

# 基于改进YOLOv5的超声钢轨缺陷检测方法

任恩璇<sup>1</sup>, 汪磊<sup>1</sup>, 鲍蟠虎<sup>2</sup>, 王振<sup>2</sup>

<sup>1</sup>上海工程技术大学, 城市轨道交通学院, 上海

<sup>2</sup>江苏三合声源超声波科技有限公司, 江苏 常州

收稿日期: 2023年3月22日; 录用日期: 2023年4月21日; 发布日期: 2023年4月28日

## 摘要

钢轨内部缺陷检测是保证列车运行安全非常重要的安全保证, 为了增加对钢轨内部缺陷的识别准确率, 减少人为不确定因素的干扰, 提出了一种用YOLOv5目标识别的方法来辅助识别, 并对YOLOv5算法进行改进, 改进了数据增强, 提升数据鲁棒性, 添加NAM注意力模块来捕获局部信息, 同时引入GhostNet, 对模型进行轻量化处理。改进的YOLOv5\_GNM算法的AP<sub>50</sub>值到达了87.7%, 参数量减少了37%, 帧数达到了66帧, 提高了缺陷识别的准确率。

## 关键词

钢轨缺陷检测, 注意力机制, YOLOv5, 超声相控阵图像

# Ultrasonic Rail Defect Target Detection Based on Improved YOLOv5

Enxuan Ren<sup>1</sup>, Lei Wang<sup>1</sup>, Panhu Bao<sup>2</sup>, Zhen Wang<sup>2</sup>

<sup>1</sup>School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai

<sup>2</sup>Jiangsu Sanheshengyuan Ultrasonic Technology Co., Ltd., Changzhou Jiangsu

Received: Mar. 22<sup>nd</sup>, 2023; accepted: Apr. 21<sup>st</sup>, 2023; published: Apr. 28<sup>th</sup>, 2023

## Abstract

Rail internal defect detection is a very important safety guarantee to ensure the safety of train operation. In order to increase the recognition accuracy of rail internal defects and reduce the interference of human uncertainties, a YOLOv5 target recognition method is proposed to assist the recognition, and YOLOv5 algorithm is improved to improve data enhancement, improve data robustness, add NAM attention module to capture local information, introduce GhostNet, and lightweight the

model. The  $AP_{50}$  value of the improved YOLOv5\_GNM algorithm reaches 87.7%, the number of parameters decreases by 37%, and the number of frames reaches 66, which improves the accuracy of defect recognition.

## Keywords

Rail Defect Detection, Attention Mechanism, YOLOv5, Ultrasonic Phased Array Image

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着国家的工业发展, 钢轨等材料产品的质量和安全也越发引起人们的重视。钢轨等零部件的内部缺陷检测识别, 已有许多方法, 但是考虑不损坏材料本身等方面, 无损检测技术表现出了独特的优势。在无损检测领域, 超声检测操作便捷、成本低, 同时超声相控阵以其高精度、无损、高安全性等优点, 被广泛应用于交通运输等领域[1]。超声相控阵技术是超声检测非常重要的一部分, 在朝着高实时性, 高智能化方向发展[2], 然而以往的超声相控阵图像多为由技术人员进行判断, 需要一定的技术门槛。基于深度学习的机器视觉技术因高效性和准确性, 用于缺陷检测中可以在一定程度上提升检测的效率[3]。

深度学习的目标检测框架被广泛划分为两种类型: 两阶段检测框架和一阶段检测框架。两阶段检测算法是一种基于检测框和分类器的二阶段算法, 最早的二阶段网络是 2014 年 Girshick 等人[4]提出的 R-CNN, 在 VOC 系列数据集上达到了 53.7% 的准确率。随后, Fast R-CNN [5]在前者基础上增设了 ROI pooling, 在主干网络上提取特征图进行分类回归, 减少了计算量。2015 年提出的 Faster R-CNN [6]是最典型的二阶段网络, 它能够自动提取目标区域, 计算性能大幅度提升。一阶段算法则包括 SDD [7]和 YOLO [8] [9] [10] [11]系列算法等, 这些算法的速度很快, 可以实现实时检测。当然, 学者们也对 YOLOv5 进行了很多改进方案。仇娇慧等设计了一种基于通道和空间的注意力机制的齿轮表面缺陷检测方法, 在 YOLOv5s 网络模型的基础上引入卷积注意力模块, 增强缺陷区域特征并抑制背景区域特征, 提高小目标的检测精度[12]。王淑青等对 YOLOv5 模型输入端改进了一种动态反馈多尺度训练的数据增强方法以提高特征训练精度, 然后用 ELU 函数替换主干网络中的激活函数, 加快训练速度[13]。将超声相控阵图像缺陷检测技术与深度学习紧密联系起来, 可以准确地检测各种缺陷的大小和位置, 减少一些不稳定因素的干扰, 以提高超声相控阵图像缺陷识别的准确率, 使设备的可靠性得到进一步加强[14]。吉春生提出了基于深度学习的方法来识别相控阵超声检测图谱中的缺陷, 基于 Faster R-CNN 网络框架, 根据焊缝缺陷的特点, 使用带有特征金字塔网络的 ResNet50 作为特征提取的主干, 以此来更好地检测缺陷图谱中的小缺陷[15]。综上所述, 本文提出一种改进的 YOLOv5 网络框架, 对钢轨缺陷的超声相控阵图像进行精准识别。

## 2. YOLOv5 算法概述

YOLOv5 算法是基于 YOLOv4 的改进版本, 经过多次改进后, YOLOv5 的检测准确度、实时性上提升很大, 这对于钢轨的实时成像检测具有非常重要的意义。YOLOv5 网络模型分为 4 个部分, 其网络结构如图 1 所示。主要结构从 Input 输入到 Detect 输出, 下方是对各个结构的细分。

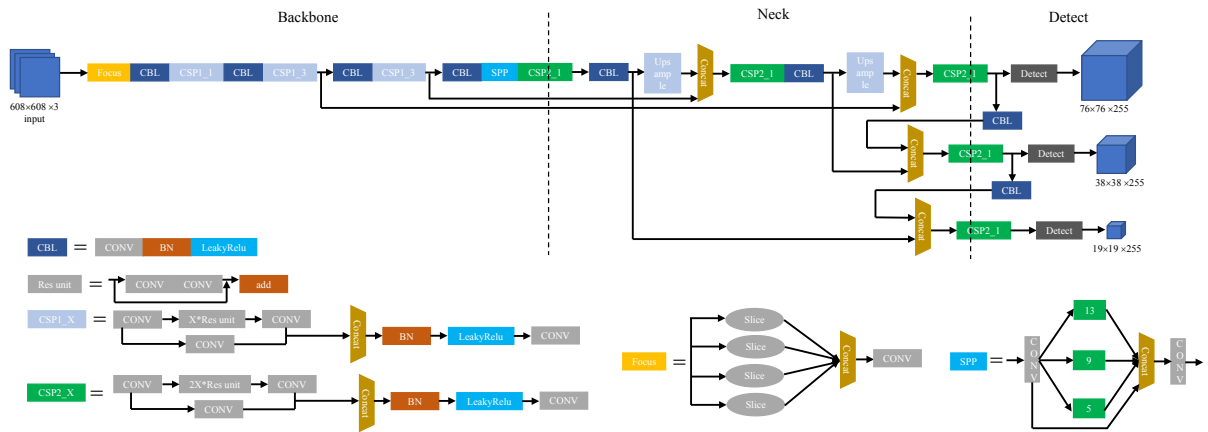


Figure 1. YOLOv5s structure  
 图 1. YOLOv5s 结构图

### 2.1. Input 端

Input 端包含 Mosaic 数据增强、自适应锚框计算和自适应图片缩放三个部分。Mosaic 数据增强是使用裁剪缩放等方式将 4 张图片拼接；自适应锚框计算则是针对不同的数据集设定初始的锚框，其初始框基础上输出预测框，计算差距后再反向更新进行迭代；自适应图片缩放则是将图像都缩放成统一的尺寸。

### 2.2. Backbone

Backbone 包括 Focus 结构和 CSPNET 结构。Focus 将输入的  $608 \times 608 \times 3$  的图像进行切片操作，然后再经过 32 个卷积核的卷积后得到  $304 \times 304 \times 32$  的特征图。

### 2.3. Neck

Yolov5 采用 FPN (Feature Pyramid Networks, 特征金字塔网络) + PAN (Pyramid Attention Network, 金字塔注意力网络) 的结构，其结构如图 2 所示。

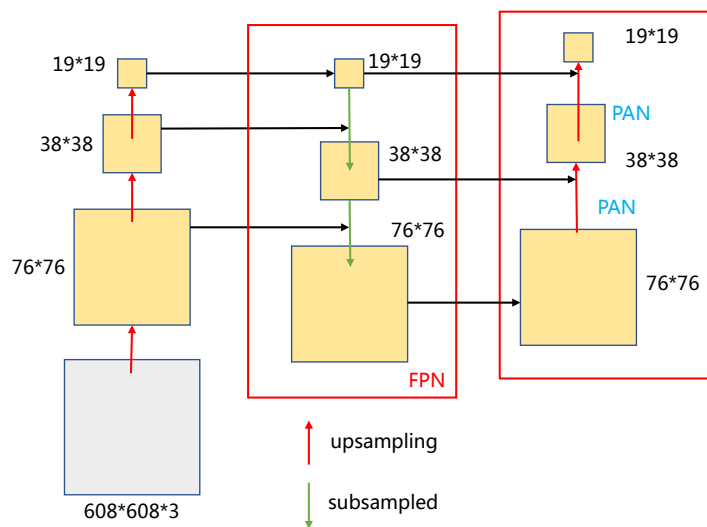


Figure 2. Neck structure  
 图 2. Neck 结构图

## 2.4. 输出端

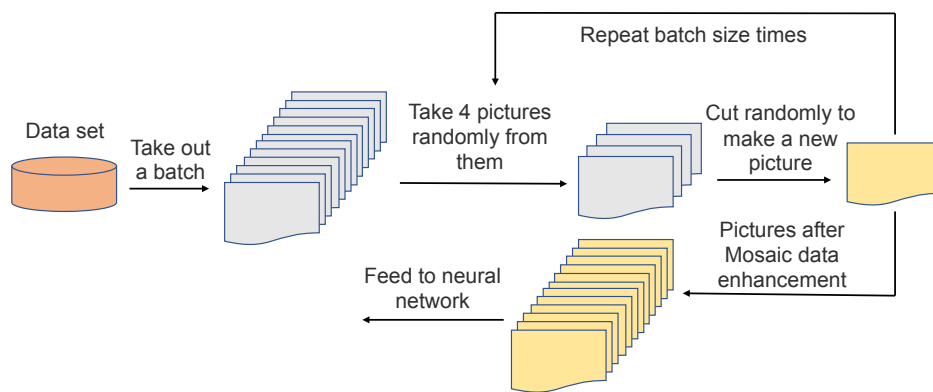
Prediction 包括 Bounding box 损失函数和 NMS 非极大值抑制。YOLOv5 使用 CIOU\_Loss 作为 Bounding box 的损失函数，提高了预测框回归的速度和精度。

## 3. YOLOv5 算法改进

### 3.1. 数据增强改进

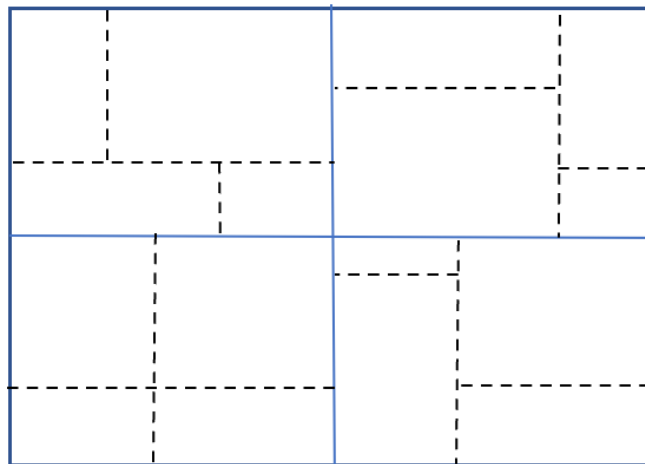
许多的神经网络的参数都是非常多而复杂的，要使得这些参数可以正确工作运行起来则需要大量的数据进行模型训练。获取数据这项工作通常是要花费很多成本，但是通过使用数据增强，对数据训练样本进行缩放、平移、翻转、色彩变换等可以降低模型对其中一些元素的依赖，从而提高模型的泛化能力。

Mosaic 数据增强的方法，丰富数据集的同时，也在一定程度上提升了网络的训练速度，如图 3 所示。



**Figure 3.** Data enhancement flowchart  
**图 3.** 数据增强流程图

由于钢轨实验样本数量少的原因，本文改进 Mosaic 数据增强方式，将原来的 4 张图片进行修改，从而对 16 张图片进行裁剪，并随机缩放，随机排列形成一张新图片。进一步增加了鲁棒性，效果如图 4 所示。



**Figure 4.** Mosaic-16 effect picture  
**图 4.** Mosaic-16 效果图

### 3.2. 引入注意力机制

在钢轨超声检测中，由于需要提高对重要的缺陷信息的关注度，降低对一些杂波的等其他信息注意力，所以引入注意力机制。

之前的很多的研究都在考虑如何通过注意力算子来获取显著性的特征，但是很少有研究去考虑权值的贡献因子的影响，其中这个因子可以进一步抑制某些不被考虑的特征。本文通过在网络结构中使用一种基于归一化的注意力模块——NAM 模块，用权值的贡献因子来提升注意力机制的效果。NAM 是 Normalization-Based Attention Module 的缩写，是 CBAM 模块的变体[16]，它使用了 Batch Normalization (BN)比例因子来表示权值的重要性，但是并不会像 SE, BAM 和 CBAM 这些结构增加网络的全连接层和卷积层，这样可以尽量少增加模型的计算量。

如图 5 所示，它包括通道注意力模块和空间注意力模块，可以通过调整通道和空间维度中训练权重的方差测量来重新加权注意力。

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_b}{\sqrt{\sigma_b^2 + \epsilon}} + \beta \tag{1}$$

其中  $\gamma$  和  $\beta$  分别表示可训练的变换参数(尺度和位移); 而  $\mu_b$  和  $\sigma_b$  分别是每个 batch  $b$  中的平均值和标准差; 其中  $\mu_b$  是均值,  $\sigma_b^2$  是 BN 中的方差, 方差越大, 意味着变化越大, 信息越丰富。对于通道注意力模块, 基于归一化相关权重  $W_\gamma$  可以将更多的注意力集中在重要通道, 同时抑制信息量较小的权重。假设  $F_1 \in R^{H \times W \times C}$  是输入特征图, 其中  $H$ 、 $W$  和  $C$  表示高度, 宽度和通道数。  $\gamma$  是每个通道的比例因子, 权重为  $W_\gamma$ , 则通道注意力的输出  $M_c$  可以表示为:

$$M_c = \text{sigmoid}(W_\gamma(BN(F_1))) \tag{2}$$

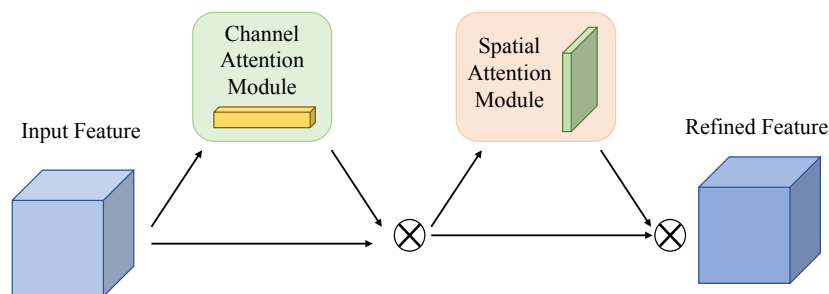


Figure 5. Normalization-based attention module  
图 5. NAM 模块

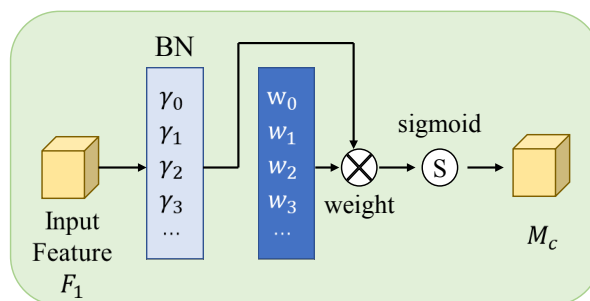


Figure 6. Channel attention module  
图 6. 通道注意力模块

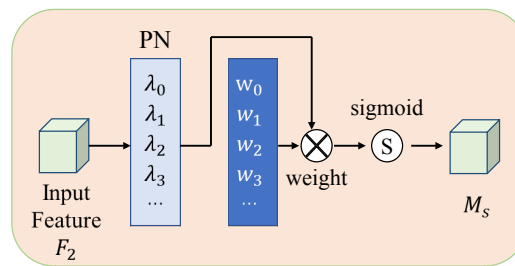


Figure 7. Spatial attention module  
图 7. 空间注意力模块

通道注意力模块如图 6 所示，空间注意力模块如图 7 所示。以通道注意模块的方式，其空间注意模块将 BN 应用于空间维度中的像素，进行像素归一化(PN)。它根据比例因子  $\lambda$  将更多的注意力集中在在信息量更大的像素上，并调整相关的权重  $W$ 。同样， $F_2 \in R^{H \times W \times C}$  是输入特征图，输出空间注意力模块的输出  $M_s$  是：

$$M_s = \text{sigmoid}(W_\lambda(PN(F_2))) \quad (3)$$

同时为了抑制不太重要的权重，向损失函数中增加正则化项。

$$\text{Loss} = \sum_{(x,y)} l(f(x,W), y) + p \sum g(\gamma) + p \sum g(\lambda) \quad (4)$$

其中， $l(\cdot)$  和  $g(\cdot)$  分别表示损失函数和范数惩罚函数； $x$  和  $y$  分别为输入和输出； $W$  是网络权重； $p$  是平衡惩罚因子。

在每个网络 block 的末端嵌入一个 NAM 模块，它重新处理并合理化从主干中提取的重要特征，以方便对象在下一步中预测。这样可以使模型更加关注图中的关键信息，提升计算效率，但由于 NAM 的轻量化，增加的计算成本并不多。

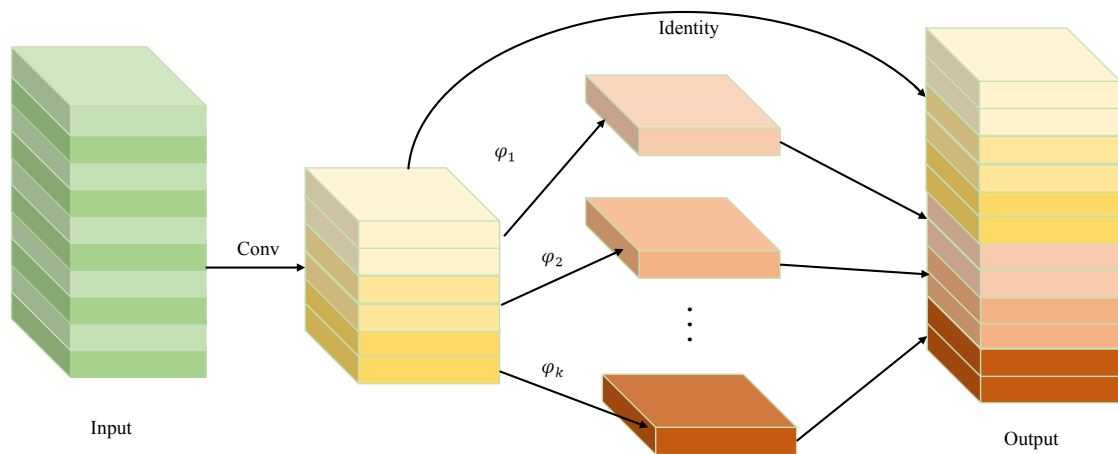


Figure 8. Ghost module  
图 8. Ghost 模块

### 3.3. 模型轻量化

在实际的超声检测设备中，为了满足三防等要求，硬件的配置不是很高，这就要求后续附加的目标检测模型的计算量有一定的要求。因此对模型进行轻量化处理，引入 GhostBottleneck 模块，它来源于

GhostNet [17], 如图 8 所示, 能够在最大限度减少网络计算损耗的同时提高网络的检测准确率。假设正常卷积是为 32 个卷积核的卷积计算, 在 Ghost 模块的第一阶段会采用其原有卷积的一半进行计算, 在第二阶段中  $\varphi$  是 3\*3 的小卷积, 通过 3\*3 的这些卷积对第一阶段所得的特征图进行逐层卷积, 来进一步优化之前卷积的计算损耗。

GhostBottleneck 模块作为一个即插即用模块, 可以保证充分的增值输入数据。此模块包含两种结构, 一种是 Stride 为 1 的 GhostBottleneck 模块, 由两个堆叠的 Ghost 模块组成, 另一种是 Stride 为 2 的 GhostBottleneck 模块, 中间连接了一个 DW 卷积层, 如图 9 所示。其中第一个 Ghost 模块的作用是增加输入特征图的通道数, 为后面的操作来做扩展处理, 而第二个 Ghost 模块的作用是减少输出特征图的通道数使其与网络中的直径结构相匹配。本文使用 GhostBottleneck 来替换原本 YOLO 网络主干部分中的 Bottleneck CSP 结构来优化网络的参数量和计算规模, 减少模型的计算量。

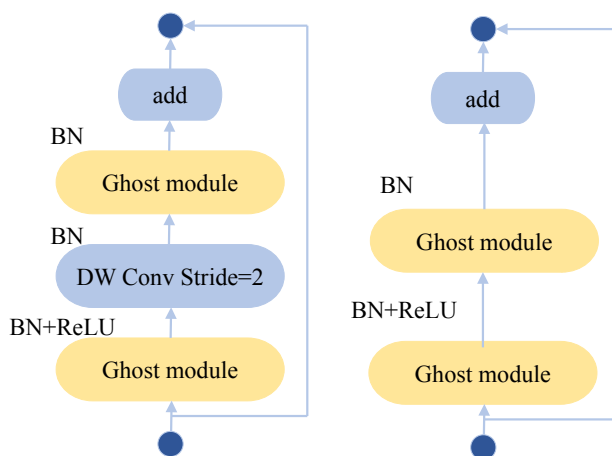


Figure 9. Two GhostBottleneck module  
图 9. 两种 GhostBottleneck 模块

## 4. 实验

### 4.1. 实验环境

本文采用自制钢轨缺陷超声相控阵数据集进行训练, 包含 1500 张图片, 搭建 PyTorch 深度学习平台, 具体测试环境: 显卡为 NVIDIA RTX 2070SUPER, 显存为 8 G, CUDA 版本 10.2, 编译语言为 Python3.7, 批处理量 batch size 设置为 16, 初始学习率设置为 0.01, 动量系数设置为 0.937, 权值衰减系数设置为 0.0005。一共训练 300 个 epochs。训练集 80%, 验证集 10%, 数据集 10%。

### 4.2. 评价标准

对于该算法的评价标准, 本文使用了结合召回率和精准率的  $mAP$  (mean Average Precision)对改进前和改进后的算法性能进行评价。其中召回率表示为在所有正类样本中的检出率, 其计算公式如下

$$R = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

其中的  $R$  代表召回率,  $TP$  表示算法将样本正类预测为正类的个数;  $FN$  则是表示了将样本正类预测为负类的个数。

精准率  $P$  可以表示未预测的正样本占有所有正样本的比例, 计算公式如下:



$$P = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

PR 曲线是以  $P$  为纵轴，每种类别通过 PR 曲线积分

$$AP = \int_0^1 p(r) dr \quad (7)$$

得到  $AP$ ，对所有类别的  $AP$  值求平均值即为  $mAP$

$$mAP = \frac{\sum_{i=1}^N AP}{N} \quad (8)$$

### 4.3. 对比实验

对原 YOLOv5s 和改进后的 YOLOv5s 进行实验对比，做 4 次实验，如下所示。

- 1) 原 YOLOv5s
- 2) YOLOv5s\_M: 改进数据增强。
- 3) YOLOv5s\_G: 引入 GhostBottleneck 模块。
- 4) YOLOv5s\_N: 引入 NAM 模块。
- 5) YOLOv5s\_GNM: 引入 NAM 模块、GhostBottleneck 模块和改进数据增强。

实验结果如表 1 和表 2 所示：

**Table 1.** Comparison table of model complexity

**表 1.** 模型复杂度对比表

Model	Parameters (M)	GFLOPs	Weight (MB)
YOLOv5s	7.07	16.4	14.4
YOLOv5s_M	7.11	16.5	14.5
YOLOv5s_G	4.17	9.7	8.7
YOLOv5s_N	7.17	16.6	14.7
YOLOv5s_GNM	4.45	10.4	9.3

**Table 2.** Model performance comparison table

**表 2.** 模型性能对比表

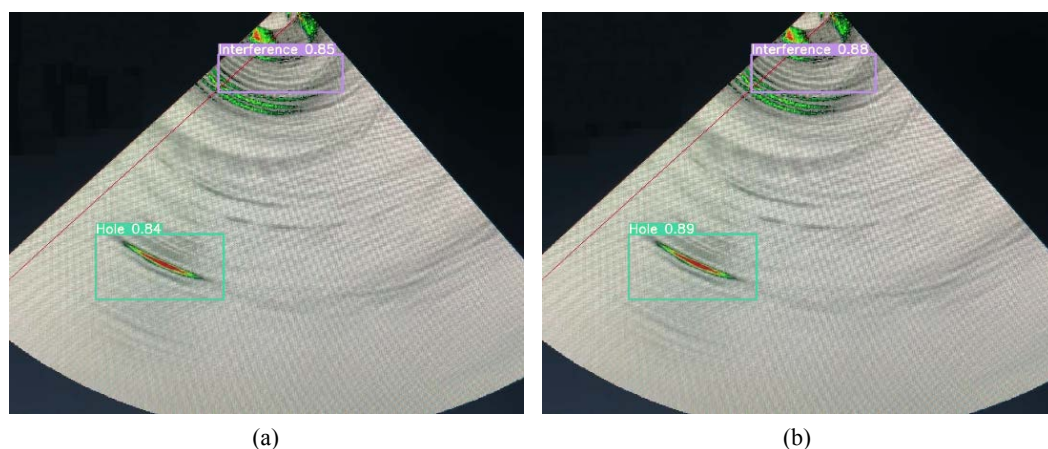
Model	mAP@0.5/%	检测速度(s/pic)
YOLOv5s	86.6	0.019
YOLOv5s_M	87.4	0.019
YOLOv5s_G	85.4	0.014
YOLOv5s_N	88.3	0.020
YOLOv5s_GNM	87.7	0.015

由表可知，YOLOv5s\_M 算法对 mosaic 数据增强进行改进后，虽然模型大小有了略微的增加，但 mAP 增长了 0.8%，证明了改进的可行性。YOLOv5s\_G 算法进行 GhostBottleneck 结构改进后，模型的参数数量减少 41%，平均计算时间少了 5 ms，帧率从 52.6 帧提升到 71.4 帧，提升了 35.8%。但 mAP 只减少了 1.2%，用少量的性能损失换来大量的参数数量的降低，实现了轻量化，这是由于 GhostNet 使用了线性运算替代部分卷积过程，因此在其得到近似于原始特征图的同时减少了大量的计算量。YOLOv5s\_N 算法增加



了 NAM 模块, 增加了注意力机制, 检测速度虽然增加了 1 ms, 但是 mAP 提升了 1.7%, 表现了注意力机制对于 YOLOv5 目标检测精确度有加强作用。YOLOv5s\_GNM 算法则是在检测精度上提升 1.1% 的同时, 模型参数减少了 37%, 帧率也从 52.6 帧提升至 66.7 帧, 实现了模型的轻量化。

改进前的 YOLOv5s 网络模型和改进后的 YOLOv5s\_GNM 网络模型的检测结果如图 10 所示, Interference 的置信度从 0.85 提升到 0.88, 缺陷 Hole 的置信度从 0.84 提升到 0.89, 改进后的检测精度有明显提升。



**Figure 10.** Comparison of test results. (a) Before improvement; (b) After improvement  
**图 10.** 检测结果对比。(a) 改进前; (b) 改进后

## 5. 结论

本文提出了一种改进 YOLOv5 算法, 并将其融合到超声相控阵钢轨缺陷检测中, 可以辅助专业人员通过超声相控阵仪器检测钢轨缺陷。结果表明:

- 1) 通过改进 Mosaic 数据增强丰富了数据集, 增加小样本目标, 实验发现, 模型的检测精度有了小幅度的上升。
- 2) 在 YOLOv5s 的网络结构中引入注意力机制, 融入 NAM 模块, 以此来强化重要的缺陷特征, 在少量增加模型计算量的基础上增加了模型检测的准确度。
- 3) 在 YOLOv5s 网络结构中引入 GhostBottleneck 结构, 模型的参数量和计算量大幅度减少, 实现了模型的轻量化, 方便后续模型部署至工控机或嵌入式设备。

## 基金项目

QNTD202108 轨道交通智能运维青年创新团队; 12172211 非饱和复合地基固结理论构建与解析模型研究。

## 参考文献

- [1] 张学强. 超声相控阵探伤中缺陷图谱处理与智能分类技术的研究[D]: [硕士学位论文]. 包头: 内蒙古科技大学, 2021.
- [2] 薛文杰. 超声相控阵全聚焦成像实现与降噪方法研究[D]: [硕士学位论文]. 南京: 东南大学, 2021.
- [3] Kong, X.X. and Li, J. (2018) Image Registration-Based Bolt Loosening Detection of Steel Joints. *Sensors (Basel, Switzerland)*, **18**, 1000. <https://doi.org/10.3390/s18041000>
- [4] Girshick, R., Donahue, J., Darrell, T., et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Seman-

- tic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [5] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [6] Ren, S., He, K., Girshick, R., et al. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [7] Liu, W., Anguelov, D., Erhan, D., et al. (2016) Ssd: Singleshot Multibox Detector. In: *European Conference on Computer Vision*, Springer, Cham, 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [8] Redmon, J., Divvala, S., Girshick, R., et al. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [9] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [10] Joseph, R. and Ali, F. (2018) YOLOv3: An Incremental Improvement.
- [11] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection.
- [12] 仇娇慧, 贝绍轶, 尹明锋, 卿宏军. 基于改进 YOLOv5s 的齿轮表面缺陷检测[J]. 现代制造工程, 2022(3): 104-113. <https://doi.org/10.16731/j.cnki.1671-3133.2022.03.016>
- [13] 王淑青, 张鹏飞, 要若天, 鲁东林, 顿伟超. 基于改进 YOLOv5 的太阳能电池片表面缺陷检测[J]. 仪表技术与传感器, 2022(5): 111-116.
- [14] 费学智, 吴军芳, 柳铭哲, 等. 超声相控阵图像中的缺陷识别算法[J]. 无损检测, 2018, 40(8): 61-66.
- [15] 吉春生. 基于改进 Faster R-CNN 的焊缝缺陷图谱识别[J]. 中国化工装备, 2022, 24(2): 26-32+36.
- [16] Liu, Y., Shao, Z., Teng, Y. and Hoffmann, N. (2021) NAM: Normalization-Based Attention Module.
- [17] Han, K., Wang, Y.H., Tian, Q., et al. (2020) GhostNet: More Features from Cheap Operations. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 1577-1586. <https://doi.org/10.1109/CVPR42600.2020.00165>