



# 计算机科学

COMPUTER SCIENCE

## 改进YOLOv5s的轻量化钢材表面缺陷检测模型

蒋博, 万毅, 谢显中

引用本文

蒋博, 万毅, 谢显中. [改进YOLOv5s的轻量化钢材表面缺陷检测模型](#)[J]. 计算机科学, 2023, 50(11A): 230900113-7.

JIANG Bo, WAN Yi, XIE Xianzhong. [Improved YOLOv5s Lightweight Steel Surface Defect Detection Model](#) [J]. Computer Science, 2023, 50(11A): 230900113-7.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于改进YOLOv4的磁瓦缺陷检测算法](#)

Magnetic Tile Defect Detection Algorithm Based on Improved YOLOv4

计算机科学, 2023, 50(11A): 230100100-7. <https://doi.org/10.11896/jsjcx.230100100>

### [基于机器视觉的超声相控阵缺陷检测研究](#)

Study on Ultrasonic Phased Array Defect Detection Based on Machine Vision

计算机科学, 2023, 50(11A): 230200150-6. <https://doi.org/10.11896/jsjcx.230200150>

### [一种面向工业产品表面缺陷图像的色调增强方法](#)

Hue Augmentation Method for Industrial Product Surface Defect Images

计算机科学, 2023, 50(11A): 230200089-6. <https://doi.org/10.11896/jsjcx.230200089>

### [基于改进的YOLOv5s列车轨道线检测方法](#)

Railway Track Detection Method Based on Improved YOLOv5s

计算机科学, 2023, 50(11A): 230200101-6. <https://doi.org/10.11896/jsjcx.230200101>

### [基于多尺度改进的YOLOv5电解槽设备及样品检测方法](#)

Electrolyzer Equipment and Sample Detection Method Based on Multi-scale Improved YOLOv5

计算机科学, 2023, 50(11A): 230200163-6. <https://doi.org/10.11896/jsjcx.230200163>

# 改进 YOLOv5s 的轻量化钢材表面缺陷检测模型

蒋博<sup>1</sup> 万毅<sup>1</sup> 谢显中<sup>2</sup>

1 重庆邮电大学通信与信息工程学院 重庆 400065

2 重庆邮电大学研究生院 重庆 400065

**摘要** 针对现有钢材表面缺陷检测模型结构复杂、参数量多、检测精度和实时性较差等问题,提出了一种改进 YOLOv5s 的轻量化钢材表面缺陷检测模型。首先采用 MobileNetv3-Small 网络替换 YOLOv5s 主干提取网络,实现模型轻量化,提升检测速度;其次在特征融合阶段采用加权双向特征金字塔网络(Bidirectional Feature Pyramid Network, BiFPN)加强特征提取,通过融合不同尺度的特征,提升检测的准确率和鲁棒性。同时引入 CBAM(Convolutional Block Attention Module)注意力机制增强模型对小尺度目标的检测能力;最后使用 K-means++ 算法聚类先验框,提高先验框聚类的准确性和收敛速度。改进后的模型在 NEU-DET 数据集上的平均精度均值(mAP@0.5)达到 77.2%,在 NVIDIA 1080Ti 上检测速度达到 102FPS。相较于原始 YOLOv5s 模型,mAP 提升 3.90%,参数量减少 58.6%,体积减小 34%,检测速度提升 29.7%。实验结果表明改进的 YOLOv5s 模型在保证轻量化的同时能够有效提升钢材表面缺陷检测的精度和速度,易于部署,满足带钢实际生产中的需求。

**关键词:** 缺陷检测; YOLOv5s; 轻量化; MobileNetv3-Small; BiFPN; CBAM; K-means++

中图分类号 TP391

## Improved YOLOv5s Lightweight Steel Surface Defect Detection Model

JIANG Bo<sup>1</sup>, WAN Yi<sup>1</sup> and XIE Xianzhong<sup>2</sup>

1 School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2 Graduate School, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

**Abstract** Aiming at the problems of complex structure, large number of parameters, poor detection accuracy and real-time performance of existing steel surface defect detection models, this paper proposes an improved YOLOv5s lightweight steel surface defect detection model. Firstly, the MobileNetv3-Small is used to replace the YOLOv5s backbone extraction network, achieving model lightweight and improving detection speed. Secondly, in the feature fusion stage, a weighted bidirectional feature pyramid network (BiFPN) is used to enhance feature extraction. By fusing features of different scales, the accuracy and robustness of detection are improved. Simultaneously, the convolutional block attention module (CBAM) attention mechanism is introduced to enhance the model's ability to detect small scale targets. Finally, the K-means++ algorithm is proposed to cluster prior boxes, improve the accuracy and convergence speed of prior box clustering. The average accuracy of the improved YOLOv5s on the NEU-DET dataset (mAP@0.5) reaches 77.2%, with a detection speed of 102 FPS on NVIDIA 1080Ti. Compared to the original YOLOv5s, the mAP is increased by 3.90%, the parameter quantity is decreased by 58.6%, the volume is decreased by 34%, and the detection speed is increased by 29.7%. Experimental results demonstrate that the improved lightweight YOLOv5s effectively improves both the accuracy and speed of steel surface defect detection. Moreover, it is easy to deploy and meet the requirements of actual production in the steel strip industry.

**Keywords** Defect detection, YOLOv5s, lightweight, MobileNetv3-Small, BiFPN, CBAM, K-means++

## 1 引言

钢材表面缺陷检测是工业生产中至关重要的一项检测工作,主要包含对缺陷的识别和定位。钢材表面的缺陷检测算法可以分为传统的缺陷检测算法和基于深度学习的缺陷检测算法。但传统的表面缺陷检测算法存在人工误判和漏检等问题,需要耗费大量的时间和人力成本,同时也难以满足高质量和高效率的要求。而基于深度学习的表面缺陷检测算法可以利用卷积神经网络(CNN)对图像进行特征提取,从而对缺陷实现更快速、更精准的识别与定位<sup>[1]</sup>,已成为缺陷检测领域的

主流,得到了广泛的应用。

目前,基于深度学习的表面缺陷检测算法包括两阶段和单阶段检测算法。两阶段算法包括特征提取和区域生成两个过程,其中,特征提取部分使用卷积神经网络,区域生成网络使用回归器和分类器进行检测和定位。典型的算法有 Faster R-CNN<sup>[2]</sup>, R-FCN<sup>[3]</sup>, Mask R-CNN<sup>[4]</sup>等。但由于算法受限于两个阶段的处理,检测速度较慢,模型体积较大,不易于部署在边缘端设备和嵌入式系统,难以满足钢材缺陷检测的实际需求。

单阶段算法采用单个神经网络模型同时进行目标位置和

类别的回归,检测速度更快,网络结构也更简洁,典型的算法有 SSD<sup>[5]</sup>和 YOLO<sup>[6-10]</sup>系列。目前大量研究工作将 YOLOv5 算法用于表面缺陷检测。文献[11]基于 YOLOv5s 提出目标检测算法(RFBG-YOLO),引入 GhostConv 卷积减小模型参数,实现模型轻量化,但改进后的模型检测速度有所降低。文献[12]基于 YOLOv5 算法,构造了一种 SPP\_Res 特征金字塔结构,同时加入了多头注意力机制(C3\_MHSA),改进的算法有效提升了模型的整体检测精度,但算法对部分小尺度目标,如裂纹类(Crazing)的检测效果仍不理想。文献[13]使用 GhostBottleneck 结构实现 YOLOv5 模型轻量化,其次在 Backbone 部分引入 SE 注意力机制,使模型整体的检测速度和精度都得到提升,但该算法对于裂纹类小尺度目标的检测精度仍然较低。文献[14]基于 YOLOv5 算法,引入 Transformer encoder block 模块和 CBAM 注意力机制,提升了检测的精度,但该算法引入较多模块导致模型体积和参数量较大,无法满足实时检测的需求。基于以上分析及存在的问题,为了满足带钢实际生产中的需求,本文对 YOLOv5s 进行改进,首先采用 MobileNetv3-Small 网络替换 YOLOv5s 主干提取网络,大幅减小模型参数量,实现模型轻量化,有效提升检测速度。特征融合阶段采用加权双向特征金字塔网络(BiFPN),以更好地捕捉不同尺度的特征,提供更全面和丰富的特征表示。同时引入 CBAM 注意力机制,提升模型的检测精度。最后使用 K-means++ 算法聚类先验框,在加快收敛速度的同时降低聚类结果的误差。最终改进的 YOLOv5s 模型更轻量化,易于部署,有效解决了检测速度和精度难以平衡的问题,为带钢实际生产提供了可参考的方案。

## 2 YOLOv5s 模型介绍

YOLOv5s 模型结构如图 1 所示。YOLOv5s 包括 Input, Backbone, Neck 和 Head 4 部分。

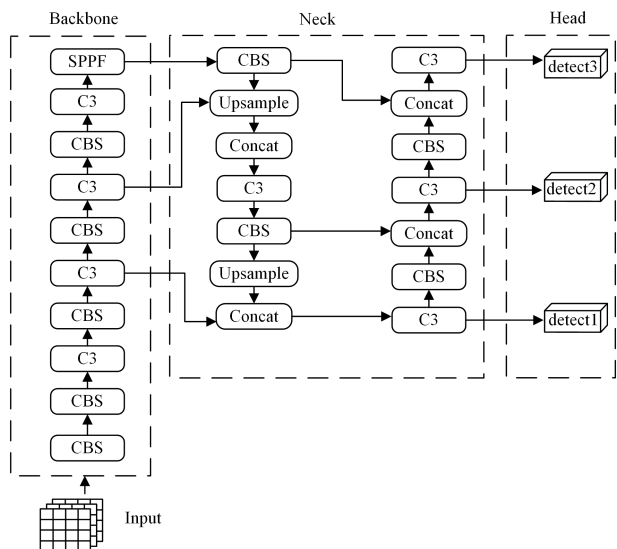


图 1 YOLOv5s 结构图

Fig. 1 Structure diagram of YOLOv5s

Input 部分进行图像预处理,输入图像大小为  $640 \times 640$ ,采用 Mosaic 数据增强、自适应锚框和自适应图像缩放等技术来处理输入数据。Backbone 部分由卷积模块(CBS)、瓶颈层(C3)、快速空间金字塔池化(Spatial Pyramid Pooling-Fast, SPPF)模块构成,用于提取不同细粒度的图像特征。Neck 部分采用了特征图金字塔网络(Feature Pyramid Networks,

FPN)和路径聚合网络(Path Aggregation Network, PAN)组合而成的 PANet 网络。具体地,FPN 自顶向下将深层特征图的语义信息传递给浅层,增强了多尺度特征图的语义表达,提高了检测精度。而 PAN 则自底向上将浅层特征图的位置信息传递给深层,增强了多尺度特征图的定位能力,提高了检测的鲁棒性和准确性。Head 部分包括 3 个尺寸为  $20 \times 20$ ,  $40 \times 40$ ,  $80 \times 80$  大小的检测分支,分别用于检测大、中、小目标。通过非极大值抑制(Non Maximum Suppression, NMS)算法来获得最优目标框,用于检测物体的位置和类别。边界框采用了 CIUO 损失函数。

YOLOv5s 虽然在检测精度上表现出了不俗的性能,但对于特定的钢材表面缺陷检测仍存在诸多问题。钢材表面缺陷通常是细小且复杂的纹理、裂纹等,对于这类小尺寸目标,YOLOv5s 可能存在一定的检测困难。而 C3 模块的结构复杂,计算复杂度较高,导致检测速度较慢。

## 3 YOLOv5s 模型改进

### 3.1 轻量化主干提取网络

为降低 YOLOv5s 计算的复杂度、减小模型体积,采用 MobileNetv3-Small 代替 YOLOv5s 主干提取网络,减少模型参数量,实现模型轻量化。

MobileNetv3<sup>[15]</sup>继承了 MobileNetv1<sup>[16]</sup>的深度可分离卷积和 MobileNetv2<sup>[17]</sup>的逆残差结构,并在此基础上提出了 SE (Squeeze-and-Excitation)注意力机制,同时采用 h-swish 激活函数。MobileNetv3 共有 Large 和 Small 两种结构,其中 Small 结构更轻量、具有更高的运行速度和更小的计算复杂度。因此,本文选用 Small 结构对 YOLOv5s 网络进行改进。Small 结构主要由 Bneck 结构组成,Bneck 结构如图 2 所示。它很好地融合了深度可分离卷积和逆残差结构。在 Bneck 结构中,首先通过  $1 \times 1$  大小的卷积升维,接着采用  $3 \times 3$  大小的深度可分离卷积提取特征信息;然后通过 SE 注意力机制调整每个通道的权重;最后使用  $1 \times 1$  的卷积实现降维,并将降维后的输出与残差边部分进行求和。这样先升维后降维的操作更有利于网络对小尺度目标的特征提取,从而提高检测精度。

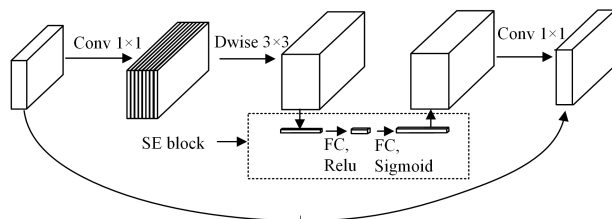


图 2 Bneck 结构

Fig. 2 Bneck structure

#### 3.1.1 深度可分离卷积

MobileNetv3-Small 网络主要采用深度可分离卷积(Depthwise Separable Convolution)提取特征,与普通卷积相比,它可以大幅减少网络的参数量。深度可分离卷积由逐通道卷积(DW)和逐点卷积(PW)组成,逐通道卷积对每个通道上的像素点分别卷积获得各自的特征,卷积核大小为  $3 \times 3$ ,卷积核的数量等于输入通道数。逐点卷积使用  $1 \times 1$  大小的卷积核在每个像素点上对不同通道上的特征进行卷积。标准卷积和深度可分离卷积分别如图 3 和图 4 所示。由图 3 可知,

假设输入特征图长为  $H$ , 宽为  $W$ , 有  $M$  个输入通道, 输出通道数为  $N$ , 卷积核大小为  $D_K \times D_K$ , 则标准卷积的计算量为:

$$S = H \times W \times M \times N \times D_K \times D_K \quad (1)$$

由图 4 可得, 深度可分离卷积的计算量为:

$$P = H \times W \times M \times D_K \times D_K + H \times W \times M \times N \quad (2)$$

可以得到深度可分离卷积与标准卷积的计算量比值如下:

$$\frac{P}{S} = \frac{1}{N} + \frac{1}{D_K^2} \quad (3)$$

由式(3)可知, 深度可分离卷积使得网络的整体计算量和参数量大幅减少, 从而实现了模型轻量化, 提高了模型的检测速度。

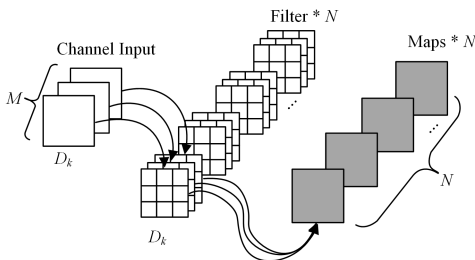


图 3 标准卷积

Fig. 3 Standard convolution

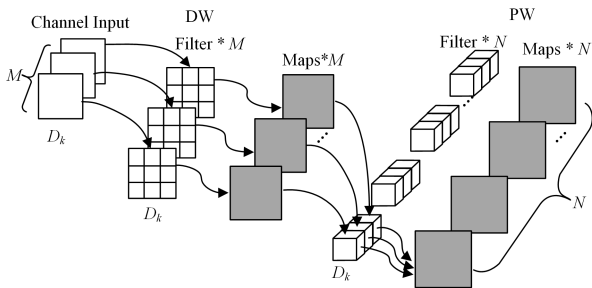


图 4 深度可分离卷积

Fig. 4 Deep separable convolution

### 3.1.2 SE 注意力机制

SE 注意力机制主要包括 Squeeze, Excitation 和 Scale 操作。如图 5 所示, Squeeze 操作对输入大小为  $H \times W \times C_2$  的特征图进行全局平均池化, 将其压缩为一个  $1 \times 1 \times C_2$  的向量。Excitation 操作则将 Squeeze 的结果通过全连接层和激活函数得到每个通道注意力权重。最后, Scale 操作将通道注意力权重作用在原始特征上, 得到经过注意力机制加权的特征。SE 注意力机制具有简单、紧凑的结构, 它的计算开销相对较小, 不会引入过多的额外参数, 因此可以更高效地使用计算资源。

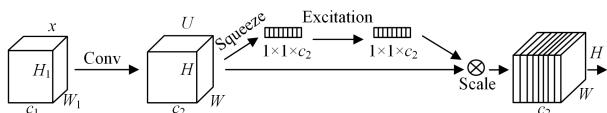


图 5 SE 注意力机制

Fig. 5 SE attention mechanism

使用 MobileNetv3-Small 网络替换 YOLOv5s 主干提取网络后, 主干提取网络详细参数如表 1 所列。From 表示当前模块来自那一层, -1 代表来自上一层, Number 代表该模块的深度, Operator 表示模块的具体操作。Conv 列和 SPPF 列对应的 Args 列的值分别表示该模块的输入通道数、输出通道

数和步长。Bneck 对应的 Args 列的值分别表示该模块的输出通道数、扩张通道数、卷积核的大小、步长、是否使用 SE 注意力机制以及是否使用 h-swish 激活函数。其中, 1 代表使用, 0 代表不使用。

表 1 主干提取网络结构

Table 1 Backbone extraction network structure

From	Number	Operator	Args
-1	1	Conv	[16,2]
-1	1	Bneck	[16,16,3,2,1,0]
-1	1	Bneck	[24,72,3,2,0,0]
-1	1	Bneck	[24,88,3,1,0,0]
-1	1	Bneck	[40,96,5,2,1,1]
-1	1	Bneck	[40,240,5,1,1,1]
-1	1	Bneck	[40,240,5,1,1,1]
-1	1	Bneck	[48,120,5,1,1,1]
-1	1	Bneck	[48,144,5,1,1,1]
-1	1	Bneck	[96,288,5,2,1,1]
-1	1	Bneck	[96,576,5,1,1,1]
-1	1	SPPF	[576,5]

### 3.2 特征融合网络

YOLOv5s 通过 PANet 网络实现特征融合, PANet 使用一系列的卷积核池化操作来聚合不同尺度的特征图, 以提取更全面的上下文信息, 增强特征的语义表达能力。但 PANet 网络融合过程中的信息均来自于前一阶段的处理, 这使得融合后的特征并没有很好地利用原始特征信息。这不利于对小尺度目标特征的提取。为了解决这一问题, 本文借鉴 BiFPN 网络的思想改进 YOLOv5s, 以提高模型的检测精度。

BiFPN<sup>[18]</sup> 网络使用特征金字塔的思想, 能够同时利用来自不同尺度的特征信息, 通过在网络中添加并行分支来处理不同尺度的特征。该网络同时具备双向特征融合能力, 在特征金字塔的每个分支上进行下采样和上采样操作。这使得低分辨率特征和高分辨率特征可以相互影响, 有效地捕捉目标的细节信息和跨尺度特征, 且可以有效地提高检测的准确性和鲁棒性。PANet 和 BiFPN 网络结构如图 6 所示。BiFPN 通过在 PANet 的基础上添加额外的边实现特征的传递和融合。通过这种方式, 可以扩大感受野, 提高目标检测的性能。

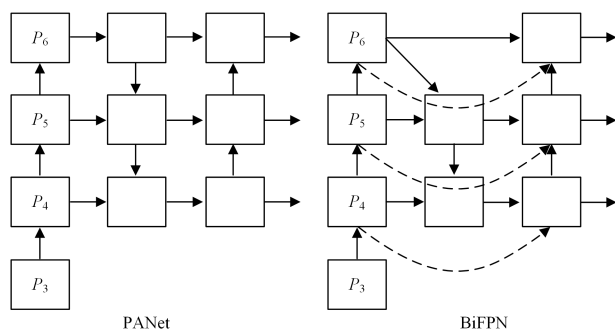


图 6 特征融合网络结构

Fig. 6 Feature fusion network structure

### 3.3 CBAM 注意力机制

在特征提取过程中, YOLOv5s 模型对不同重要程度的特征采取相同的加权方式, 这样不利于对小尺度目标信息的提取。为解决这一问题, 引入了 CBAM 注意力机制。CBAM<sup>[19]</sup> 注意力机制结合了空间注意力模块和通道注意力模块, 主要目的是提高神经网络的特征表达能力和泛化能力。如图 7 所示, CBAM 注意力机制首先通过通道注意力模块对不同特征进行加权, 然后将加权后的特征送入空间注意力模块, 得到最后的结果。

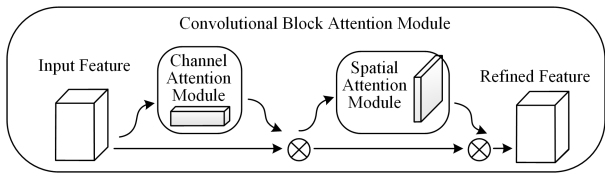


图7 CBAM 注意力机制

Fig. 7 CBAM attention mechanism

最终改进后的 YOLOv5s 模型的结构如图 8 所示。

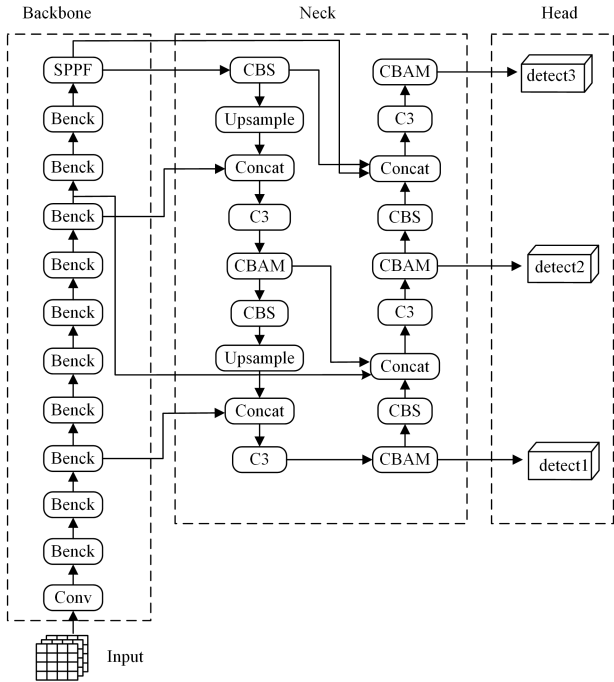


图8 改进后的 YOLOv5s 模型结构

Fig. 8 Improved YOLOv5s model structure

### 3.4 K-means++ 算法

YOLOv5s 采用  $K$ -means<sup>[20]</sup> 算法对先验框进行聚类,  $K$ -means 算法通过随机指定聚类中心, 很难避免聚类中心的相似性问题, 并不完全适合于本文数据集。因此, 本文选用  $K$ -means++<sup>[21]</sup> 算法对先验框进行聚类,  $K$ -means++ 算法能更好地选择初始聚类中心, 得到更优的先验框尺寸。具体步骤如下:

- (1) 随机选择数据集中的一个样本作为第一个聚类中心;
- (2) 对于数据集中的每个样本, 计算它与当前已选择的聚类中心的最近距离  $D(X)$ ;
- (3) 根据概率分布函数  $P$ , 以概率  $P$  选择下一个聚类中心, 函数表达式如式(4)所示:

$$P = \frac{D(X)^2}{\sum_{x \in X} D(X)^2} \quad (4)$$

- (4) 重复步骤(2)和步骤(3)直到选择出  $K$  个聚类中心为止。

根据得到的  $K$  个聚类中心, 使用  $K$ -means++ 算法进行迭代得到最终结果, 将每个样本点和聚类中心 IoU(交并比)的最大值的平均作为评价最终聚类效果的标准, 计算方法如式(5)所示:

$$Accuracy = \frac{1}{n} \sum_{i=1}^n \max \left( \frac{box_i \cap center_1}{box_i \cup center_1}, \dots, \frac{box_i \cap center_k}{box_i \cup center_k} \right) \quad (5)$$

其中,  $box$  为预测框,  $center$  为真实框。数据集聚类中心  $K$  的取值则是由聚类中心的平均交并比确定。如图 9 所示, 数据集聚类中心  $K$  等于 9 时平均交并比达到最大值, 说明此时聚类效果最佳, 因此, 本文数据集的  $K$  值选择 9。

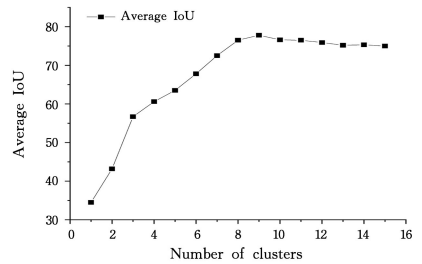


图9 聚类中心数与平均交并比变化曲线

Fig. 9 Change curve of cluster centers and average intersection and union ratio

最终通过  $K$ -means++ 算法聚类的结果如表 2 所列。由表 2 可知, 通过  $K$ -means++ 算法聚类先验框的 Accuracy 达到 77.8%, 相比  $K$ -meas 算法提升了 5.5%。

表2 K-means++ 聚类结果

Table 2 K-means++ clustering results

聚类方法	特征图尺寸	先验框尺寸	Accuracy/%
K-means	80×80	[10,13,16,30,33,23]	72.3
	40×40	[30,61,62,45,59,119]	
	20×20	[116,90,156,198,373,326]	
K-means++	80×80	[12,25,28,47,46,42]	77.8
	40×40	[23,82,48,55,89,110]	
	20×20	[97,118,221,166,354,246]	

## 4 实验与结果分析

### 4.1 实验数据集

带钢是一种窄而长的扁平断面钢材, 在工业领域应用广泛。本文所用数据集是由东北大学发布的带钢表面缺陷数据集 (NEU-DET)<sup>[22]</sup>, 该数据集共包含 6 种类别的缺陷, 分别是裂纹 (crazing)、夹杂物 (inclusion)、斑块 (patches)、点蚀面 (pitted\_surface)、氧化轧皮 (rolled-in\_scale)、划痕 (scratches)。缺陷图片为 200×200 尺寸的灰度图, 每种类型 300 张, 一共 1800 张图片。使用 Labelme 工具对缺陷图片的真实框进行标注, 得到 xml 格式的标签文件。实验按照 8:2 的比例将数据集划分为训练集和测试集, 其中训练集 1440 张, 测试集 360 张。部分数据集示例如图 10 所示。

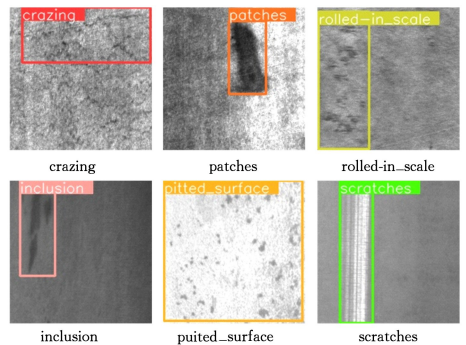


图10 部分数据集示例

Fig. 10 Example of partial dataset

### 4.2 实验环境

本文使用的实验环境如表 3 所列。

表 3 实验环境

Table 3 Experimental environment

名称	参数
GPU	NVIDIA GeForce GTX 1080Ti 11 GB
CPU	Intel Core i7-9700K@ 3.60 GHz
操作系统	Windows10
深度学习框架	Pytorch+cuda10.2+ cudnn8.2.1
编译软件	PyCharm

### 4.3 训练策略

epoch 设置为 300,训练时采用在线增强的方法处理数据集,通过在每个 epoch 中对数据集进行平移、缩放以及饱和度和亮度变换等操作增加训练数据的多样性,进一步提升模型的泛化能力,并减少过拟合的可能性。使用 K-means 十算法聚类初始先验框的大小。batch size 大小设置为 32,图片大小设置为  $640 \times 640$ ,优化器选择 SGD,初始学习率为 0.01。为了充分利用训练过程中的信息,采用余弦退火策略降低学习率,这种策略基于余弦函数的形式调整学习率,加速训练过程并提高模型的准确性。衰减系数设置为 0.0005,学习率动量设置为 0.937,IoU 阈值设置为 0.5。当预测框与真实框的 IoU 超过该阈值时,则认为检测到了目标物体。

### 4.4 评价指标

本文从模型体积(MB)、模型参数量(Parameters)、计算量(GFLOPs)、平均精确度均值 mAP(mean Average Precision)以及每秒传输帧数 FPS(Frames Per Second)这 5 方面来评价模型的性能。模型的体积表示模型训练完成后保存的权重文件的大小;模型参数量表示训练过程中总的参数量;计算量即浮点运算次数,用来衡量模型的复杂度;mAP 表示所有类别的检测准确率的均值,由精确率  $P$ (Precision)、召回率  $R$ (Recall)以及  $AP$  计算得到,其中精确率是指正确预测为正样本占全部预测为正样本的比例,召回率是指正确预测为正样本占全部实际为正样本的比例, $AP$  为各类别的检测精度。 $P, R, AP$  以及 mAP 的计算式如式(6)~式(9)所示:

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$AP = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \quad (9)$$

其中,  $TP$  表示模型预测为正的样本,  $FP$  表示模型预测为正的负样本,  $FN$  表示模型预测为负的正样本。  $N$  为检测的类别数。

### 4.5 实验结果分析

#### 4.5.1 改进前后的模型对比

本文分别对改进前后的 YOLOv5s 模型进行训练,两种模型的检测精度对比如图 11 所示。由图 11 可知,改进后的 YOLOv5s 模型 mAP@0.5 为 77.2%,较 YOLOv5s 提升了 3.9%。6 种缺陷的检测精度均有不同程度的提升。其中裂纹类(crazing)提升最高,达到 7.3%,夹杂物(inclusion)提升 6.6%,斑块(patches)提升 0.9%,点蚀面(pitted\_surface)提升 1.2%,氧化轧皮类(rolled\_in\_scale)提升 2.3%,划痕

(scratches)提升 5.0%。结果显示改进后的模型检测精度更高,对小尺度目标的定位和识别更加准确。

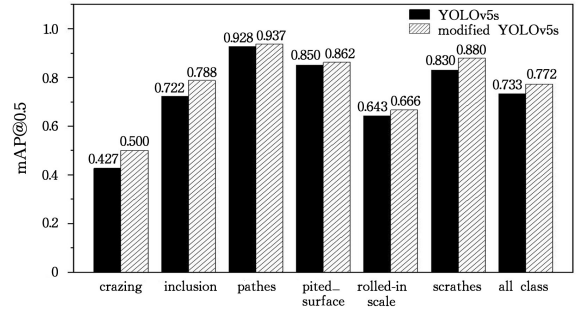


图 11 改进前后的检测精度对比

Fig. 11 Comparison of detection accuracy before and after improvement

本文还对比了模型改进前后的参数量、体积、检测速度和计算量。如表 4 所列。由表 4 可知,改进的 YOLOv5s 模型整体性能优于 YOLOv5s。改进后的模型的参数量较 YOLOv5s 减少 58.6%,模型体积减小 34%,计算量减少 76.8%。检测速度达到 102FPS,较改进前提升了 29.7%。

表 4 改进前后性能对比

Table 4 Performance comparison before and after improvement

模型	参数量	体积/MB	计算量/GFLOPs	FPS
YOLOv5s	7 035 811	22.4	16.0	78.6
改进的 YOLOv5s	2 906 889	14.8	3.7	102.0

#### 4.5.2 模型训练对比

改进前后的模型训练效果如图 12 所示。Total loss 为 box\_loss, obj\_loss 和 cls\_loss 三者之和的平均值。其中 box\_loss 用于监督检测框的回归,计算预测框和真实框之间的误差;obj\_loss 用于监督检测框内是否存在物体,计算网络的置信度;cls\_loss 用于监督物体分类类别,计算先验框与对应的标注类别是否正确。由图 12 可知,改进后的模型更快达到收敛,检测精度也更高。

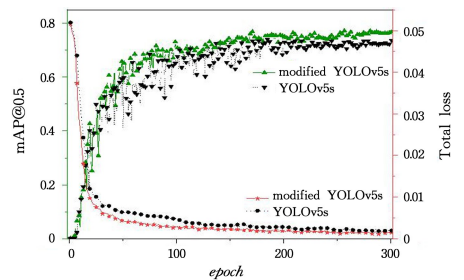


图 12 模型训练对比

Fig. 12 Comparison of model training

#### 4.5.3 模型轻量化对比

为了检测本文改进模型的可靠性,使用当前流行的轻量化网络替换 YOLOv5s 主干提取网络进行实验对比。其中包括 ShuffleNet<sup>[23]</sup> 和 GhostNet<sup>[24]</sup>。由表 5 可知,YOLOv5s-MobileNet 的计算量和参数量最少,检测速度最快。虽然其检测精度略低于 YOLOv5s-ShuffleNet,但在牺牲一定的检测精度下换取大幅降低模型参数数量的方案是可取的。因此,本文选用 MobileNetv3-Small 网络替换 YOLOv5s 主干提取网络。

表5 模型轻量化对比

Table 5 Comparison of model lightweight

模型	参数量	计算量/GFLOPs	FPS	mAP@0.5
YOLOv5s	7 035 811	16.0	78.6	0.733
YOLOv5s-MobileNetv3	<b>1494433</b>	<b>2.6</b>	<b>111.0</b>	0.693
YOLOv5s-ShuffleNet	3 802 347	7.9	88.2	<b>0.699</b>
YOLOv5s-GhostNet	2 377 908	5.2	101.0	0.684

## 4.5.4 不同模型对比

为了进一步验证本文模型的有效性,在相同数据集基准和实验环境下,将其与其他模型进行对比实验,实验结果如表6所列。首先对比了YOLOv3、YOLOv4、YOLOv5s、Mask R-CNN、文献[13]以及文献[14]的模型。其中改进的YOLOv5s模型的mAP@0.5分别提升了10.8%,6.8%,1.5%,3.9%,0.4%,2.4%;检测速度分别提升了45.6FPS,34.2FPS,56FPS,23.4FPS,58.7FPS,33.5FPS;参数量分别减少 $51.8 \times 10^6$ , $43.2 \times 10^6$ , $92.9 \times 10^6$ , $7.6 \times 10^6$ , $13.4 \times 10^6$ , $28.9 \times 10^6$ 。同时本文也对比了最新的YOLOv7, YOLOv7-tiny以及YOLOv8s模型。其中YOLOv7模型的检测精度最高,达到77.5%。相比YOLOv7模型,改进的YOLOv5s模型的检测精度略低,但检测速度快21.8FPS,体积减少 $23.2 \times 10^6$ 。相比YOLOv8s模型,改进的模型精度提升3.5%,速度提升8FPS,体积减少 $9.8 \times 10^6$ 。YOLOv7-tiny模型的体积最小,检测速度达到最高的124FPS,但检测精度较低。对于钢材表面缺陷检测问题,应在保证较高检测精度的前提下,尽可能减小模型体积,提高检测速度。对比结果说明,本文改进的YOLOv5s模型在检测速度和精度之间取得了平衡,是可取的有效方案。

表6 不同模型对比结果

Table 6 Comparison results of different models

模型	mAP@0.5	FPS	体积/MB
YOLOv3	0.664	56.4	62.90
YOLOv4	0.704	67.8	54.30
Mask R-CNN	0.757	46.0	104.00
YOLOv5s	0.733	78.6	22.40
文献[13]	0.768	43.3	28.20
文献[14]	0.748	68.5	43.70
YOLOv7	<b>0.775</b>	80.2	38.00
YOLOv7-tiny	0.738	<b>124.0</b>	<b>7.22</b>
YOLOv8s	0.746	94.0	24.60
改进的YOLOv5s	0.772	102.0	14.80

## 4.5.5 消融实验

为了验证不同改进方案对模型性能的影响,以YOLOv5s为基线进行消融实验。实验环境及配置参数均保持不变。结果如表7所列。①代表使用MobileNetv3-Small网络替换YOLOv5s主干提取网络,②代表采用BiFPN网络,③代表加入CBAM注意力机制,④代表使用K-means++算法替换K-means算法,“√”代表使用此方案。由表7可知,使用MobileNetv3-Small网络替换主干提取网络后,虽然平均检测精度有所下降,但网络的检测速度较改进前提升了41.2%,说明使用MobileNetv3-Small替换骨干网络可以大幅减少模型的参数,实现模型轻量化。使用BiFPN网络融合特征后,平均检测精度提升了2.1%。引入CBAM注意力机制后,检测速度有所下降,平均检测精度提升了3.8%,达到预期效果。

使用K-means++替换K-means算法后,平均检测精度提升了0.7%,检测速度提升了6.9%。使用4种方案共同改进YOLOv5s模型后,改进后模型的检测精度与检测速度达到平衡,mAP@0.5达到77.2%,检测速度达到102FPS。可见,实验所用改进方案是可取的。

表7 消融实验

Table 7 Ablation experiment

①	②	③	④	mAP@0.5	FPS
				0.733	78.6
√				0.693	<b>111.0</b>
	√			0.754	72.7
		√		0.771	40.5
			√	0.740	84.0
√	√	√	√	<b>0.772</b>	<b>102</b>

## 4.5.6 缺陷检测效果

为了更直观地比较模型改进前后的对比效果,本文利用改进前后的模型对测试集进行检测,部分检测结果如图13所示。由图13可得,原YOLOv5s模型对部分缺陷目标存在明显的漏检问题,并且检测精度不高,尤其对于小目标缺陷裂纹类(crazing)的检测精度较低。而改进后的YOLOv5s模型整体检测效果明显优于改进前,对裂纹类缺陷的检测也有了明显的提升。检测结果验证了本文改进的YOLOv5s模型的有效性,为钢材表面缺陷检测提供了实用的参考方案。

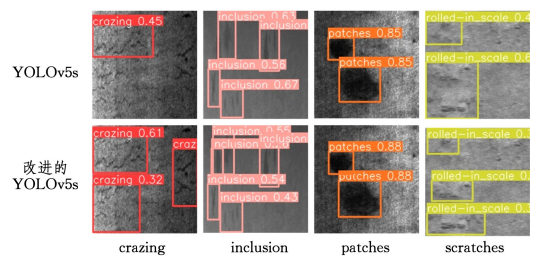


图13 部分检测结果

Fig. 13 Partial test results

**结束语** 为了解决现有钢材表面缺陷检测模型结构复杂、体积较大、检测精度较低、不易于部署于边缘端设备和嵌入式系统等问题,本文对YOLOv5s模型进行了改进。改进方法包括采用MobileNetv3-Small网络替换YOLOv5s主干提取网络、使用BiFPN网络进行特征融合、引入CBAM注意力机制以及使用K-means++算法聚类先验框。实验结果表明,改进后的模型在实现轻量化的同时提升了检测精度和速度,易于部署,满足带钢实际生产中的需求。下一步工作将研究如何进一步提高模型对裂纹类(crazing)缺陷的检测精度,同时研究模型在边缘端设备和嵌入式系统上的部署与应用。

## 参考文献

- [1] ZHANG R, WANG Y Y, DUAN Y Q, et al. Real time object detection and localization algorithm for flight robotic arms[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2022, 54 (1): 27-33.
- [2] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015:1440-1448.
- [3] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. Advances in Neural Information Processing Systems, 2016, 29: 379-387.

- [4] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017;2961-2969.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// Computer Vision — ECCV 2016: 14th European Conference, Amsterdam, The Netherlands (Part I 14). Springer International Publishing, 2016;21-37.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;779-788.
- [7] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv:1804.02767, 2018.
- [8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv:2004.10934, 2020.
- [9] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv:2107.08430, 2021.
- [10] 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]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023;7464-7475.
- [11] YANG J H, LI H, DU Y Y, et al. Lightweight object detection algorithm based on improved YOLOv5s[J]. Electro Optics and Control, 2023, 30 (2): 24-30.
- [12] 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.
- [13] LI X, WANG C, LI B, et al. Improvement of YOLOv5 steel surface defect detection algorithm [J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(2): 26-33.
- [14] WU D, LI M H, MA W K, et al. Steel surface defect detection based on improved YOLOv5[J]. Journal of Shanxi University of Science & Technology, 2023, 41 (2): 162-169.
- [15] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetv3[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019;1314-1324.
- [16] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861, 2017.
- [17] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018;4510-4520.
- [18] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020; 10781-10790.
- [19] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018;3-19.
- [20] MODHA D S, SPANGLER W S. Feature weighting in k-means clustering[J]. Machine learning, 2003, 52: 217-237.
- [21] ARTHUR D, VASSILVITSKII S. K-means++ the advantages of careful seeding[C]// Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. 2007; 1027-1035.
- [22] 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, 2019, 69(4): 1493-1504.
- [23] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018;6848-6856.
- [24] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020; 1580-1589.



**JIANG Bo**, born in 1998, postgraduate. His main research interests include machine learning and image processing.