

文章编号 1004-924X(2024)08-1227-14

改进轻量化 VTG-YOLOv7-tiny 的 钢材表面缺陷检测

梁礼明, 龙鹏威*, 冯 耀, 卢宝贺

(江西理工大学 电气工程与自动化学院, 江西 赣州 341000)

摘要: 针对钢材表面缺陷形态多样、结构复杂且存在检测目标漏检现象和算法参数数量过大等问题, 提出一种轻量化 VTG-YOLOv7-tiny 的钢材缺陷检测算法。该方法一是设计 VoVGA-FPN 网络, 以减少信息传递过程中的丢失, 增强网络特征融合能力; 二是构建三重坐标注意力机制, 提升模型对空间和通道信息的特征提取能力; 三是引入鬼影混洗卷积, 在提高精度的同时降低模型参数数量和计算量; 四是增加大目标检测层, 改善特征图中部分缺陷占比较大, 导致检测精度低的问题。在 NEU-DET 和 Severstal 钢材缺陷数据集进行实验验证, 改进后算法与原模型相比, mAP 分别提升 5.7% 和 8.5%; 参数数量和计算量分别降低 0.61 M 和 4.2 G; 精确度和召回率分别提升 7.1%, 1.8% 和 8.9%, 7.0%。实验结果表明, 改进后的算法更好地平衡了检测精度和轻量化, 为边缘终端设备提供了参考。

关键词: 缺陷检测; 轻量化 YOLOv7-tiny; VoVGA-FPN 网络; 三重坐标注意力; 鬼影混洗卷积

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/OPE.20243208.1227

Improving the lightweight VTG-YOLOv7-tiny for steel surface defect detection

LIANG Liming, LONG Pengwei*, FENG Yao, LU Baohe

(School of Electrical Engineering and Automation, Jiangxi University of Science and
Technology, Ganzhou 341000, China)

* Corresponding author, E-mail: 2637018663@qq.com

Abstract: To address the problems of diverse and complex shapes of steel surface defects, detection target missing, and large number of algorithm parameters, a lightweight VTG-YOLOv7-tiny steel defect detection algorithm was proposed. The method first designed VoVGA-FPN network to reduce the loss of information during information transmission and enhance the network feature fusion ability; second, it constructed a triple coordinate attention mechanism to improve the model's feature extraction ability of spatial and channel information; third, it introduced ghost shuffle convolution to reduce the model parameters and computation while improving the accuracy; fourth, it added a large target detection layer to improve the problem that some defects in the feature map occupy a large proportion, resulting in low detection accuracy. The improved algorithm was verified on the NEU-DET and Severstal steel defect datasets. Compared with the original model, the mAP of the improved algorithm is increased by 5.7% and 8.5%, re-

收稿日期: 2023-10-16; 修订日期: 2023-12-01.

基金项目: 国家自然科学基金资助项目 (No. 51365017, No. 6146301); 江西省自然科学基金资助项目 (No. 20192BAB205084); 江西省教育厅科学技术研究重点项目 (No. GJJ170491)

spectively; the parameters and computation are reduced by 0.61 M and 4.2 G, respectively; the accuracy and recall are increased by 7.1%, 1.8% and 8.9%, 7.0%, respectively. The experimental results show that the improved algorithm better balances the detection accuracy and lightweight, and provides a reference for edge terminal devices.

Key words: defect detection; Lightweight YOLOv7-tiny; VoVGA-FPN network; Triplet Coordinate Attention(TCA); Ghost Shuffle Convolution(GSConv)

1 引 言

钢材工业在现代化社会中具有重要地位,其在建筑、制造业和交通运输等领域扮演着关键角色。然而,由于生产过程中不可避免会出现工艺偏差和材料缺陷,钢材产品可能存在各种类型的缺陷,如裂缝、突起和划痕等^[1]。如果不对其进行及时检测和处理,则容易导致严重安全隐患和产品质量问题。因此,研究一种自动识别缺陷检测算法具有重要意义。

为解决上述问题,诸多学者提出了基于传统机器学习的方法和基于深度学习的方法^[2-4]。其中传统检测方法主要包括人工视觉检查、频闪法^[5]和涡流检测法^[6]等。然而,上述传统方法存在人力成本高、操作繁琐和效率低的问题。基于深度学习检测方法通过对大量标记样本的学习,能够自动学习和提取关键特征,实现对钢材缺陷的准确检测。基于深度学习的缺陷检测算法主要包括一阶段和二阶段两种主流算法。其中二阶段检测算法比较知名的有 Faster R-CNN^[7]和 Mask R-CNN^[8]等。这类算法在准确率上具有较好的表现,但计算量和参数量较大,速度较慢。一阶段检测算法是通过直接预测目标的位置和类别来实现检测,代表性算法包括 YOLO^[9-10]系列、SSD^[11]和 RetinaNet^[12]等。这些算法具有快速和端到端的特点,能够实时检测目标,但精度有所欠缺。相比于二阶段检测算法,一阶段的 YOLO 算法能够更好地平衡检测精度与速度,但也存在对小目标缺陷难以检测、模型参数量和计算量过大等问题,不少研究者对其进行改进。Lin 等^[13]提出特征金字塔网络 (Feature Pyramid Networks, FPN), 利用自上而下的结构处理上下文信息,增加对小目标特征映射的分辨率,以获得更多有用信息,但单向传递结构存在限制信息流问题。Liu 等^[14]构建一种路径聚合网络 (Path Ag-

gregation Network, PANet), 在 FPN 的基础上增加一条自下而上的路径, 缩短信息路径、增强小目标语义特征的流动, 但自下而上的路径反而会使浅层特征的信息在传递和交互过程中丢失。Zhang 等^[15]设计一种轻量化 YOLOv5 算法, 将主干网络换成轻量级的 MobileNet v2 网络, 同时采用剪枝技术优化了模型参数量, 但以牺牲精度为代价。Ma 等^[16]提出改进 YOLOv5 的钢材表面缺陷检测算法, 采用 Transformer 与 BiFPN 结合等方法增强浅层与深层信息的融合, 较好地均衡检测数度与精度, 但参数量随之增加。Li 等^[17]通过减少 YOLO 每层的参数量与残差层构建一种轻量化模型, 但却以牺牲精度的方式换来速度的提升。Zhou 等^[18]构建一种多尺度轻量化注意力的钢材缺陷检测, 采用轻量化的注意力模块与金字塔结合等方法, 有效提高检测精度, 但参数量与原模型相比并未减少。因此, 在提升小目标精度和优化模型参数量上还有待改进。Li-an 等^[19]应用知识蒸馏和预测尺度设计一种改进 YOLOv4 的缺陷检测算法, 有效提高检测精度。Han 等^[20]提出一种高效轻量的卷积结构, 鬼影卷积 (Ghost Convolution, GConv), 有效降低参数量和计算成本。Xu 等^[21]构建一种 YOLOv5-Ghost 网络的交通识别算法, 充分利用 GConv 轻量高效的特性, 在提升精度的同时有效降低模型参数量。

以上方法在模型轻量化与检测精度之间无法做到均衡。因此本文设计一种轻量化 VTG-YOLOv7-tiny (VoVGA-FPN+TCA+GSConv) 算法用于钢材表面缺陷检测, 并实现对缺陷目标的较高精度检测, 且本文主要研究工作如下:

① 在网络特征融合部分, 设计 VoVGA-FPN 网络替换原 PANet 网络, 增强网络特征融合能力;

② 设计三重坐标注意力 (Triplet Coordinate Attention, TCA) 模块嵌入主干特征提取部分,

提升模型对空间和通道语义信息的提取;

③引入鬼影混洗卷积(Ghost Shuffle Convolution, GSConv)替换网络中的标准卷积,在提升检测精度的同时降低模型参数量与计算量;

④在原模型基础上增加一层检测层,改善部分缺陷在特征图中占比较大导致检测精度低的问题。

2 YOLOv7-tiny 模型

Wang等^[22]提出YOLOv7-tiny算法,该模型具有检测快、准确率高和参数量低的特点,故本文选用YOLOv7-tiny作为主干检测模型。YOLOv7-tiny网络主要由特征提取骨干(Backbone)、特征融合网络(Neck)和检测头(Head)三部分组成。YOLOv7-tiny在Backbone中采用ELAN替代E-ELAN的长程聚合网络层,融合卷积和池化层组成;在Neck中沿用PANet网络进

行特征聚合;在Head中使用CBL替代RepConv卷积对输出的通道数进行调整。

3 本文模型

3.1 VTG-YOLOv7-tiny 模型

针对钢材表面缺陷复杂易漏检和算法参数量大的问题,本文提出VTG-YOLOv7-tiny模型。VTG-YOLOv7-tiny首先在模型的Neck部分,设计VoVGA-FPN网络替换PANet网络,提高模型对目标特征的融合能力;其次在Backbone部分构建TCA模块,提升主干网络对空间和通道语义信息的提取能力;再次采用GSConv卷积替换标准卷积,在保证检测精度与速度的同时降低模型参数量和计算量;最后在Head部分增加检测层,以改善特征图中部分缺陷占比较大导致检测精度低的问题。其中红框部分为改进部分,改进后的模型如图1所示(彩图见期刊电子版)。

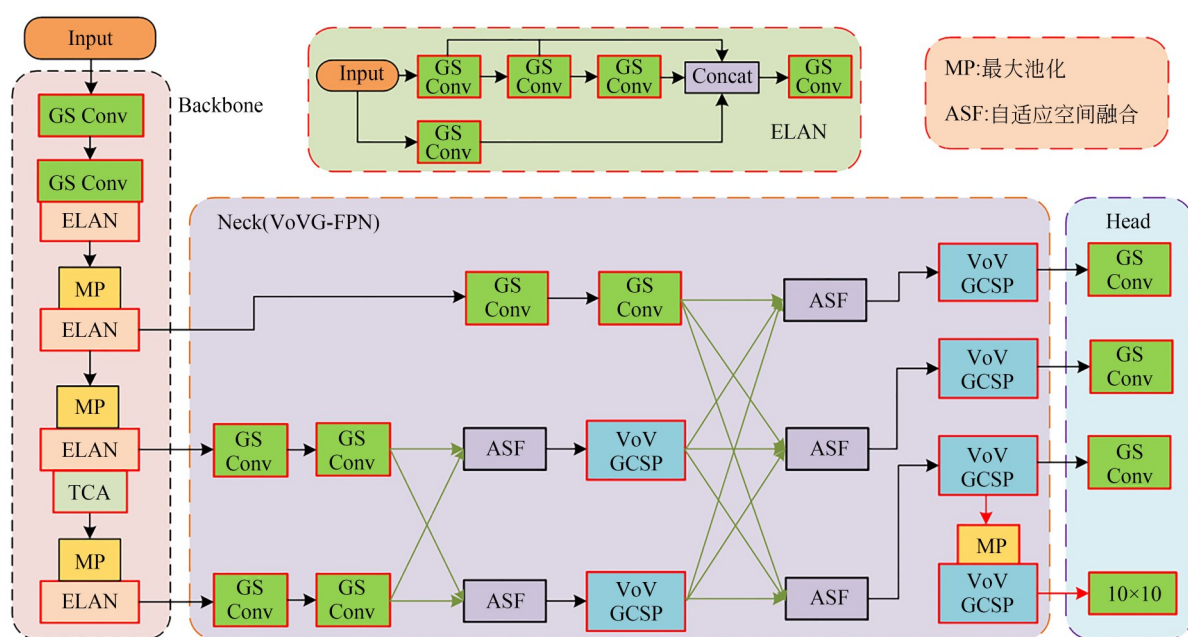


图1 VTG-YOLOv7-tiny模型结构

Fig. 1 VTG-YOLOv7-tiny model structure

3.2 VoVGA-FPN 融合网络

YOLOv7-tiny的Neck部分沿用文献[14]提出的PANet结构,该结构在FPN的基础上增加一条自下而上的路径,虽然解决FPN中单向路径结构存在限制信息流问题,但自下而上的路径反

而会使浅层特征信息在传播和交互过程中丢失。Yang等^[23]提出渐进特征金字塔(Asymptotic Feature Pyramid Network, AFPN)结构,通过将两个不同分辨率的浅层特征组合再与深层特征融合的方式来缩小非相邻层次之间的语义差距,从而

减少信息在传递过程中的丢失,但AFPN中的ELAN结构采用大量的标准卷积,容易增加网络的计算复杂性。文献[20]构建的GConv能够有效降低计算成本。本文受文献[20]和文献[23]的启发,首先采用GConv构建Gbottleneck模块,然后在Gbottleneck基础上设计VoVGCSPC模块替代AFPN中的ELAN结构,进而设计VoVGA-FPN融合网络,最后将其替换PANet网络。VoVGA-FPN网络结构如图2所示。

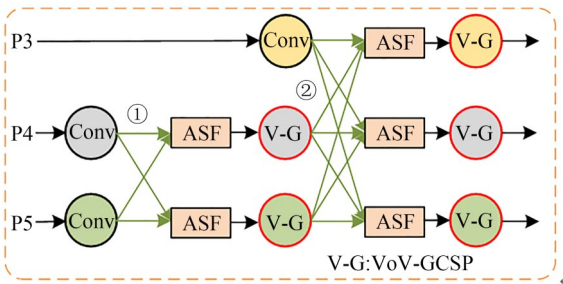


图2 VoVGA-FPN网络结构

Fig. 2 VoVGA-FPN network structure

为了避免元素求和时因不同层次之间可能存在信息矛盾,VoVGA-FPN网络通过引入自适应空间融合(Adaptively spatial fusion, ASF)操作^[24],对多级融合过程特征进行过滤。ASF由特征向量 $x_{ij}^{1 \rightarrow l}$, $x_{ij}^{2 \rightarrow l}$ 和 $x_{ij}^{3 \rightarrow l}$ 进行线性组合,其融合过程可表示为:

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

其中: y_{ij}^l 表示结果特征向量; $x_{ij}^{n \rightarrow l}$ 表示特征向量从 n 层到 l 层; α, β, γ 分别是三层特征图的权重,且 $\alpha + \beta + \gamma = 1$ 。这三个参数可以被网络自定义学习,并定义为:

$$\alpha_{ij}^l = \frac{e^{\lambda_{ij}^l}}{e^{\lambda_{ij}^l} + e^{\lambda_{ij}^l} + e^{\lambda_{ij}^l}}. \quad (2)$$

GConv结构如图3所示。与标准卷积(SC)相比,GConv卷积首先通过SC和恒等映射操作得到本征特征图;然后通过线性变换得到鬼影特征图;最后将不同的特征图进行拼接得到新的输出。该过程与SC得到的输出一致,其中 $\phi_1, \phi_2, \dots, \phi_k$ 表示 3×3 或 5×5 卷积。

计算分析:假设输入特征图和输出特征图尺寸大小为 $H \times W \times C, H' \times W' \times N$,卷积核大小为 $k \times k$ 。采用浮点操作(Floating-point Opera-

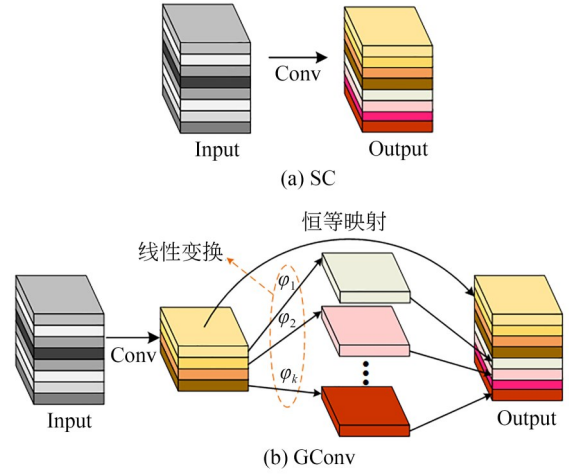


图3 GConv和SC网络结构

Fig. 3 GConv and SC network structure

tions, FLOPs)衡量复杂度,则标准卷积和鬼影卷积的计算复杂度分别为:

$$F_{SC} = H' \times W' \times N \times C \times k \times k, \quad (3)$$

$$y_1 = \frac{N}{S} \times H' \times W' \times C \times k \times k, \quad (4)$$

$$y_2 = \frac{N}{S} \times H' \times W' \times (s-1) \times d \times d, \quad (5)$$

$$F_{GC} = y_1 + y_2, \quad (6)$$

其中: s 为鬼影特征图的个数, N/S 为特征图通道数、 $d \times d$ 为鬼影卷积变换时的内核尺寸,大小与 $k \times k$ 相似。则鬼影卷积的压缩比可表示为:

$$r = \frac{F_{SC}}{F_{GC}} \approx \frac{s \times C}{s + C - 1} \approx s, \quad (7)$$

其中, s 远小于 C 。

Gbottleneck和VoVGCSP模块结构如图4所示。Gbottleneck模块主要由鬼影卷积和深度可

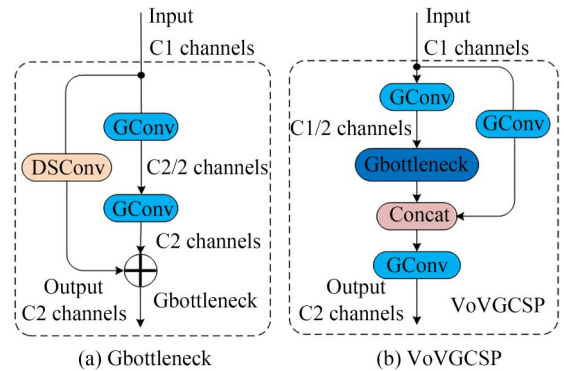


图4 Gbottleneck和VoVGCSP网络结构

Fig. 4 Gbottleneck and VoVGCSP network architecture

分离卷积 (Depthwise Separable Convolution, DSC) 组成。将输入分为两部分,分别经过 GConv 和 DSC 操作,再将输出信息融合。VoVGCSP 模块是在 Gbottleneck 结构的基础上使用一次性聚合方法设计的跨级部分网络。

3.3 三重坐标注意力

为了捕获特征图表面待检测特征的语义信息且平衡检查精度和计算成本,Misra 等^[25]提出三重注意力机制 (Triplet Attention, TA),TA 具有参数量小和实现跨纬度交互等优点,但对空间信息提取不充分,导致检测效果不如 SE^[26],CBAM^[27]和 GA^[28]等注意力。Hou 等^[29]提出坐标注意力机制 (Coordinate Attention, CA),CA 与 TA 相比不仅能获取通道特征,而且可以提取空间位置信息,更好地定位和识别目标,但 CA 仅使用 1×1 卷积对维度 C, W 和 H 之间的交互信息进行提取,导致对通道和空间的交互信息提取不充分。本文受文献[25]和文献[29]的启发,结合三重注意力和坐标注意力的优点,设计三重坐标注意力 (TCA) 模块,其网络结构如图 5 所示。

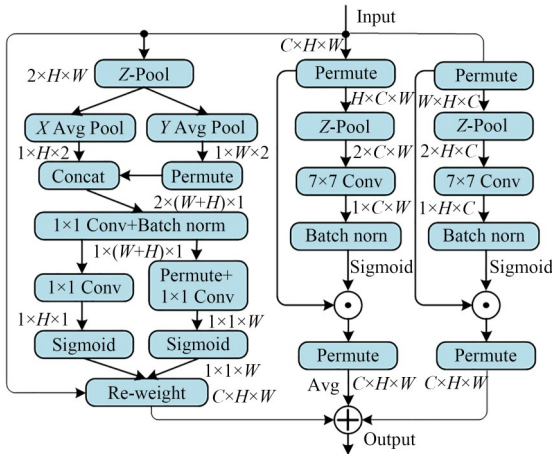


图 5 TCA 网络结构

Fig. 5 TCA network structure

三重坐标注意力的左边第一条分支是坐标注意力,并在该分支引入 $Z-Pool$ 操作,以避免通道维度对空间信息的干扰、充分提取空间和通道的交互信息,同时减少计算量。首先对特征图 $\chi (C \times H \times W)$ 进行全局平均池化和全局最大池化,然后进行拼接,在数学上可表示为:

$$X_1 = AvgPool_c(\chi), \quad (8)$$

$$X_2 = MaxPool_c(\chi), \quad (9)$$

$$Z-Pool(\chi) = Cat[X_1, X_2]. \quad (10)$$

具体而言,第一条分支是空间交互模块,首先在空间水平和垂直方向使用全局最大池化操作提取空间位置信息,并对垂直方向信息进行旋转操作;然后将旋转后的信息与水平方向信息进行拼接,并进行卷积操作以提取两个方向的空间交互信息;最后通过卷积和激活函数,重新分配权重信息。第二和第三分支是通道和空间交互模块,负责提取通道和空间的交互特征,首先对输入特征图 $C \times H \times W$ 进行旋转得到 $W \times C \times H$; 然后经过 $Z-Pool$ 、卷积和激活函数生成注意力特征,该注意力特征与旋转后的特征进行点乘运算;最终将三个分支得到特征信息进行加权平均融合。

3.4 鬼影混洗卷积

原模型在结构上使用大量标准卷积堆叠而成,这种密集卷积会占用过多的参数和计算资源。为保证模型的检测精度,同时降低参数量和计算量,本文引入鬼影混洗卷积 (GSConv)^[30] 替换原模型中的标准卷积,GSConv 结构如图 6 所示。

GSConv 一种轻量高效的卷积结构,由 SC、DSC 和通道混洗 (Shuffle) 模块构成。GSConv 通过压缩通道数来减少计算量。给定输入通道数为 c_1 , 输出通道数为 c_2 。首先对输入通道数为 c_1 的信息进行 SC 操作得到信息 A, 此时通道数变 $c_2/2$; 然后经 DSC 操作得到信息 B; 再将两个信息进行拼接得到通道数为 c_2 的信息; 最后通过 Shuffle 输出通道数为 c_2 的信息。

计算分析:假设卷积核尺寸和输出特征图尺寸分别为 $X \times Y, x \times y$ 。采用 FLOPs 来衡量复杂度,则 SC 与 DSC 的复杂度分别为:

$$F_{SC} = X \times Y \times c_1 \times c_2 \times x \times y, \quad (11)$$

$$F_{DSC} = X \times Y \times c_1 \times x \times y + c_1 \times c_2 \times x \times y. \quad (12)$$

经 DSC 操作后可将 SC 计算量压缩为 $F_{SC_{DSC}}$, 其计算式如下:

$$\varphi = \frac{X \times Y \times c_1 \times x \times y + c_1 \times c_2 \times x \times y}{X \times Y \times c_1 \times c_2 \times x \times y}, \quad (13)$$

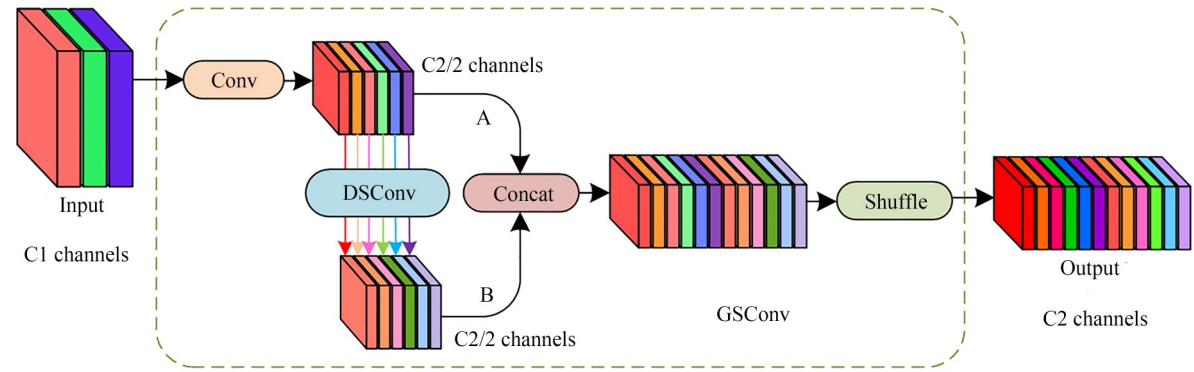


图 6 GSConv 网络结构
Fig. 6 GSConv network structure

$$F_{SC_{dsc}} = \varphi = \frac{1}{c2} + \frac{1}{X \times Y}. \quad (14)$$

GSConv 中的卷积核由主卷积核和鬼影卷积核组成,其中主卷积核负责主要计算、鬼影卷积核用来补全主卷积核计算,从而达到分组卷积的效果,将 SC 替换成 GSConv 可大幅度降低计算量,其计算式为:

$$F_{GSC} = (X \times Y)(n) = \sum_{i=1}^{n-1} f(i)g(n \oplus i \oplus r(i)), \quad (15)$$

其中: n 表示输入长度、 \oplus 表示异或操作、 $r(i)$ 表示输入第 i 个元素。

3.5 添加检测层

原模型采用三个 IDetect 检测器作为检测头,当输入尺寸为 640×640 时,三个检测层输出特征图的尺寸为: 80×80 , 40×40 和 20×20 ,分别用于检测小、中、大三个目标。钢材表面缺陷虽然小目标占比居多,但也存在部分缺陷在特征图中占比较大。为了提高这类目标的检测精度,在原模型基础上增加 10×10 尺寸检测层,并使用 K-means 聚类算法对 NEU-DET 和 Severstal 数据集重新设定锚定框,改进后的锚定框大小如表 1 所示。为避免参数量和计算量过大,省略检测层与 Neck 之间的卷积操作。

表 1 改进后各检测层锚定框大小

Tab. 1 Improved anchoring frame sizes for each detection layer

特征图尺寸	感受野	NEU-DET 数据集锚定框大小	Severstal 数据集锚定框大小
10×10	超大	(361,266)(246,560)(490,540)	(478,238)(204,502)(606,613)
20×20	大	(117,517)(217,355)(438,190)	(428,122)(267,235)(112,576)
40×40	中	(51,529)(176,193)(524,77)	(48,573)(159,187)(157,318)
80×80	小	(61,130)(135,111)(82,238)	(60,121)(132,111)(76,239)

4 实验结果及分析

4.1 实验环境与参数设置

实验环境如下: Windows11 操作系统,深度学习框架为 pytorch1.11.0, CPU 为 Intel Core i7-12700H, GPU 为 RTX4060, 显卡为 8 GB, 优化器为 Adam, 初始学习率为 0.01, 动量为 0.973, 权重衰减因子为 0.000 5。其输入图像尺寸为

640×640 , batchsize 为 16, epoch 为 300 轮。

4.2 数据集

本文采用东北大学 NEU-DET^[18] 钢材缺陷数据集和来自 Kaggle 钢铁缺陷比赛数据集, 由谢韦尔钢铁公司 (Severstal) 提供。其中 NEU-DET 和 Severstal 数据集图片分别为 1 800 和 1 120 张图片, 由于 Severstal 数据集图片过少, 本文采用水平翻转、垂直翻转和水平垂直翻转对其进行扩充, 比例

为1:3,扩充后的数据集共4 480张图片。两个数据集的训练集和验证集均按8:2随机进行划分,均包

含裂缝、夹杂物、斑块、点蚀表面、氧化皮和划痕六种缺陷,原图和扩充后的图片如图7所示。

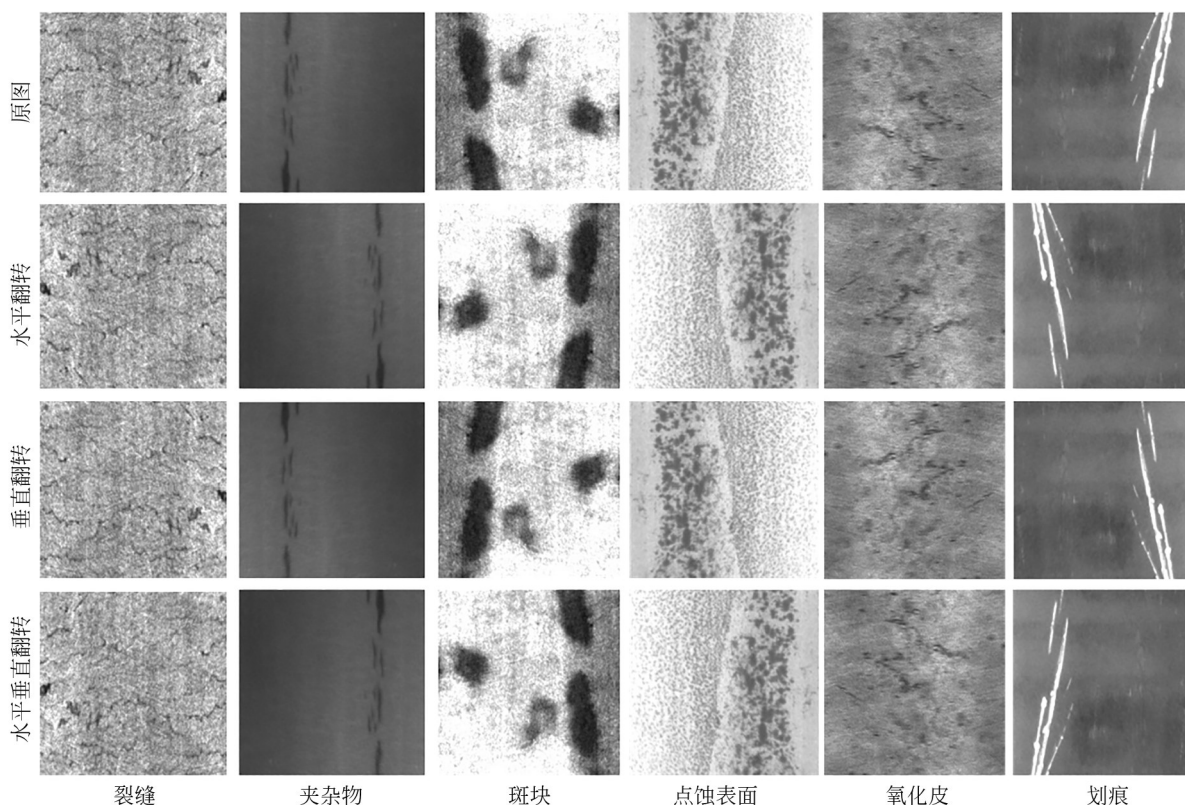


图7 钢材表面各种缺陷图像

Fig. 7 Image of various defects on the steel surface

4.3 评价指标

本实验采用平均精度(Average Precision, AP)、平均精度均值(mean AP, mAP)、每秒检测帧数(Frames Per Second, FPS)、参数量(Params)、浮点操作(Floating-point Operations, FLOPs)和模型大小(Size)作为评价指标。衡量上述指标的计算式分别为:

$$AP = \int_0^1 P dR, \quad (16)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}, \quad (17)$$

$$Precision = \frac{TP}{TP + FP}, \quad (18)$$

$$Recall = \frac{TP}{TP + FN}, \quad (19)$$

$$FPS = \frac{Frameum}{ElapsedTime}, \quad (20)$$

其中:AP表示P-R曲线坐标轴围成的面积;mAP是对所有类别的AP求平均值所得; N 表示检测的类别数,本实验 $N=6$;TP,FP和FN分别表示正确检测到目标框、误检框和漏检框的数量;Frameum是需要检测图片的总数,ElapsedTime则表示检测所需花费的总时间。

4.4 各模块对比实验

4.4.1 改进VoVGA-FPN对比实验

为验证改进VoVGA-FPN融合网络的有效性,在NEU-DET数据集上与PANet和AFPN网络进行对比,实验结果如表2所示,其中加粗为最优值。

从表2实验结果可知,对四个指标综合考虑,本文设计的VoVGA-FPN融合网络相比于改进前的AFPN网络mAP和FPS值分别提升0.5%和3帧、参数量和计算复杂度分别降低0.76 M和1.4 G,说明本文设计的融合网络在具备模型参

表 2 改进 VoVGA-FPN 对比实验

Tab. 2 Improved VoVGA-FPN comparison experiments

Model	mAP/%	Params/M	FLOPs/G	FPS/ms
PANet	68.7	6.02	13.1	108
AFPN	70.7	7.11	14.1	108
Ours	71.2	6.35	12.7	111

数量优势的同时,拥有更好的检测性能和更快的速度,虽然参数量比 PANet 网络增加 3.3%,但 mAP, FLOPs 和 FPS 指标相比于 PANet 网络均为最优。表明改进后的 VoVGA-FPN 网络具有优越性。

4.4.2 注意力对比实验

为确定 TCA 模块加在 YOLOv7-tiny 主干网络的所处位置对各项指标的提升效果最好,在 NEU-DET 数据集上,针对以下 3 种情况进行实验。其中,情况 1,2,3 分别表示只在主干的第三、四、五个特征提取后加入 TCA 模块,实验结果如表 3 所示,其中加粗为最优值。

表 3 TCA 模块不同位置实验

Tab. 3 Experiments with different positions of the TCA module

Location	mAP/%	Params/M	FLOPs/G
Baseline	68.7	6.020 1	13.1
情况 1	70.1	6.020 4	13.2
情况 2	70.8	6.020 5	13.2
情况 3	69.8	6.020 8	13.2

从表 3 实验结果可知, TCA 模块嵌入在主干三个位置所带来的计算量相当,虽然情况 1 增加的参数量最少,但 mAP 值比情况 2 低,情况 2 的 mAP 值最优且参数量与情况 1 相差甚微,情况 3 的检测性能和模型参数量等均不如情况 2。综上所述,情况 2 在牺牲较少参数量和计算复杂度的情况下换来更佳的检测性能。因此,情况 2 的效果更佳。

同时,为验证改进 TCA 模块的有效性,在 NEU-DET 数据集和同一实验条件下与 SE, GC, CA, CBAM 和 TA 注意力进行对比,实验结果如表 4 所示,其中加粗为最优值。

从实验结果可知,对三个指标综合考虑, TCA 注意力相比于改进前的 TA 和 CA 注意力,

表 4 TCA 对比实验

Tab. 4 TCA comparison experiments

Model	mAP/%	Params/M	FLOPs/G
Baseline	68.7	6.020 1	13.1
SE	70.4	11.570	30.8
GC	70.1	6.037 3	13.2
CA	68.3	6.034 5	13.5
CBAM	70.1	6.028 6	13.2
TA	69.9	6.020 5	13.2
TCA(Ours)	70.8	6.020 5	13.2

mAP 值最优达到 70.8%,所带来的参数量和计算量最少。虽然 SE 在性能上 mAP 值也达到 70.4%,但参数量和计算复杂度却接近 TCA 的两倍,不利于轻量化部署。其余注意力机制虽然在参数量和计算量上与 TCA 相当,但检测效果均不如 TCA,实验表明改进后 TCA 模块的合理性,具备更佳的检测精度和轻量化。

4.4.3 GSConv 替换位置对比实验

为确定 GSConv 替换原模型 SC 的所处位置对指标的提升效果最好,在 NEU-DET 数据集上,针对以下 3 种情况进行验证。其中, A 表示原模型, B, C, D 分别表示替换主干、颈部和全部,实验结果如表 5 所示,其中加粗为最优值。

表 5 GSConv 模块不同位置实验

Tab. 5 Experiments with different positions of the GSConv module

Location	mAP/%	Params/M	FLOPs/G	FPS/ms
A	68.7	6.02	13.1	108
B	70.1	4.82	9.9	94
C	70.3	4.31	10.3	121
D	72.0	3.12	7.1	100

从表 5 实验结果可知,对四个指标综合考虑,情况 C 所带来的检测速度最快,达到 121 帧,但相比于情况 D,在检测精度和轻量化上均不具备优势,虽然情况 D 的 FPS 相比于原模型减少 8 帧,但其他指标均为最优, mAP 值达到 72%、参数量和计算复杂度低,分别为 3.12M 和 7.1 G,在具备检测能力的同时更利于轻量化部署,故情况 D 的效果最佳。

4.4.4 检测层有无连接卷积对比实验

本文对所增加检测层与 Neck 网络之间是否需要经过卷积操作,在 NEU-DET 数据集上进行验证。其中,A 表示原模型、B 和 C 分别表示检测层与 Neck 无卷积连接和有卷积连接。实验结果如表 6 所示,加粗为最优值。

从表 6 实验结果可知,对四个指标综合考虑,虽然将检测层与 Neck 连接时加入卷积后 mAP 值最高,达到 73%,但以牺牲参数量、计算

量和检测速度为代价换来精度的提升,与本文轻量化的目的相反。而检测层与 Neck 无卷积连接时模型参数量和计算量相比于原模型只增加 1.85 M 和 0.4 G,且检测速度为最优,达到 111 帧,mAP 值相比于原模型提高 3.4%,检测性能的提升效果也较好,综上所述,选择不经过卷积连接。

4.5 消融实验

本文针对 YOLOv7-tiny 原始模型进行四个方面改进,实验为验证各个模块的有效性,分别在 NEU-DET 和 Severstal 数据集上,通过逐一消融的方式进行验证。其中,M1 表示原模型、M2 在 M1 基础上添加 TCA,M3 在 M2 基础上添加检测层、M4 在 M3 基础上添加 VoVGA-FPN 融合网络、M5 在 M4 基础上添加 GSConv,即本文所提 VTG-YOLOv7-tiny 算法。实验结果如表 7 所示,加粗为最优值。

表 6 新增检测层与 Neck 有无卷积实验
Tab. 6 New detection layer with Neck with and without convolution experiments

Model	mAP%	Params/M	FLOPs/G	FPS/ms
A	68.7	6.02	13.1	108
B	72.1	7.87	13.5	111
C	73.0	10.2	13.9	80

表 7 NEU-DET 和 Severstal 数据集消融实验
Tab. 7 Ablation experiments on NEU-DET and Severstal datasets

Dataset	Method	mAP%	Params/M	FLOPs/G	FPS/ms	P%	R%
NEU-DET	M1	68.7	6.02	13.1	108	61.4	72.7
	M2	70.8	6.02	13.2	119	63.8	73.7
	M3	72.5	7.87	13.5	113	66.7	74.2
	M4	73.6	7.04	12.7	76	67.1	72.0
	M5	74.4	5.41	8.9	87	68.5	74.5
Severstal	M1	65.6	6.02	13.1	61	58.1	65.2
	M2	67.6	6.02	13.2	79	65.8	70.2
	M3	71.5	7.87	13.5	70	64.5	72.1
	M4	72.7	7.04	12.7	52	66.8	67.7
	M5	74.1	5.41	8.9	78	67.0	72.2

从表 7 实验结果可知,轻量级的 TCA 模块能够有效提取通道和空间交互的语义信息,提升网络 mAP、精确度和召回率,同时 TCA 通过自适应地调整不同分支的权重,更好地处理不同尺度和方向的缺陷目标,有效地缩小网络的感受野,降低冗余计算量,进而提升检测速度;增加检测层 mAP 值提升较大,说明该方法能进一步改善部分缺陷在特征图中占比较大,导致检测精度低的问题;添加 VoVGA-FPN 网络,除检测速度外各项指标均有所提升,表明该网络能有效避免语义信息在传递过程中的丢失,提升网络特征融合能

力;最后轻量级的 GSConv 模块在有效降低模型参数量和计算量的同时能进一步权衡 mAP 值。在 NEU-DET 和 Severstal 数据集上,VTG-YOLOv7-tiny 算法相比于 YOLOv7-tiny 模型 mAP 值分别提升 5.7% 和 8.5%、参数量和计算量分别 5.41 M 和 8.9 G,相比于原模型分别降低 0.61 M 和 4.2 G,仅为 YOLOv7-tiny 的 89.8% 和 67.9%,同时保证了算法的实时性。两个数据集的检测速度存在明显差异,主要是因为 NEU-DET 数据集为 200×200 pixels,分辨率较低,而 Severstal 数据集则为 256×1 600 pixels,分辨率

较高,且前者的样本数量比后者少,故相对而言 NEU-DET 数据集的检测速度更快。消融实验验证了本文所提模块的有效性和合理性。

同时,对于 Severstal 数据集,为验证其各项指标是因数据增强所达到的效果还是本文算法的性能所导致。在相同实验条件下,将本文算法和原模型算法在扩充前后的数据集上进行实验对比,其中,SeverstalA,B 分别表示扩充前后的数据集,实验结果如表 8 所示,加粗为最优值。

表 8 Severstal 数据集扩充前后实验对比

Tab.8 Comparison of experiments before and after enrichment of the Severstal dataset

Dataset	Model	mAP/%	Params/ M	FLOPs/ G	FPS/ ms
SeverstalA	M1	71.5	6.02	13.1	89
	M5	75.2	5.41	8.9	94
SeverstalB	M1	65.6	6.02	13.1	61
	M5	74.1	5.41	8.9	78

从表 8 实验结果可知,Severstal 数据集扩充前后在 YOLOv7-tiny(M1)模型上的 mAP 值分别为 71.5% 和 65.6%,而在本文算法(M5)上 mAP 值分别达到 75.2% 和 74.1%,故可以看出 mAP 值的提升是本文模型提升的效果。扩充后的数据集数量为原来的三倍,导致推理和计算量增加,使得 FPS 值有所降低,但依然满足实时检测性。

本文算法与原模型在 NEU-DET 和 Severstal 数据集上对检测各类缺陷的平均精度值如图 8 所示,从图中可以看出本文算法对各类缺陷的平均检测精度值均有所提升,在 NEU-DET 数据集上对裂缝和氧化皮的提升效果最佳,分别提高 5.5% 和 18.2%;而在 Severstal 数据集上则对裂缝和划痕两类缺陷的检测精度提升最好,分别提高 8.1% 和 23.1%,对于两个数据集的其余类别缺陷原模型已达到较高的检测精度,所以提升效果略低,实验结果表明改进后的算法能有效提高对钢材表面缺陷目标的检测精度。

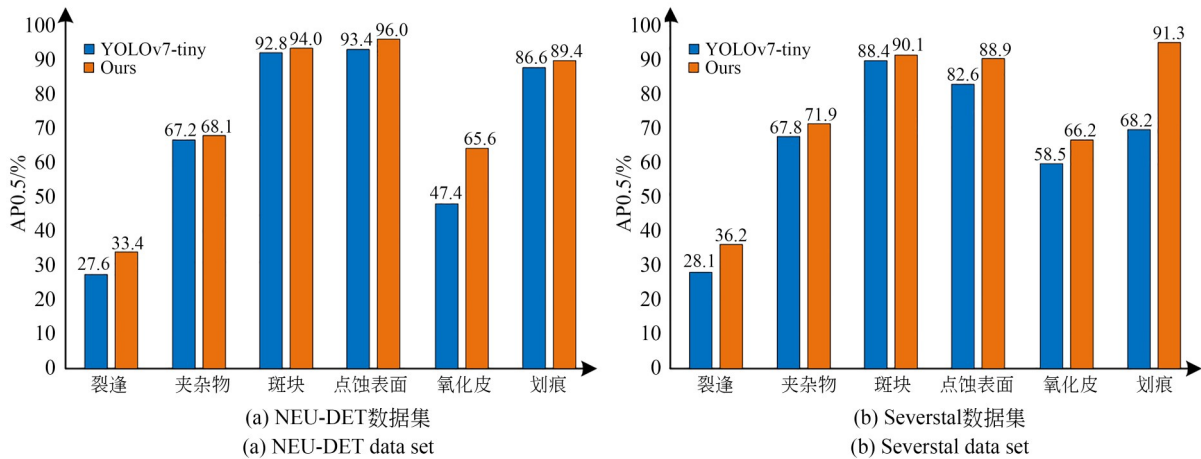


图 8 本文算法与原模型对各类缺陷 AP 值对比

Fig. 8 Comparison of AP values of this paper's algorithm and the original model for various types of defects

4.6 对比实验分析

本文在相同实验条件下,分别在 NEU-DET 和 Severstal 数据集上,将改进后的算法与 Faster R-CNN, SSD, YOLOv3, YOLOv3-tiny, YOLOv4, YOLOv4-tiny, YOLOv5s, YOLOX-s, YOLOv7, YOLOv8s 以及文献[1]、[16]和[18]算法进行对比。由于 Severstal 数据集扩充后图片数量过多,本文实验设备的计算能力和内存容

量有限,除 YOLOv3-tiny, YOLOv4-tiny, YOLOv5s 和 YOLOX-s 算法外,其余算法参数量和计算量过大,故无法进行实验,而文献[1]、文献[16]和文献[18]算法的检测对象是 NEU-DET 数据集,故无法在 Severstal 数据集上进行对比分析,实验结果如表 9 所示,其中加粗为最优值。

从表 9 实验结果可知,在 NEU-DET 数据集上虽然文献[16]和文献[18]的 mAP 值较高,分

表 9 NEU-DET 和 Severstal 数据集对比实验
Tab. 9 Experiments comparing NEU-DET and Severstal datasets

Dataset	Model	mAP0.5/%	Params/ M	FLOPs/G	Size/MB
NEU-DET	Faster R-CNN	65.7	72.0	167.3	108.0
	SSD	61.0	41.1	145.3	93.1
	YOLOv3	67.0	61.5	155.0	117.0
	YOLOv3-tiny	46.5	8.67	12.9	16.6
	YOLOv4	51.0	52.5	119.8	102.6
	YOLOv4-tiny	54.6	5.90	16.18	22.5
	YOLOv5s	70.1	7.07	16.4	13.7
	YOLOX-s	71.8	8.0	21.6	16.3
	YOLOv7	70.0	37.2	104.8	74.8
	YOLOv7-tiny	68.7	6.02	13.1	11.7
	YOLOv8s	72.1	11.1	28.4	21.4
	文献[1]	74.1	23.9	—	—
	文献[16]	82.4	15.37	—	29.7
	文献[18]	77.2	7.03	—	—
Severstal	VTG-YOLOv7-tiny(Ours)	74.4	5.41	8.9	10.8
	YOLOv3-tiny	56.4	8.67	12.9	16.6
	YOLOv4-tiny	59.6	5.9	16.18	22.5
	YOLOv5s	72.4	7.07	16.4	13.7
	YOLOX-s	73.8	8.0	21.6	16.3
	YOLOv7-tiny	65.6	6.02	13.1	11.7
	YOLOv8s	71.8	11.1	28.4	21.4
	VTG-YOLOv7-tiny(Ours)	74.1	5.41	8.9	10.8

别为 82.4% 和 77.2%, 但该指标与实验设备、数据集划分以及实验参数设置也具有相关性, 而参数量和模型大小在不同实验设备中差异略微, 文献[16]的参数量和模型大小接近本文算法的三倍, 文献[1]、[16]和[18]均无法有效地平衡轻量化和检测精度。Faster R-CNN, SSD, YOLOv3 和 YOLOv4 算法的 mAP 值较低, 无法有效对钢材表面缺陷进行检测, 且参数量、计算复杂度过大, 不利于设备的部署。YOLOv3-tiny, YOLOv4-tiny, YOLOv5s 和 YOLOX-s 算法在轻量化上得到一定优化, 但检测精度也随之降低。在 Severstal 数据集上本文算法相比于其余对比算法各项指标均较优, 在达到较高检测精度的基础上具备轻量化优势。综合对比可见本文算法 (VTG-YOLOv7-tiny) 在 NEU-DET 和 Severstal 数据集上检测性能较优, VTG-YOLOv7-tiny 算法在 NEU-DET 数据集上 mAP 值达到 74.4%, 在 Severstal 数据集上最优, 达到 74.1%, 说明其

检测效果较好。参数量、计算复杂度和模型大小则衡量模型的部署能力, VTG-YOLOv7-tiny 的参数量、计算复杂度和模型大小分别为 5.41 M, 8.9 G 和 10.8 MB 为最优, 说明本文算法易于在边缘终端设备部署。对比实验结果表明本文算法相比于其他算法更好地平衡检测精度和轻量化, 泛化性能更好。

4.7 可视化对比

本文算法与原模型算法对 NEU-DET 和 Severstal 数据集钢材表面 6 种缺陷的检测效果如图 9 所示。对于不同的缺陷类别分别用不同颜色的锚定框来表示。从图中可以看出, YOLOv7-tiny 存在钢材缺陷漏检现象, 而本文 VTG-YOLOv7-tiny 对目标缺陷区域的检测能力更强, 且检测钢材缺陷更多。从图 8 本文算法和原模型对两个数据集各类缺陷的检测精度对比图可知, 在 NEU-DET 和 Severstal 数据集上, 本文算法分别对裂缝、氧化皮和点蚀表面三类缺陷以及裂缝和

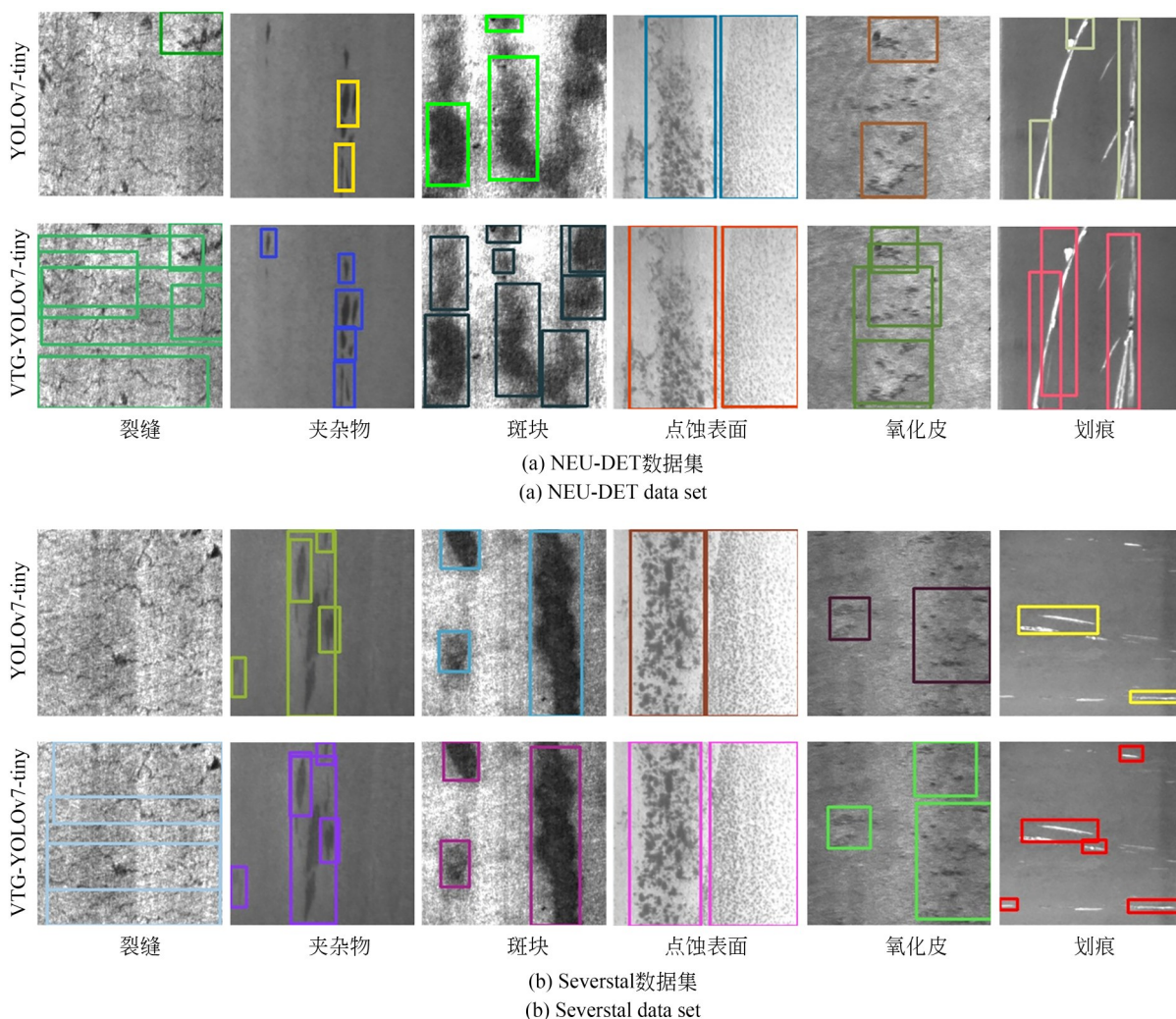


图9 本文算法与原模型检测效果对比

Fig. 9 Comparison of the detection effect of this paper's algorithm and the original model

划痕两类缺陷的检测精度提升较大,从图9(a) NEU-DET数据集和图9(b)Severstal数据集中也可以直观看出本文算法对所提类别的缺陷能识别出更多有效的检测框。说明改进后的算法有更强的特征提取和特征融合能力,在一定程度上可以避免漏检情况,能有效提升钢材表面缺陷的检测精度。

5 结 论

本文提出轻量化VTG-YOLOv7-tiny的钢材缺陷检测算法,有效平衡钢材表面缺陷精度和模型轻量化的问题。首先设计VoVGA-FPN网络减少语义信息丢失,提升模型特征融合能力;其次构建TCA模块提高网络对通道和空间信息的

提取能力;再次引入GSConv,利用其高效的特征提取能力和轻量化特性,在提升精度的同时有效减少模型参数量和计算量;最后增加大目标检测层,有效改善部分缺陷在特征图中占比较大,导致检测精度低的问题。在NEU-DET和Severstal数据集上验证其有效性,mAP值分别提升5.7%和8.5%、参数量和计算量仅为原模型的89.8%和67.9%。结果表明,本文方法为在边缘设备部署提供了参考,对工业钢材表面缺陷检测有一定的应用价值。但也存在不足之处,比如:VTG-YOLOv7-tiny算法在NEU-DET数据集上的检测速度相比于原模型算法有所降低,针对该问题,未来可以通过调整阈值、改变锚点、调整网络结构和尺寸、采用数据增强和优化训练等方法进一步平衡检测精度和速度。

参考文献:

- [1] 曹义亲, 伍铭林, 徐露. 基于改进YOLOv5算法的钢材表面缺陷检测[J]. 图学学报, 2023, 44(2): 335-345.
CAO Y Q, WU M L, XU L. Steel surface defect detection based on improved YOLOv5 algorithm [J]. *Journal of Graphics*, 2023, 44(2): 335-345. (in Chinese)
- [2] ZHOU L S, LIANG L M, SHENG X Q. GA-Net: ghost convolution adaptive fusion skin lesion segmentation network [J]. *Computers in Biology and Medicine*, 2023, 164: 107273.
- [3] 郭峰, 朱启兵, 黄敏, 等. 基于改进YOLOV4的陶瓷基板瑕疵检测[J]. 光学精密工程, 2022, 30(13): 1631-1641.
GUO F, ZHU Q B, HUANG M, *et al.* Defect detection in ceramic substrate based on improved YOLOV4 [J]. *Opt. Precision Eng.*, 2022, 30(13): 1631-1641. (in Chinese)
- [4] 陈亚芳, 廖飞, 黄新宇, 等. 多尺度YOLOv5的太阳能电池缺陷检测[J]. 光学精密工程, 2023, 31(12): 1804-1815.
CHEN Y F, LIAO F, HUANG X Y, *et al.* Multi-scale YOLOv5 for solar cell defect detection [J]. *Opt. Precision Eng.*, 2023, 31(12): 1804-1815. (in Chinese)
- [5] 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.
LI S B, YANG J, WANG Z, *et al.* Review of development and application of defect detection technology [J]. *Acta Automatica Sinica*, 2020, 46(11): 2319-2336. (in Chinese)
- [6] 朱玉龙, 赵迎松, 方阳, 等. 孔边裂纹的旋转涡流检测[J]. 中国机械工程, 2023, 34(8): 883-891.
ZHU Y L, ZHAO Y S, FANG Y, *et al.* Rotating eddy current testing for inspection of cracks at hole edge [J]. *China Mechanical Engineering*, 2023, 34(8): 883-891. (in Chinese)
- [7] GIRSHICK R. Fast R-CNN [C]. 2015 *IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile. IEEE, 2015: 1440-1448.
- [8] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy. IEEE, 2017: 2980-2988.
- [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)*. Las Vegas, NV, USA. IEEE, 2016: 779-788.
- [10] HUSSAIN M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection [J]. *Machines*, 2023, 11(7): 677.
- [11] QIAN H M, WANG H L, FENG S, *et al.* FES-SD: SSD target detection based on feature fusion and feature enhancement [J]. *Journal of Real-Time Image Processing*, 2023, 20(1): 2.
- [12] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy. IEEE, 2017: 2999-3007.
- [13] 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)*. Honolulu, HI, USA. IEEE, 2017: 936-944.
- [14] 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*. Salt Lake City, UT, USA. IEEE, 2018: 8759-8768.
- [15] 张利红, 蔡敬菊. 基于轻量化Yolov5算法的目标检测系统 [J]. 计算机技术与发展, 2022, 32(11): 134-139.
ZHANG L H, CAI J J. Target detection system based on lightweight Yolov5 algorithm [J]. *Computer Technology and Development*, 2022, 32(11): 134-139. (in Chinese)
- [16] 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 [J]. 电子测量与仪器学报, 2022, 36(8): 150-157.
MA Y T, ZHAO H D, YAN C, *et al.* Strip steel surface defect detection method by improved YOLOv5 network [J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(8): 150-157. (in Chinese)
- [17] 李成跃, 姚剑敏, 林志贤, 等. 基于改进YOLO轻量化网络的目标检测方法 [J]. 激光与光电子学进展, 2020, 57(14): 141003.
LI C Y, YAO J M, LIN Z X, *et al.* Object detection method based on improved YOLO lightweight

- network [J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141003. (in Chinese)
- [18] 周彦, 孟江南, 王冬丽, 等. 基于多尺度轻量化注意力的钢材缺陷检测[J]. *控制与决策*, 2024, 39(3): 901-909.
- ZHOU Y, MENG J N, WANG D L, *et al.* Steel defect detection based on multi-scale lightweight attention [J]. *Control and Decision*, 2024, 39(3): 901-909. (in Chinese)
- [19] LIAN J W, HE J H, NIU Y, *et al.* Fast and accurate detection of surface defect based on improved YOLOv4 [J]. *Assembly Automation*, 2022, 42(1): 134-146.
- [20] HAN K, WANG Y H, TIAN Q, *et al.* Ghost-Net: more features from cheap operations [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA. IEEE, 2020: 1577-1586.
- [21] 徐正军, 张强, 许亮. 一种基于改进 YOLOv5s-Ghost 网络的交通标志识别方法[J]. *光电子·激光*, 2023, 34(1): 52-61.
- XU Z J, ZHANG Q, XU L. A traffic sign recognition method based on improved YOLOv5s-Ghost network [J]. *Journal of Optoelectronics·Laser*, 2023, 34(1): 52-61. (in Chinese)
- [22] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada. IEEE, 2023: 7464-7475.
- [23] YANG G Y, LEI J, ZHU Z K, *et al.* AFPN: Asymptotic Feature Pyramid Network for Object Detection [EB/OL]. 2023: *arXiv*: 2306.15988. <http://arxiv.org/abs/2306.15988>
- [24] LIU S T, HUANG D, WANG Y H. Learning Spatial Fusion for Single-Shot Object Detection [EB/OL]. 2019: *arXiv*: 1911.09516. <http://arxiv.org/abs/1911.09516>
- [25] MISRA D, NALAMADA T, ARASANIPALAI A U, *et al.* Rotate to attend: convolutional triplet attention module [C]. 2021 *IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA. IEEE, 2021: 3138-3147.
- [26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 2018: 7132-7141.
- [27] 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.
- [28] LIU Y C, SHAO Z R, HOFFMANN N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions [EB/OL]. 2021: *arXiv*: 2112.05561. <http://arxiv.org/abs/2112.05561>
- [29] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [30] LI H L, LI J, WEI H B, *et al.* Slim-Neck by GSConv: a Better Design Paradigm of Detector Architectures for Autonomous Vehicles [EB/OL]. 2022: *arXiv*: 2206.02424. <http://arxiv.org/abs/2206.02424>

作者简介:



梁礼明(1967—),男,江西吉安人,硕士,教授,硕士生导师,主要研究方向为机器学习、模式识别与图像处理等。E-mail:9119890012@jxust.edu.cn

通讯作者:



龙鹏威(1998—),男,江西赣州人,硕士研究生,主要研究方向为机器学习、模式识别与图像处理等。E-mail:2637018663@qq.com