Neck 特征融合端以及 Head 输出端。YOLOv5 结构如图 1 所示。

在图片输入网络前,会在输入端进行 mosaic 数据增强、自适应锚框计算和自适应图片缩放操作,对数据进行预处理。

在 Backbone 主干网络部分,相较于早期版本,新版 YOLOv5 把网络的第 1 层 Focus 模块换成了一个 6×6 大小的卷积层,两者在理论上是等价的,但是对于某些 GPU 设备,新模块相较于 Focus 模块更加高效;C3 模块替换了早期的 BottleneckCSP 模块,通过残差连接实现特征融合并减小计算量;SPFP 模块采用多个小尺寸池化核级联代替 SPP 模块中单个大尺寸池化核,在保留原有功能的前提下进一步提高运行速度。

在 Neck 特征融合部分,借鉴 FPN (feature pyramid network) 与 PANet 的思想,将深层的语义信息与浅层的位置信息相互传递,增强多个尺度的信息融合。

在 Head 网络输出部分,接收 Neck 传递的不同尺度的特征图,采用非极大值抑制筛选目标框,形成 3 个预测层,用来预测和回归目标。

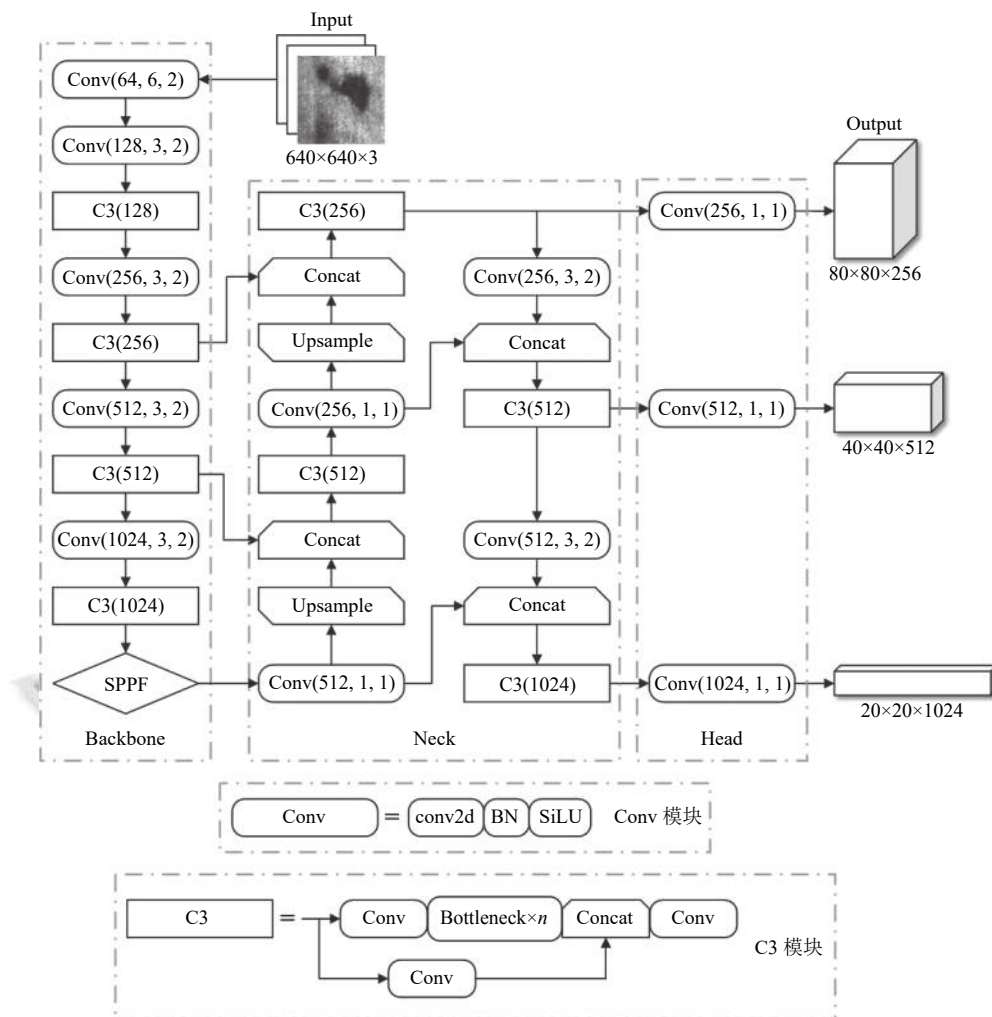


图1 YOLOv5 结构

2 模型改进与优化

2.1 轻量化特征提取网络

ShuffleNet^[21]是由旷视科技提出的一种轻量级的特征提取网络.当前主流的轻量化网络主要使用深度可分离卷积(depthwise separable convolution)和分组卷积(group convolution),以降低网络的计算量,但这两种卷积操作没有办法改变特征通道数,因此还需要1x1卷积改变通道维度.在轻量化网络中大量的1x1卷积在一定程度上降低了模型的精度,为了进一步降低模型计算量,ShuffleNet1提出了通道混洗(channel shuffle)的思想,通过通道混洗完成通道间信息的融合.通道混洗操作如图2所示.

在ShuffleNet1中采用了大量的分组卷积与逐点卷积,这会导致模型速度变慢,于是在ShuffleNet2中提出了一种新的通道分割(channel split)操作.图3所

示为ShuffleNet2的基本单元.

具体操作是将输入特征分为两部分,一部分进行深度可分离计算,再将计算结果与另一部分进行concat操作,最后进行通道混洗,完成信息的交流.

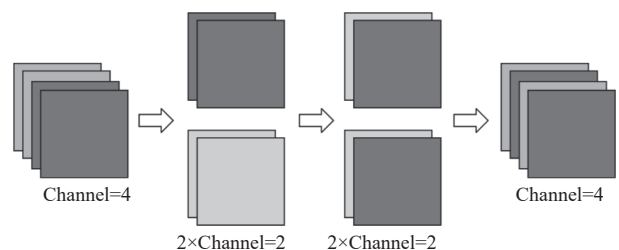


图2 通道混洗

为了满足检测准确性与模型轻量化,本文将YOLOv5的主干网络替换为ShuffleNet2的两个基本单元.首

先在网络第一层增加 CBRM 结构,对输入图像做一次卷积、批量归一化、最大池化和激活操作,然后将单元 1 与单元 2 交替堆叠,形成新的主干网络结构,在遵循 ShuffleNetv2 提出的 4 条准则的前提下,保证模型的泛化能力并降低模型参数。

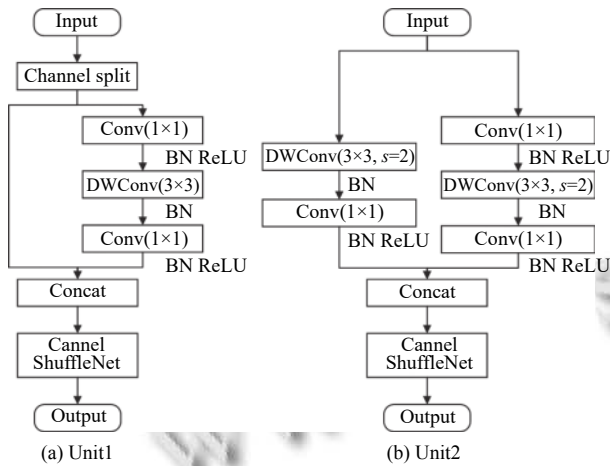


图 3 ShuffleNetv2 基本单元

2.2 GSConv 卷积层

大型深度学习模型很难部署在工业场景下的嵌入式设备上,并且大多轻量级网络使用大量的深度可分离卷积,无法达到足够的准确性,即使将主干网络替换成 ShuffleNetv2,但依然存在大量的 1×1 卷积操作,这种密集卷积的操作反而会占用更多资源,即使经过了通道混洗操作,效果仍然不好.因此本文将 GSConv 模块嵌入到特征融合阶段,显著降低模型参数量.本文没有将 GSConv 模块引入主干网络,这是因为它会加深网络层数,不利于空间信息流动,进而影响推理速度. GSConv 模块原理如图 4 所示。

GSConv 模块主要由 Conv 模块、DWConv 模块、Concat 模块和通道混洗模块组成.假设输入特征图通道数为 c_1 ,期望特征图通道数为 c_2 ,首先对输入特征图进行标准卷积操作,得到通道数为 $c_2/2$ 的特征图;接下来对它进行深度可分离卷积操作,得到另一个通道数为 $c_2/2$ 的特征图;然后将两个特征图进行 concat 拼接,得到通道数为 c_2 的拼接特征图;最后通过通道混洗操作输出目标通道数的特征图。

GSConv 模块采用深度可分离卷积结合标准卷积的形式,可以通过深度可分离卷积降低计算复杂度,同时通过标准卷积缓解由于深度可分离卷积的特征提取

和融合能力弱而造成识别精度低的问题。

2.3 CARAFE 上采样算子

为了将输入特征图的输出尺寸与原图进行匹配,同时实现不同尺度的特征融合,上采样操作在目标检测领域应用十分广泛.但常用的上采样操作存在着一些问题,比如仅依靠像素点的空间位置信息决定上采样核,容易丢失重要的特征信息.并且在预测上采样核的过程中,会引入大量的计算量与参数量,这对网络的轻量化设计有着不利的影响.因此本文引入 CARAFE 轻量级上采样算子,在不引入较多参数和计算量的前提下,增大感受野,并将输入信息与特征图的语义信息进行关联,实现比主流上采样操作更好的性能. CARAFE 原理图如图 5 所示。

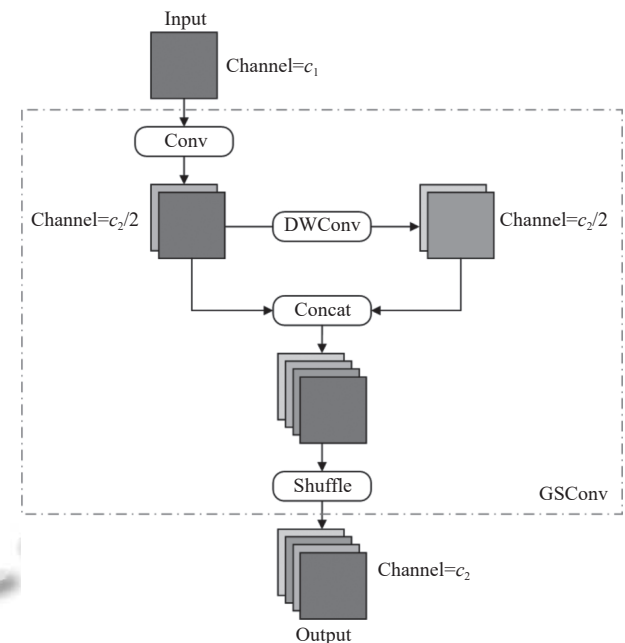


图 4 GSConv 模块

CARAFE 将上采样分成两个模块,分别是上采样核预测模块与特征重组模块.在上采样预测模块中,对于 $H \times W \times C$ 的输入图像,首先使用 1×1 卷积对输入的特征图通道压缩到 $H \times W \times C_m$,减小计算量;然后为了对卷积进行编码,对通道数量进行重新分配,其中, σ 是上采样系数;接下来进行像素重组,扩大上采样的接受域;最后归一化处理,减少操作中的参数量.在特征重组模块中,使用上采样核预测模块得到的特征图和普通上采样得到的特征图进行点乘,得到输出值.两个模块的公式如式 (1) 所示。

$$\begin{cases} K_{\text{encoder}} = k_{\text{up}} - 2 \\ W_0 = \psi(N(X_1, K_{\text{encoder}})) \\ Z_0 = \varphi(N(X_1, k_{\text{up}}), W_0) \end{cases} \quad (1)$$

其中, 每个目标位置都需要一个 $k_{\text{up}} \times k_{\text{up}}$ 的上采样核. ψ 代表上采样核预测模块, K_{encoder} 是编码卷积的卷积核, $N(X_1, k)$ 是以该位置为中心的 X 的子区域, W_0 是预测模块的输出, φ 是特征重组模块, Z_0 是上采样模型的总输出.

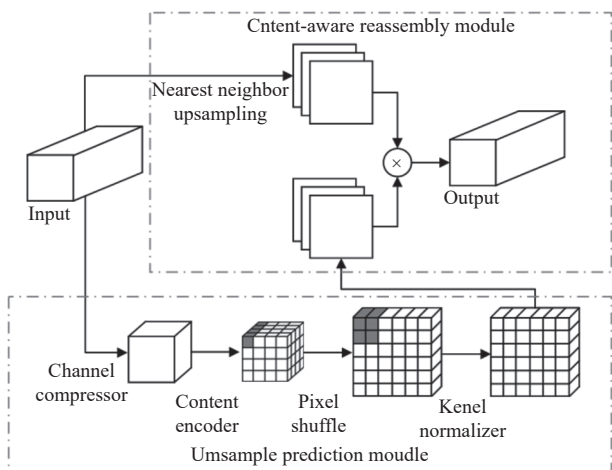


图5 CARAFE原理图

2.4 跨层级特征融合机制

随着特征提取和特征融合网络层数的加深, 信息的流通与传递会变得更加困难. 相比于其他目标检测算法, YOLOv5 采用了FPN^[23]的思想, 设计了一种自顶向下的特征融合网络, 同时兼顾了浅层特征与深层特征, 一定程度上解决了目标检测精度低的问题, 但这种结构对浅层特征的提取仍有一定的不足. 本文设计了一种新的跨层级特征融合机制, 将浅层特征与深层特征进行进一步融合. 改进后的网络结构如图6所示.

具体来说, 第1个特征融合层与主干网络第4层相连, 接收的特征图经过卷积核大小为 3×3 , 步长为2的卷积操作; 然后与特征融合网络经过GSConv操作的特征图进行concat连接, 形成通道数为512的特征图; 最后与检测头相连, 经过C3操作后输出结果. 同理, 第2个特征融合层与主干网络第2层相连, 经过类似操作后输出结果.

由于浅层特征感受野更小, 更加注重细节信息, 能够提取到更多的细节, 因此增加两个相似的特征融合结构, 起到浅层网络与深层网络全局融合的作用, 进一步增强了网络检测能力.

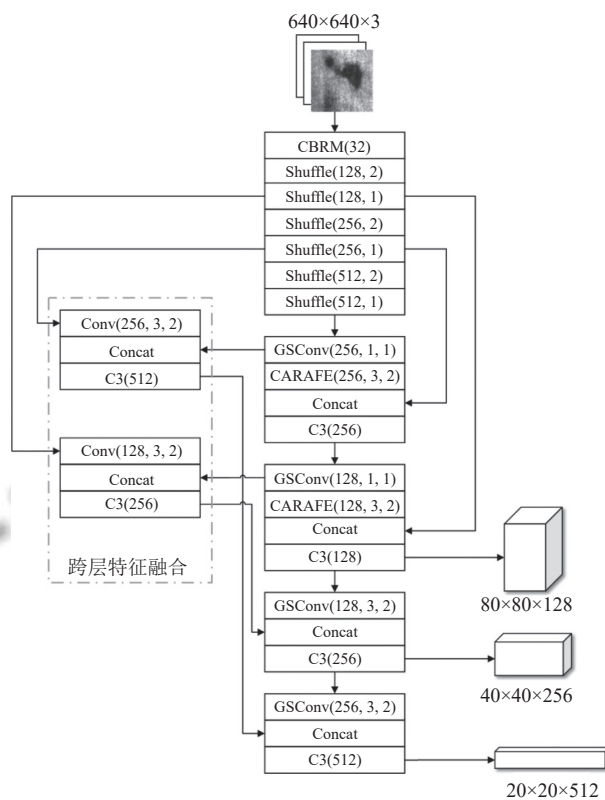


图6 改进后的YOLOv5网络结构

3 实验结果及分析

3.1 数据集

本文使用东北大学公开的带钢NEU-DET数据集作为训练测试数据. 其中, 包含了6种不同类型的表面缺陷: 划痕 (scratches, Sc)、裂纹 (crazing, Cr)、斑块 (patches, Pa)、夹杂 (inclusion, In)、麻点 (pitted surface, Ps) 以及压入氧化铁皮 (rolled-in scale, Rs). 共有1800张灰度图片, 每种缺陷各300张, 图片大小为 200×200 . 本文按照训练集、验证集和测试集8:1:1的比例抽取数据, 即训练集1440张, 验证集和测试集各180张. 各类缺陷如图7所示.

3.2 实验环境与参数设定

实验使用的环境如表1所示.

在训练过程中, 加载预训练模型“yolov5s.pt”, 设置epoch=200, batch size=64, 动量为0.937, 采用SGD算法和余弦退火学习率衰减方式进行训练.

3.3 评价指标

在目标检测领域中, 通常采用召回率 (recall, R)、精确率 (precision, P)、平均精度 (average precision,

AP) 和平均精度均值 (mean average precision, mAP) 对性能进行评定. 召回率是指正确预测为正的占全部实际为正的比例, 精确率是指正确预测为正的占全部预测为正的比例. 计算公式如式 (2) 所示.

$$\begin{cases} R = \frac{TP}{TP+FN} \\ P = \frac{TP}{TP+FP} \end{cases} \quad (2)$$

其中, TP (true positive) 表示一个实例是正类并且被认定为正类; FN (false negative) 表示一个实例是正类但被判定为假类; FP (false positive) 表示一个实例是假类但被判定为正类. 以 P 为纵坐标, R 为横坐标绘制的曲线被称为 $P-R$ 曲线, $P-R$ 曲线与坐标轴围成的面积被称为 AP . 所有类别 AP 的均值被称为 mAP , 两者越接近 1, 表示模型检测越准确.

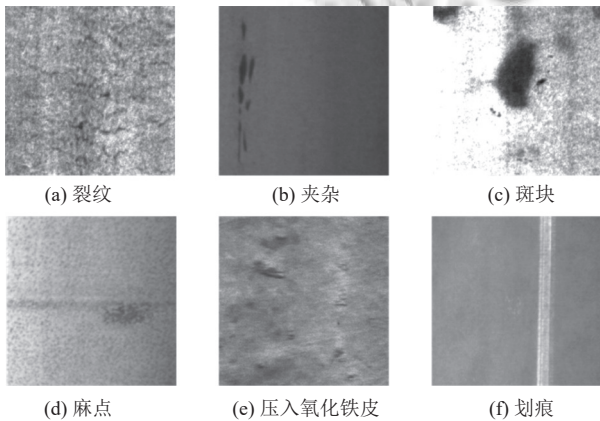


图7 带钢表面缺陷

表1 实验环境

名称	版本型号
CPU	Intel (R) Xeon (R) CPU E5-2680 v4@2.40 GHz
GPU	NVIDIA GeForce RTX 3090-24 GB
操作系统	Ubuntu 20.04
框架	PyTorch 1.11.0 + Cuda 11.3

评价目标检测模型, 除了精确度指标外, 模型复杂度也必须考虑. 通常用正向推理的浮点运算次数 (floating points of operations, FLOPs) 和参数量 (parameters) 来描述模型的复杂度. 其中, FLOPs 越小, 说明模型计算复杂度越小; parameters 越少, 模型体积越小.

本文采用 mAP 和参数量作为模型的主要评价指标, 另外采用 FLOPs 和每秒传输帧数 (frames per second, FPS) 综合衡量模型检测速度.

3.4 消融实验及结果

为了体现不同改进策略对带钢表面缺陷检测准确

性的影响, 进行消融实验, 其中, “√”表示采用此模型进行设计. 实验结果如表2所示.

表2 消融实验结果

模型	组1	组2	组3	组4	组5	组6
YOLOv5s	√	√	√	√	√	√
ShuffleNetv2	—	√	√	√	√	√
GSCConv	—	—	√	—	√	√
CARAFE	—	—	—	√	√	√
特征融合	—	—	—	—	—	√
mAP (%)	77.1	75.1	77	76.2	78	78.5
GFLOPs	16.0	8.0	7.5	8.5	7.9	10.3
FPS (f/s)	45.5	47.3	45.6	46.1	48.2	49

本实验以原始 YOLOv5s 模型为基线. 由第 2 组与第 1 组对比可知, 使用轻量化主干网络 mAP 降低到 75.1%, GFLOPs 降低到 8.0, FPS 提升到 47.3f/s, 说明轻量化网络由于参数量的降低, 会降低模型检测精度, 但模型检测速度提高. 第 2 组和第 3 组进行对比可知, 增加 GSCConv 模块后, 增强了网络的特征融合能力, mAP 提高了 1.9%, GFLOPs 降低 0.5, FPS 降低 1.7 f/s, 检测速度相比于原始 YOLOv5s 有所提高, 但检测精度仍然低于原始 YOLOv5s. 第 2 组和第 4 组对比可以发现, 更换上采样层为 CARAFE 后, 增大了感受野, 提高了语义信息的关联能力, mAP 提升了 1.1%, GFLOPs 提高 0.5, FPS 降低 1.2 f/s. 第 2 组与第 5 组对比可知, 同时引入 GSCConv 模块与 CARAFE 上采样层后, 模型结合两个模块的优点, mAP 提升了 2.9%, GFLOPs 降低 0.1, FPS 提高 0.9 f/s. 第 2 组与第 6 组对比可知, 在增加特征融合层后, 更有利于浅层信息与深层信息的融合交流, mAP 提升了 0.5%, GFLOPs 提高 2.3, FPS 提高 1.7 f/s. 综合第 1 组与第 6 组的比较结果, 改进后的模型的 mAP 、GFLOPs 和 FPS 均优于原始 YOLOv5 模型.

3.5 对比实验及结果

为了进一步验证本文算法的有效性, 将其与 SSD、Faster-RCNN、YOLOv3、YOLOX^[24]、YOLOv7^[25] 和 YOLOv5 在内的主流目标检测算法在 NEU-DET 数据集上进行对比实验. 实验结果如表3所示.

根据实验结果可以看出, 本文提出的算法 mAP 为 78.5%, FPS 为 49 f/s, 参数量为 5.88×10^6 , 与 YOLOv5 算法相比, mAP 提高 1.4%, FPS 提升 3.5 f/s, 参数量降低 1.15×10^6 . 相比于参数量很大的 Faster-RCNN、SSD、YOLOv3、YOLOX、YOLOv7 算法, mAP 都有明显提高. 除与 SSD 相比 FPS 降低外, 都高于其他算法. 结果

表明, 本文算法在实现了模型轻量化的同时, *mAP* 和 *FPS* 得到有效提升. 总体而言, 本文算法对目标具有较高的准确率, 可以实现在工业场景下的部署, 效果优于其他主流的目标检测算法.

表3 对比实验结果

算法	<i>mAP</i> (%)	<i>FPS</i> (f/s)	Params ($\times 10^6$)
Faster-RCNN	64.9	25.3	56.81
SSD	75.9	97.6	48.83
YOLOv3	74.5	37.4	61.55
YOLOX	75.8	44.1	18.01
YOLOv5	77.1	45.5	7.03
YOLOv7	77	29.1	37.20
本文算法	78.5	49	5.88

如图8所示, 为原始 YOLOv5s 与本文算法在各类缺陷上的检测结果对比情况. 其中, 标注框上标注了缺陷类型以及置信度.

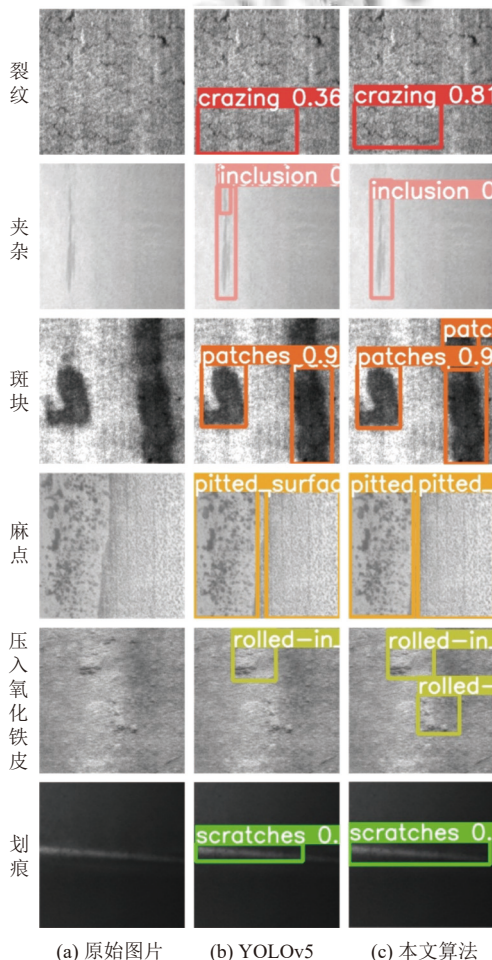


图8 检测效果图

根据图8的检测结果可以看出, 经过改进后的网络在检测相同位置的缺陷时, 置信度更高; 相较于原始

YOLOv5 算法对于部分目标检测不到的情况也有所改善.

针对带钢表面的缺陷检测问题, 前人进行了很多研究. 为了验证本文算法的综合性能, 选取部分文献与本文算法进行对比. 对比结果如表4所示.

根据对比, 在带钢表面缺陷检测领域的研究中, *mAP* 普遍偏低, 主要原因是由于带钢缺陷中的裂纹缺陷特征不明显, 并且在 NEU-DET 数据集中, 存在数据标注不完全的问题. 这些原因都会对检测时的速度和精度产生影响.

其中, 文献[18]的 *mAP* 最高, 主要原因将检测类别由6类改为5类, 删除了裂纹类别; 文献[26]的 *mAP* 为82.4%, 但参数量过多; 与文献[27]和文献[28]对比, 本文算法 *mAP* 分别提高4.4%和1.7%, 同时参数量有着明显下降. 与其他算法相比, 在检测精度方面, 本文算法 *mAP* 虽然低于文献[18]和文献[26], 但参数量最低, 满足了检测模型的轻量化要求; 而在检测速度方面, 理论上 *FPS* 高于30 f/s, 即满足实时检测的要求, 本文算法 *FPS* 虽然低于文献[26]和文献[27], 但仍可以保证检测时的流畅.

表4 与其它部分文献的对比

算法	<i>mAP</i> (%)	<i>FPS</i> (f/s)	Params ($\times 10^6$)
文献[18]	85.5	27.7	>8.1
文献[26]	82.4	65.4	15.37
文献[27]	74.1	75	23.9
文献[28]	76.8	43.3	6.26
本文	78.5	49	5.88

4 结论与展望

对于带钢的表面缺陷检测问题, 本文对 YOLOv5s 算法进行改进. 采用轻量化的 ShuffleNetv2 网络作为主干网络, 引入 CARAFE 轻量级上采样算子增大特征图感受野, 采用 GSConv 层保留语义信息, 引入多层级连的特征融合机制, 增强顶层特征与底层特征的融合. 通过实验的检测结果, 可以发现改进网络在提升检测精度的前提下, 实现了检测网络的轻量化. 但现有数据集对裂纹类缺陷的标注不完全, 测试时会将检测结果中部分正确的检测判定为假正例. 在接下来的研究中, 会对数据进行清洗, 调整标注信息, 从而进一步提升检测精度.

参考文献

- 曹家乐, 李亚利, 孙汉卿, 等. 基于深度学习的视觉目标检

- 测技术综述. 中国图象图形学报, 2022, 27(6): 1697–1722. [doi: 10.11834/jig.220069]
- 2 吴平川, 路同浚, 王炎. 带钢表面自动检测系统研究现状与展望. 钢铁, 2000, 35(6): 70–75. [doi: 10.3321/j.issn:0449-749X.2000.06.018]
 - 3 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587. [doi: 10.1109/CVPR.2014.81]
 - 4 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448. [doi: 10.1109/ICCV.2015.169]
 - 5 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
 - 6 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386–397. [doi: 10.1109/TPAMI.2018.2844175]
 - 7 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37. [doi: 10.1007/978-3-319-46448-0_2]
 - 8 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788. [doi: 10.1109/CVPR.2016.91]
 - 9 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525. [doi: 10.1109/CVPR.2017.690]
 - 10 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
 - 11 Fu GZ, Sun PZ, Zhu WB, *et al.* A deep-learning-based approach for fast and robust steel surface defects classification. Optics and Lasers in Engineering, 2019, 121: 397–405. [doi: 10.1016/j.optlaseng.2019.05.005]
 - 12 He D, Xu K, Zhou P, *et al.* Surface defect classification of steels with a new semi-supervised learning method. Optics and Lasers in Engineering, 2019, 117: 40–48. [doi: 10.1016/j.optlaseng.2019.01.011]
 - 13 He Y, Song KC, Meng QG, *et al.* An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493–1504. [doi: 10.1109/TIM.2019.2915404]
 - 14 Lv XM, Duan FJ, Jiang JJ, *et al.* Deep metallic surface defect detection: The new benchmark and detection network. Sensors, 2020, 20(6): 1562. [doi: 10.3390/s20061562]
 - 15 代小红, 陈华江, 朱超平. 一种基于改进 Faster RCNN 的金属材料工件表面缺陷检测与实现研究. 表面技术, 2020, 49(10): 362–371.
 - 16 翁玉尚, 肖金球, 夏禹. 改进 Mask R-CNN 算法的带钢表面缺陷检测. 计算机工程与应用, 2021, 57(19): 235–242. [doi: 10.3778/j.issn.1002-8331.2010-0446]
 - 17 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. 电子学报, 2020, 48(7): 1284–1292. [doi: 10.3969/j.issn.0372-2112.2020.07.006]
 - 18 孙泽强, 陈炳才, 崔晓博, 等. 融合频域注意力机制和解耦头的 YOLOv5 带钢表面缺陷检测. 计算机应用, 2023, 43(1): 242–249.
 - 19 Ma NN, Zhang XY, Zheng HT, *et al.* ShuffleNet v2: Practical guidelines for efficient CNN architecture design. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 122–138.
 - 20 Wang JQ, Chen K, Xu R, *et al.* CARAFE: Content-aware reassembly of features. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 3007–3016.
 - 21 Li HL, Li J, Wei HB, *et al.* Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. arXiv:2206.02424, 2022.
 - 22 Zhang XY, Zhou XY, Lin MX, *et al.* ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6848–6856. [doi: 10.1109/CVPR.2018.00716]
 - 23 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.
 - 24 Ge Z, Liu ST, Wang F, *et al.* YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.
 - 25 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv:2207.02696, 2022.
 - 26 马燕婷, 赵红东, 阎超, 等. 改进 YOLOv5 网络的带钢表面缺陷检测方法. 电子测量与仪器学报, 2022, 36(8): 150–157. [doi: 10.13382/j.jemi.B2205354]
 - 27 曹义亲, 伍铭林, 徐露. 基于改进 YOLOv5 算法的钢材表面缺陷检测. 图学学报. <http://kns.cnki.net/kcms/detail/10.1034.T.20221012.1820.004.html>. (2022-10-13).
 - 28 李鑫, 汪诚, 李彬, 等. 改进 YOLOv5 的钢材表面缺陷检测算法. 空军工程大学学报 (自然科学版), 2022, 23(2): 26–33.

(校对责编: 牛欣悦)