

基于改进 YOLOv7 的钢材表面缺陷检测

付帅, 凌铭[†], 楚东港

(上海工程技术大学 电子与电气工程学院, 上海 201620)

摘要: 钢材表面缺陷对于钢材行业来说是一个巨大的挑战。针对传统的钢材缺陷检测方法存在着效率低、检测精度不高等问题, 基于 YOLOv7 设计了一种 AFSD-YOLOv7 模型进行实时的钢材表面缺陷检测。首先, 在 YOLOv7 模型中使用一种轻量化卷积结构替换标准卷积结构, 以加速模型的推理过程; 然后采用快速空间金字塔池化结构替换原始空间金字塔池化结构, 以加速网络的特征提取过程; 最后添加改进的 ECA-Net 注意力机制, 以提升模型检测精度。实验结果表明, AFSD-YOLOv7 能够对钢材缺陷进行有效识别, 相比 YOLOv7 模型, 计算量减少了 54.8%, mAP 提高了 3.2%, 对于钢材表面缺陷检测具有实际应用价值。

关键词: 钢材; 缺陷检测; YOLOv7; 神经网络; 深度学习; 注意力机制; 标准卷积

中图分类号: TP39

文献标识码: A

Steel Surface Defect Detection Based on Improved YOLOv7

FU Shuai, LING Ming[†], CHU Donggang

(School of Electronic and Electrical Engineering, Shanghai University
of Engineering Science, Shanghai 201620, China)

Abstract: Surface defects on steel are a significant challenge for the steel industry. Traditional steel defect detection methods suffer from low efficiency and accuracy. To address these issues, an AFSD-YOLOv7 model has been designed for real-time steel surface defect detection. First, a lightweight convolutional structure was used to replace the standard convolutional structure in the YOLOv7 model, speeding up the inference process. Next, a fast spatial pyramid pooling structure was used to replace the original spatial pyramid pooling structure to accelerate the network's feature extraction process. Finally, an improved ECA-Net attention mechanism was added to enhance the model's detection accuracy. Experimental results show that AFSD-YOLOv7 can effectively identify steel defects. Compared to the YOLOv7 model, AFSD-YOLOv7 reduces computation by 54.8% and improves mAP by 3.2%, indicating significant practical value for steel surface defect detection.

Key words: steel; defect detection; YOLOv7; neural networks; deep learning; attention mechanism; standard convolution

随着工业化和数字化技术的不断发展, 钢材表面缺陷检测已经成为生产过程中不可或缺的一部分。传统的缺陷检测方法存在着检测效率低、检测精度不高、易受主观因素影响等问题, 无法满足现代工业对高效、精准、自动化生产的要求。因

而, 钢材表面缺陷作为钢材生产过程中的一个重要质量问题, 对钢材的加工和使用都具有巨大的影响^[1]。据悉, 钢材表面缺陷种类繁多, 包括裂痕、夹杂、斑点、点蚀、锈蚀以及划痕等多种类型。这些缺陷在加工过程中容易导致钢材表面损坏, 使其质量

降低,从而严重影响钢材的使用效果和寿命。因此,开展钢材表面缺陷的自动化检测研究尤为必要^[2]。

近年来,钢材表面缺陷的检测越来越受到关注。基于传统机器视觉的缺陷检测方法也被广泛应用,包括阈值法^[3]、边缘检测法^[4]、纹理分析法^[5]、卷积神经网络^[6]等。尽管这些方法都能检测出钢材表面的某些缺陷,但并不能覆盖生产过程中可能出现的所有缺陷类型,难以实现精确检测,并且检测实时性较差。

针对上述问题,本文基于 YOLOv7 提出了一种 AFSD-YOLOv7 模型的钢材表面缺陷检测方法。使用改进的卷积结构替换标准卷积结构,以加

速模型的推理;采用快速空间金字塔池化结构,加速主干网络的特征提取过程;最后添加改进的 ECA-Net 注意力机制,以提升模型的检测效果。

1 钢材表面缺陷检测算法

1.1 AFSD-YOLOv7 网络结构

YOLOv7 是一种快速的单阶段目标检测模型,它在速度和精度方面优于大部分已知的目标检测网络^[7]。本文对 YOLOv7 算法进行改进,提出 AFSD-YOLOv7 (accurate and fast inspection of steel surface defects-YOLOv7) 钢材表面缺陷检测算法,其网络结构如图 1 所示。

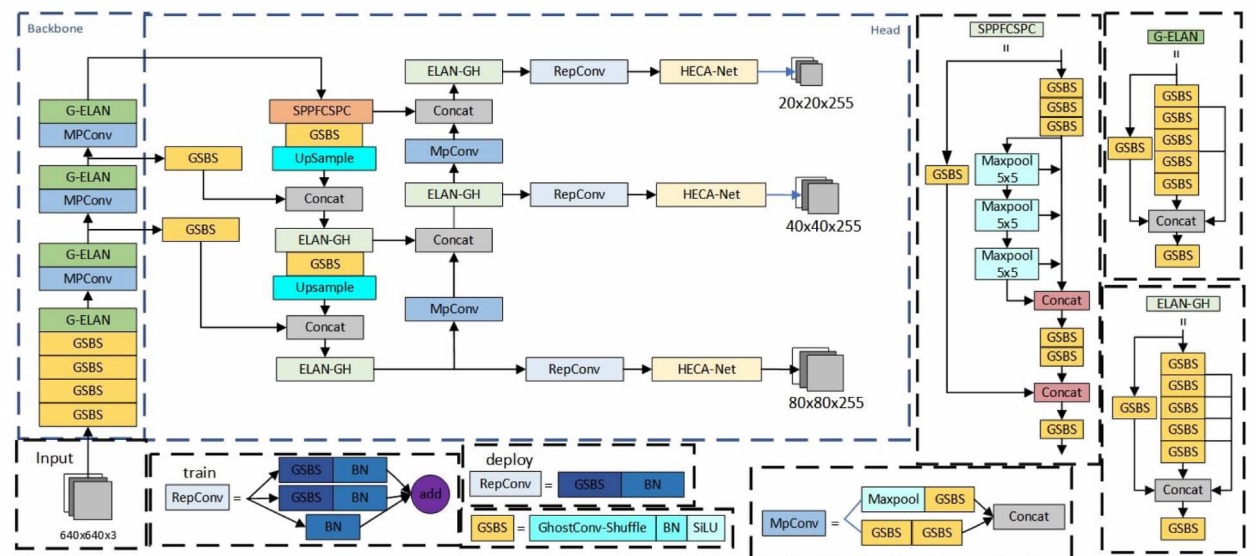


图 1 AFSD-YOLOv7 网络结构

1.2 GhostConv-Shuffle 卷积结构

传统的特征提取方法存在着大量参数的问题,这是因为它通过多层卷积对输入特征映射的所有通道进行卷积映射,这会消耗大量计算资源。尽管这些方法能够生成丰富的特征图,但同时也会不可避免地造成资源浪费。本文设计了 GhostConv-Shuffle(GS)模块替代网络中的标准卷积,它是一种轻量化卷积结构,能在输出大小与标准卷积一样的情况下,降低通用卷积层的计算成本,并保持与标准卷积相似的功能^[8]。它的结构如图 2 所示。

GhostConv 的表达式如式(1)所示。

$$Y' = X \times W' \tag{1}$$

其中,输入特征的格式为 $X \in R^{h \times w \times l}$, W' 表示 GhostConv 的卷积核,它的格式为 $R^{c \times k \times k \times m}$, m 表示 GhostConv 卷积核的通道数, $m \ll n$, n 为最后所需输出的通道数,为了方便介绍,此处省略偏置

项。 Y' 表示输出特征,它的结构为 $Y' \in R^{h' \times w' \times m}$,其中 h' 和 w' 分别为输出特征图的高度和宽度。此外,其他超参数,比如填充、步幅和卷积核大小都与标准卷积一样,以确保输出特征的大小与标准卷积一致。为了进一步获得所需的特征,对 Y' 进行一系列的线性运算,如式(2)所示。

$$Y_{ij} = \varphi_{i,j}(Y'_i), \forall i = 1, \dots, m, j = 1, \dots, s \tag{2}$$

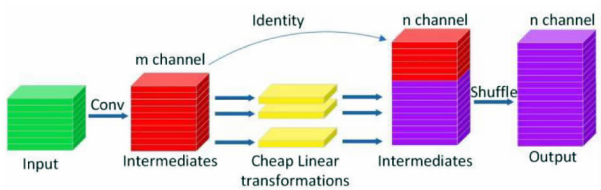


图 2 GhostConv-Shuffle 结构

其中, Y'_i 表示 Y' 中的第 i 个特征图, $\varphi_{i,j}$ 表示对第 j 个(最后一个除外)特征图 Y'_i 进行线性运算的函数, 最后一个 $\varphi_{i,j}$ 函数, 是对特征图保留固有特征的映射, 于是可以得到 $n = m \times s$ 个特征图。Shuffle 操作的重要作用是增强模型的表达能力。通过对输入特征图进行 Shuffle 操作, 不同组的通道可以相互交流, 从而促进特征的信息交换和整合。这种交换和整合的过程有助于提高模型的表达能力, 并使得模型更加适合处理复杂的输入数据。为比较 GhostConv-Shuffle 与标准卷积(standard convolution, SC)的计算复杂度, 采用 FLOPs 作为衡量标准, 标准卷积的模型复杂度如式(3)所示。

$$FLOP_{s^{SC}} = n \cdot h' \cdot w' \cdot c \cdot k \cdot k \quad (3)$$

而 GhostConv-Shuffle 卷积模块的 FLOPs 如式(4)所示。

$$FLOP_{s^{GS}} = \frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + (s-1) \cdot$$

$$\frac{n}{s} \cdot h' \cdot w' \cdot d \cdot d \quad (4)$$

于是, 压缩比可以表示为式(5)

$$r_s = \frac{FLOP_{s^{SC}}}{FLOP_{s^{GS}}} = \frac{n \cdot h' \cdot w' \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot h' \cdot w' \cdot d \cdot d} = \frac{c \cdot k \cdot k}{\frac{1}{s} \cdot c \cdot k \cdot k + \frac{s-1}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \quad (5)$$

假设线性卷积核大小 $d \times d$ 与标准卷积核大小 $k \times k$ 相似, 而 $s \ll c$, 则 GhostConv 模块的计算量仅为标准卷积模块的 $\frac{1}{s}$, 从理论上证明了 GhostConv-Shuffle 的优越性。

1.3 快速空间金字塔池化模块

SPPCSPC 空间金字塔池化层, 能够将任意大小的特征图转换成固定大小的特征向量, 它不同尺度下对特征图进行池化, 以提取具有不同空间范围的特征, 提升网络的感受野。本文使用一种效率更高的 SPPCSPC 结构, 替换原始 YOLOv7 网络中的 SPPCSPC 的结构。在 SPPFCSPC 结构中, 使用基于相同大小的最大池化层的 SPPF 结构, SPPCSPC 结构和 SPPFCSPC 结构分别如图 3 所示。

SPPCSPC 结构中, 使用了三个具有不同大小池化内核的独立池化层来计算空间金字塔池化结

构, 三次池化层的输入相同, 于是可以将 SPPCSPC 中的池化结构表达如式(6)。

$$Y = \text{Maxpool}_5(X) \oplus \text{Maxpool}_9(X) \oplus \text{Maxpool}_{13}(X) \quad (6)$$

其中, Y 表示池化后的输出, $\oplus$ 代表张量缝合运算, X 代表输入特征图。在 SPPCSPC 中, 使用三个尺寸相同的池化核, 并将上一层池化的输出作为下一层池化的输入, 最后将三层池化的结果拼接在一起。这种优化结构在保持感受野不变的同时, 加速了计算过程, 并得到更具代表性和更准确的池化结果。它的池化结构如式(7)所示。

$$\begin{cases} Y_1 = \text{Maxpool}_5(X) \\ Y_2 = \text{Maxpool}_5(Y_1) \\ Y_3 = \text{Maxpool}_5(Y_2) \\ Y = Y_1 \oplus Y_2 \oplus Y_3 \oplus X \end{cases} \quad (7)$$

其中, Y_1 、 Y_2 和 Y_3 分别表示第一层、第二层和第三层的池化输出, X 表示原始特征图, Y 表示张量连接后的池化结果。

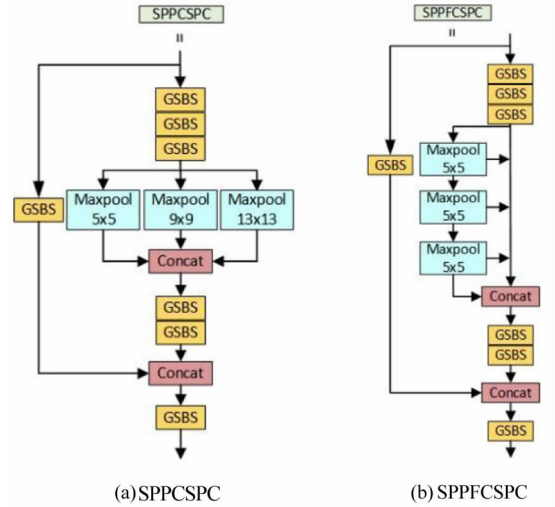


图 3 SPPCSPC 和 SPPFCSPC 结构

1.4 HECA-Net 注意力机制

在 YOLOv7 中, 注意力机制被广泛应用于加强网络的特征表达能力。然而, 传统的注意力机制往往会引入过多的网络复杂度, 导致网络的运行速度变慢。ECA-Net (efficient channel attention net) 作为一种轻量级的注意力机制, 可以在不增加额外计算负担的情况下显著提高特征表达能力, 从而提高目标检测的准确性^[9]。使用 H-Swish 激活函数替换 Sigmoid 激活函数, 提出 HECA-Net 注意力机制, 它的结构如图 4 所示。

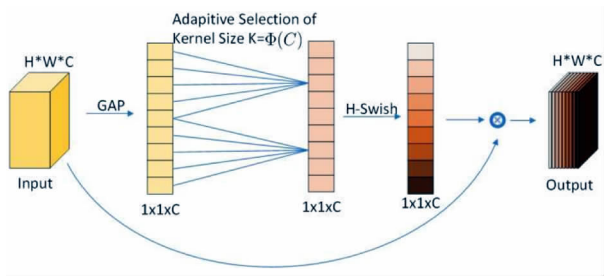


图 4 HECA-Net 网络结构

首先原始特征图在不降低维度的情况下,进行逐通道的全局平均池化,然后通过内核大小为 K 的快速的一维卷积,值得注意的是,在 ECA-Net 中内核 K 的大小决定了交互的覆盖范围,不同的通道数和不同的卷积神经网络架构的卷积块可能会有所不同。 K 与通道维数 C 有关,通道尺寸越大,得到的全连接层越长,长期交互作用越强,而通道尺寸越小,短期交互越强,即存在通道维数 C 和卷积内核 K 的某种关系映射,ECA-Net 使用一种自适应的卷积核大小,它的表达式如式(8)所示。

$$k = \Phi(C) = \left\lceil \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rceil_{\text{odd}}$$

(8)

其中, $\lceil \cdot \rceil_{\text{odd}}$ 表示取最近的奇数运算,ECA-Net 中, γ 和 b 分别设置为 2 和 1。然后将卷积后的结果通过 H-Swish 激活函数^[10],它的表达式如式(9)所示。

$$f(x) = x \cdot \min[\max(0, x + 3), 6]/6$$

(9)

在此选择 H-Swish 作为激活函数主要有以下几个原因。

(1)Sigmoid 函数的计算涉及指数运算,因此计算速度较慢。而 H-Swish 函数中的乘法和加法运算非常简单,计算速度更快。

(2)Sigmoid 函数在输入接近 0 或 1 时,梯度会变得非常小,这种现象称为梯度消失。而 H-Swish 函数的梯度没有这个问题,因此在深度神经网络中使用 H-Swish 可以更容易地进行训练。

(3)H-Swish 与 Swish^[11] 相比,在保持相同计算成本的情况下,具有更好的性能表现。在许多图像分类任务中,使用 H-Swish 可以获得更高的准确率。

将原始特征逐通道与经过 H-Swish 激活函数后权值相乘,得到输出特征。通过添加 HECA-Net 注意力机制,可以提高深度卷积神经网络的特征提取性能。

2 实验数据集与模型评估指标

2.1 实验数据集

NEU-DET 钢材缺陷检测数据集是美国东北大学的研究人员开发的一个用于钢材缺陷检测的数据集^[12]。该数据集包含 6 个不同类型的钢材表面缺陷,即裂痕、夹杂、斑点、点蚀、锈蚀和划痕。数据集中的每个图像都是由真实的钢材表面图像拍摄而来,并且都经过了专业人员的手工标注,标注了每个缺陷的位置和类型。该数据集包含 1800 张灰度,图片分辨率为 200×200 ,每种缺陷包含 300 张图片。为了满足实验要求,将数据集按照 7 : 2 : 1 的比例分别划分为训练集、验证集和测试集。相关数据集的示例图片如图 5 所示。根据图片,光照及制造过程的影响导致钢材表面缺陷具有相似性和多样性,因此在缺陷检测中容易出现漏检和误检的问题。

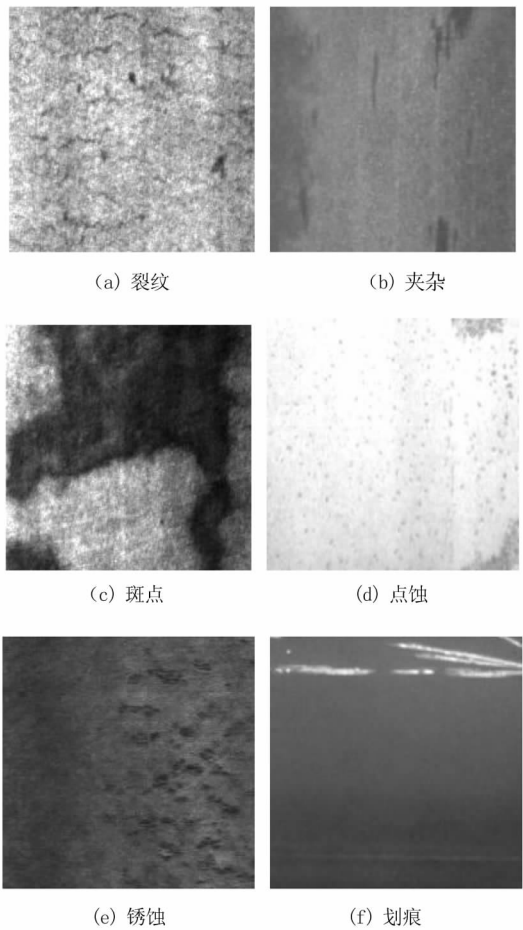


图 5 数据集示例图片

2.2 模型评估指标

为了评估模型的有效性,采用精度(P)、召回率(R)、平均精确度(mAP)作为模型性能的评价指标。其中,精确度表示被模型预测为正类的样本中,真正为正类的样本数占有所有被预测为正类的样本数的比例;召回率表示真正为正类的样本中,被模型预测为正类的样本数占有所有真正为正类的样本数的比例;AP 表示对于单个目标的预测性能,考虑了检测器在不同的召回率下的表现;平均精确度表示所有类别中 AP 的平均值,根据交并比的不同可以分为 mAP@0.5 和 mAP@0.5:0.95,它们的计算如式(10)所示。

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN}$$
$$AP = \int_0^1 P(R) dR$$
$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

(10)

其中,TP 表示模型正确预测为正例的样本数;FP 表示模型错误地预测为正例的样本数;FN 表示模型错误地预测为反例的样本数;AP 为不同召回率下精度的积分;N 为检测到的所有类被数。

3 实验结果与分析

3.1 实验设置

在训练过程中,采用 Adam 优化算法,动量因子为 0.937,同时将初始学习率设置为 0.1,权重衰减系数为 0.0005,输入图像尺寸为 640 × 640,单词训练图像数量为 4,将 IoU 阈值设置为 0.25,最大迭代轮次为 300,训练平台配置如表 1 所示。

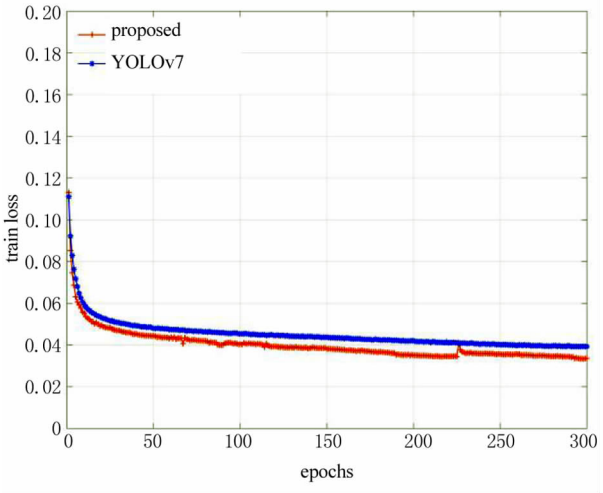
表 1 训练平台配置

名称	参数
处理器	Intel Core i5-12400F
显卡	NVIDIA GeForce RTX 3060Ti
操作系统	Windows10
内存	32G
语言	Python 3.8
CUDA	CUDA 11.6
CuDNN	CuDNN 10.0
Matlab	Matlab2022a

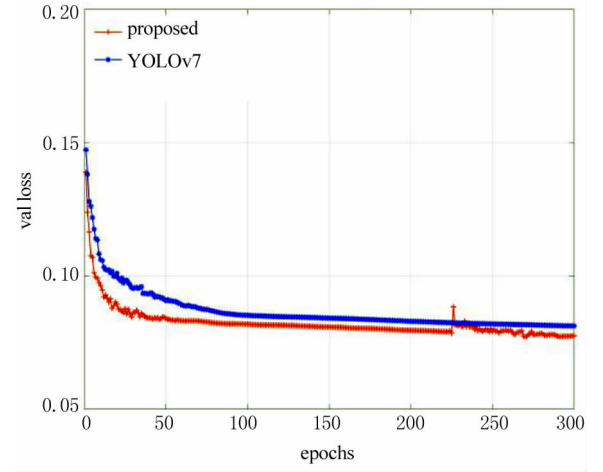
3.2 训练过程与分析

在相同的实验设置下,分别对 AFSD-

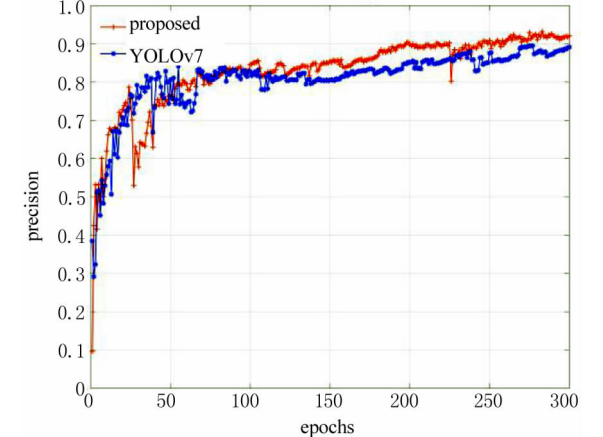
YOLOv7 和 YOLOv7 进行训练,在训练过程中记录训练数据,使用 Matlab 对训练数据进行解析。训练集损失、验证集损失、精度、召回率、mAP@.5 如图 6 所示。



(a) 训练集损失



(b) 验证集损失



(c) 精度

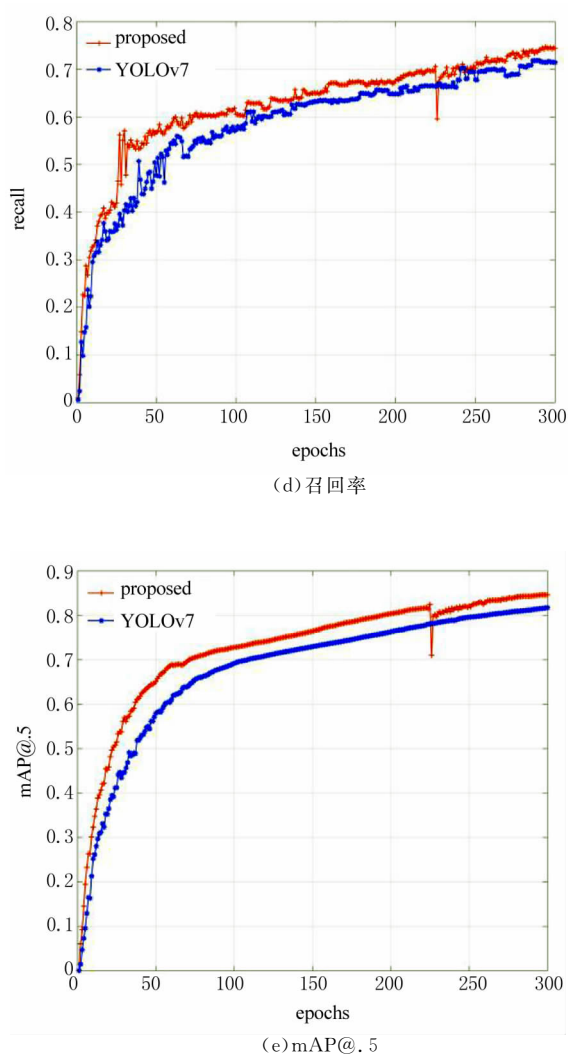


图 6 各项指标对比

表 2 不同算法下缺陷种类检测精度对比(%)

算法	裂痕	夹杂	斑点	点蚀	锈蚀	划痕
YOLOv4	34.5	83.7	91.6	87.2	66.8	92.5
YOLOv5s	35.2	82.1	89.5	84.5	63.4	90.4
YOLOv7	43.6	83.9	93.5	88.3	68.5	94.6
YOLOv7-e6	45.7	84.2	95.6	90.4	69.7	94.8
YOLOv7-d6	47.2	86.4	95.2	93.2	71.3	96.2
AFSD-YOLOv7	48.6	87.3	96.4	94.8	73.5	97.4

表 3 不同算法下模型性能对比

算法	mAP@.5/%	FPS/(帧/s)	FLOPs/G	模型大小/MB
YOLOv4	76.1	51.4	198.3	248
YOLOv5s	74.2	56.9	16.5	28.7
YOLOv7	78.7	60.2	105.3	73.8
YOLOv7-e6	80.1	26.7	144.8	190.8
YOLOv7-d6	81.6	17.2	198.8	261.9
AFSD-YOLOv7	82.9	87.2	27.6	33.5

3.4 模型检测效果比较

本节旨在通过展示 AFSD-YOLOv7 算法在实际图片中的检测效果来印证实验结果。选择了代表

随着训练轮次的迭代,模型的各项指标逐渐趋于平稳,在训练集损失和验证集损失上,AFSD-YOLOv7 相较于 YOLOv7 损失值较少;对于精度、召回率、mAP@.5 指标,AFSD-YOLOv7 相较于 YOLOv7 显著增加,证明了 AFSD-YOLOv7 具有更好的精确性和优越性。

3.3 网络性能比较

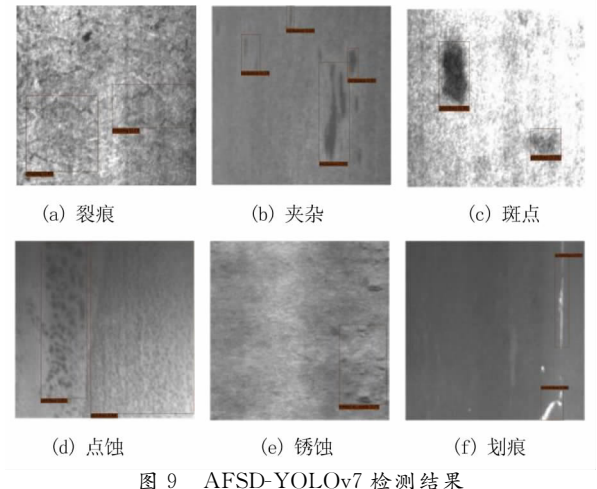
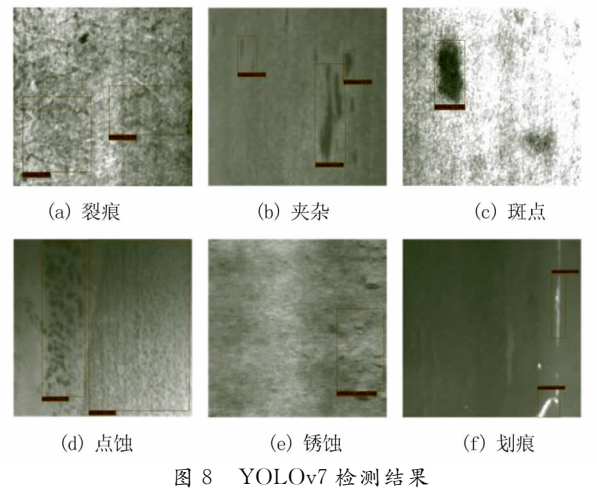
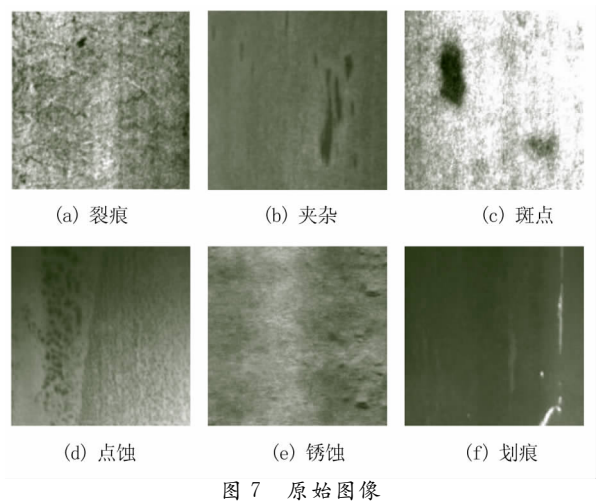
为进一步评估 AFSD-YOLOv7 模型的性能,在实验设置和数据集划分相同的条件下,选择 YOLOv7、YOLOv7-d6、YOLOv7-e6^[13]、YOLOv5s^[14]、YOLOv4^[15] 目标检测算法进行对比实验,实验结果如表 2、表 3 所示。

表 2 显示,AFSD-YOLOv7 模型在不同类型的缺陷检测方面均表现最佳,包括裂纹、夹杂、斑点、点蚀、锈蚀和划痕。具体而言,AFSD-YOLOv7 算法在所有缺陷类型的检测准确性方面都表现出最高的性能,例如裂痕的检测准确率为 48.6%,夹杂为 87.3%,斑点为 96.4%。

表 3 中的比较结果表明,所提出的 AFSD-YOLOv7 算法在模型复杂度、检测准确度和实时性能方面均表现最佳。该算法通过提高 mAP@.5 达到更好的检测准确度,减少模型大小达到更低的复杂性,提高 FPS 以实现更好的实时性能。其中,与原始 YOLOv7 相比,mAP@.5 提高了 3.2%,模型大小减少了 54.8%。

性的、包含不同种类缺陷的图片,并将它们分别用 AFSD-YOLOv7 和 YOLOv7 算法进行检测。通过比较不同算法的检测结果,进一步验证了 AFSD-

YOLOv7 算法的性能优势,如图 7~图 9 所示。



对比图 8 和图 9 可知, YOLOv7 存在着部分漏检、定位效果差以及精度不够的问题: ①如图 8 (b)和图 9(b)所示, AFSD-YOLOv7 能够精确定位夹杂, 而 YOLOv7 存在漏检在边缘处的夹杂的问题。②如图 8(c)所示, YOLOv7 未能识别出右下角的斑点。③如图 8(f)所示, YOLOv7 的划痕定

位效果不佳。综上所述, AFSD-YOLOv7 算法相比 YOLOv7 具有更高的精度、更准确的定位能力以及更好的检测效果。

4 结 论

钢材表面缺陷对钢材外观和性能带来巨大影响, 例如裂纹、气泡、夹杂、疤痕等会降低钢材的强度、韧性和塑性等力学性能, 进而影响其使用寿命和安全性。本文提出了 AFSD-YOLOv7 钢材表面缺陷检测算法, 该算法使用一种轻量级卷积结构替换标准卷积结构, 使用更高效的空间金字塔池化结构, 最后添加改进的 ECA-Net 注意力机制, 进一步提升模型的检测精度。将 AFSD-YOLOv7 与一些经典的目标检测算法对比, 突出了本文缺陷检测精度高、实时性强、定位精度好等优点。本文提出的算法对于保障钢材性能和品质具有深刻的意义。

参考文献

[1] 刘晓妍, 田琳, 边振虎. 合理应用连铸用耐火材料以提高钢材质量[J]. 耐火与石灰, 2022(5): 33-38+48.

[2] 王元清, 张元元, 石永久. 钢材缺陷及焊接对 Z 向性能的影响分析[J]. 低温建筑技术, 2013(7): 1-4.

[3] 李应果, 杨洁, 董春雷. 木材表面缺陷特征轮廓提取算法研究[J]. 西北林学院学报, 2021(4): 204-208+281.

[4] 石冬阳, 张俊林, 杨欣, 等. 一种基于 Canny 边缘检测算法的零件缺陷图像处理办法[J]. 重庆科技学院学报(自然科学版), 2022(3): 52-56.

[5] 周瑞旭, 高文胜, 刘卫东. 基于图像处理—纹理分析的 HVDC GIL 缺陷识别方法[J]. 高电压技术, 2022(12): 4719-4730.

[6] 赵京鹤, 修大元, 王金龙, 等. 绝缘纸板典型缺陷局放特征的 CNN 识别[J]. 哈尔滨理工大学学报, 2022(5): 79-88.

[7] WU D, JIANG S, ZHAO E, et al. Detection of camellia oleifera fruit in complex scenes by using YOLOv7 and data augmentation [J]. Applied Science, 2022, 12: 11318.

[8] LIU Y, LV Z, HU Y, et al. Improved cotton seed breakage detection based on YOLOv5s [J]. Agriculture, 2022, 12: 1630.

[9] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[10] CHEN L, SONG Y, ZHANG H. Extraction of impervious surface from high-resolution remote sensing image based on deep learning[C]. In 2021 International Conference on Wireless Communications and Smart Grid (ICWCSG), 2021.

[11] EGER S, YOUSSEF P, GUREVYCH I. Is it time to swish? comparing deep learning activation functions across NLP tasks [C]. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

[12] HE Y, SONG K, MENG Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493-1504.

[13] WANG C Y, Bochkovskiy A, LIAO H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv e-prints, 2022.

[14] YING Z P, LIN Z T, WU Z Y, et al. A modified YOLOv5s model for detection of wire braided hose defects[J]. Measurement, 2022(190): 1-11.

[15] 陈为, 钟欣童, 张婧, 等. 基于 YOLOv4 的航空发动机叶片凸台目标检测[J]. 计算机仿真, 2022, 39(7): 5.