文章编号:1671-4598(2025)07-0046-08 DOI:10.16526/j.cnki.11-4762/tp.2025.07.006 中图分类号:TP391.41 文献标识码:A

基于改进 YOLOv5s 的金属表面缺陷检测

顾嘉炜[1],焦良葆[1,2],焦 波[1],孟 琳[1],李笑笑[1]
(1. 南京工程学院 人工智能产业技术研究院,南京 211167;
2. 江苏省智能感知技术与装备工程研究中心,南京 211167)

摘要:金属表面缺陷在飞机和航天器制造等领域易致结构失效,带来巨大的事故风险;针对传统 YOLOv5s 算法检测金属表面冲孔、丝状斑点、月牙形缺口等性能差的问题,现提出一种改进的 YOLOv5s 算法;该算法将 Ghost 网络加入到 Backbone 部分;在 Neck 网络中增加了 MHSA 注意力机制;将原损失函数替换为 DIOU;经实验验证,改进后的网络与原网络相比,*mAP*@0.5 提升了 3.2%;提高了模型的检测精度并且误检、漏检率较低,证实了该方法在金属表面缺陷检测上的有效性。

关键词:目标检测;金属表面缺陷;深度学习;损失函数;注意力机制

Metal Surface Defect Detection Based on Improved YOLOv5s

GU Jiawei[1], JIAO Liangbao[1,2], JIAO Bo[1], MENG Lin[1], LI Xiaoxiao[1]
(1. AI Industrial Technology Research Institute, Nanjing Institute of Technology, Nanjing 211167, China;
2. Jiangsu Intelligent Perception Technology and Equipment Engineering Research Center, Nanjing 211167, China)

Abstract: Metal surface defects can easily lead to structural failure in the field of aircraft and spacecraft manufacturing, which brings great risk of accidents. Traditional YOLOv5s algorithms have the the disadvantages of detecting metal surface punching, filamentous spots, and crescent notches, to solve these issues, an improved YOLOv5s algorithm is proposed. This algorithm adds the Ghost network to the Backbone. The multi-head self-attention (MHSA) attention mechanism is added in the Neck network. The original loss function is replaced with the distance intersection over union (DIOU). Experimental results show that compared with the original network, the improved network increases the *mAP*@0.5 by 3.2%, improving the detection accuracy of the model, with low false detection and missed detection rates, which proves the effectiveness of the proposed method in metal surface defect detection.

Keywords: object detection; metal surface defects; deep learning; loss function; attention mechanism

0 引言

金属材料作为重要的工业材料,在航空航天领域有着不可替代的地位[1]。然而在材料加工、生产过程中,由于各种因素的影响,工件会产生表面冲孔、丝状斑点、月牙形缺口等缺陷,这不仅会成为安全隐患,导致结构失效,严重可能造成人员伤亡[2]。因此,航空航天工业采用先进、准确、快速的金属表面缺陷检测技术是十分必要的。然而依靠人工进行的传统检测方式限制重重,传统方法检测速度缓慢、效率低下,无法满足流水线上对高效、快速检测的需求。因此,在工业生产中提升缺陷检测的效率和准确性是一个亟待解决的问题[3]。

凭借深度学习强大的特征表示和学习能力,其已经成为了诸多人工智能任务中的重要“推动力”[4]。目前主流的检测方法大致分为两大类:一阶段(One-Stage)检测方法和二阶段(Two-Stage)检测方法。一阶段检测方法直接从输入图像预测目标的类别和位置,不需要单独的区域提议步骤。这使得一阶段方法通常比二阶段方法运行速度更快,更适合实时应用场景。典型的一阶段检测方法包括:YOLO(You Only Look Once)、SSD(Single Shot MultiBox Detector)、RetinaNet。二阶段检测方法首先创建目标的候选区域,随后对这些区域执行

收稿日期:2024-05-14; **修回日期**:2024-06-07。
基金项目:江苏省产学研合作项目(BY20230656)。
作者简介:顾嘉炜(2000-),男,硕士研究生。
焦良葆(1972-),男,教授,硕士研究生导师。
孟 琳(1989-),女,副教授,硕士研究生导师。
引用格式:顾嘉炜,焦良葆,焦 波,等.基于改进 YOLOv5s 的金属表面缺陷检测[J].计算机测量与控制,2025,33(7):46-53,63.

分类和边框回归以得到最终的检测结果。这种方法通常能够达到更高的准确率，但速度较慢。典型的二阶段检测方法包括：R-CNN（Regions with Convolutional Neural Network）、FastR-CNN（Fast Region-Based Convolutional Neural Network）。

在单阶段目标检测中，检测模型存在因没有像二阶段方法那样先通过区域提议步骤精细化地选择感兴趣的区域而导致的准确度问题、缺乏专门的细粒度的区域提议阶段导致模型难以精确控制边界框的大小和位置而导致的定位精度不足的问题等。为解决上述问题，文献[5]通过引入 DRA（Dimension Related Attention）模块来解决 C3 模块信息丢失问题，增强主干网络提取图像特征的能力；针对感受野大而导致的定位困难问题，引入新的 SIOU（Scale-Invariant Overlap Union）损失函数，在提高边界框的定位精确度的同时优化模型的推理速度，间接提升模型的性能。文献[6]提出添加 DyHead（Dynamic Head）检测头、更换 aLRPLoss（active Learning by Region Proposal Loss）损失函数、C3-Faster 代替网络中的 C3 模块，提升了对于大小缺陷目标的检测效果。

为了有效提高金属表面缺陷的检测准确度与精度，本文将对 YOLOv5 的第 6.0 版本进行改进。将 Ghost Net 架构应用到 Backbone 网络，通过使用少量的卷积操作生成“幽灵”特征图，大大减少计算成本；将 MHSA（Multi-HeadSelf-Attention）注意力机制加入到 Neck 网络中，模型能够捕捉到序列内各元素之间的复杂关系，提高模型的表征能力；通过将损失函数修改为 DIoU _ Loss（Distance Intersection Over Union Loss），考虑了重叠面积和中心点距离，当目标框包裹预测框的时候，直接度量 2 个框的距离，大大提高了收敛速度。

1　YOLOv5 网络结构

YOLOv5 模型在 YOLO 系列中表现出较高的检测速度和精确度[7]。该模型根据网络的宽度和深度不同，分为五个版本：YOLOv5l、YOLOv5m、YOLOv5n、YOLOv5s 和 YOLOv5x。其中，YOLOv5s 版本具有最小的模型尺寸，因此提供了最快的检测速度[8]。

YOLOv5 网络结构主要包括输入端、Backbone 网络、Neck 网络、输出端[9]。YOLOv5 的输入端采用了与 YOLOv4 输入端一样的 Mosaic[10] 数据增强方式，而 Mosaic 数据增强技术是基于 2019 年末提出的 CutMix 方法而来。不同于 CutMix 的两张图片拼接，Mosaic 技术通过将 4 张图片以随机缩放、裁剪、排列的方式组合，旨在丰富数据集并降低 GPU 负载。YOLOv4 在 YOLOv3 的基础上进行了锚框的自适应计算方法的改进。YOLOv4 能够自动根据不同的数据集计算初始锚框的值，无需依赖于外部程序，这一优化极大地简化了模型的预处理步骤。而在 YOLOv5 中，这一技术得到了进一步的发展，通过每次训练基于当前数据集动态优化锚框尺寸，以更好地适配目标的具体形状和大小。这种方法显著提升了模型在特定数据集上的目标检测精确度。此外，YOLOv5 还引入了一种先进的自适应图片缩放技术。这项技术根据模型的输入需求，动态地调整图像到统一的标准尺寸，以优化检测网络的性能。考虑到原始图片的长宽比多样性，缩放过程中常常涉及在图像边缘添加填充，以维持比例一致性。然而，不均匀的填充可能在图像的两侧形成黑边，这不仅增加了数据的冗余，也可能因过多的无效像素而对模型的推理速度产生负面影响。

6.0 版本的 Backbone 网络与 5.0 版本有些许不同，如图 1 为 YOLOv5 6.0 版本网络结构图。5.0 版本在 Backbone 开头部分使用 Focus 层，其关键操作是将图片切片[11]。而在 6.0 版本中则使用 Conv（Convolution）代替，目的是便于导出其他框架。Backbone 的核心组成包括 CBS 模块、C3 模块及 SPPF 模块。其中，CBS（Conv-BN-SiLU）模块由卷积层、批量归一化层和 Sigmoid 线性单元组成[12]。C3 模块，也称为 CSP 模块，设计上分为两种。在主干网络中的 CSP1 _ X 模块，特征图输入后分为两个路径：一个路径通过 CBS 模块后，

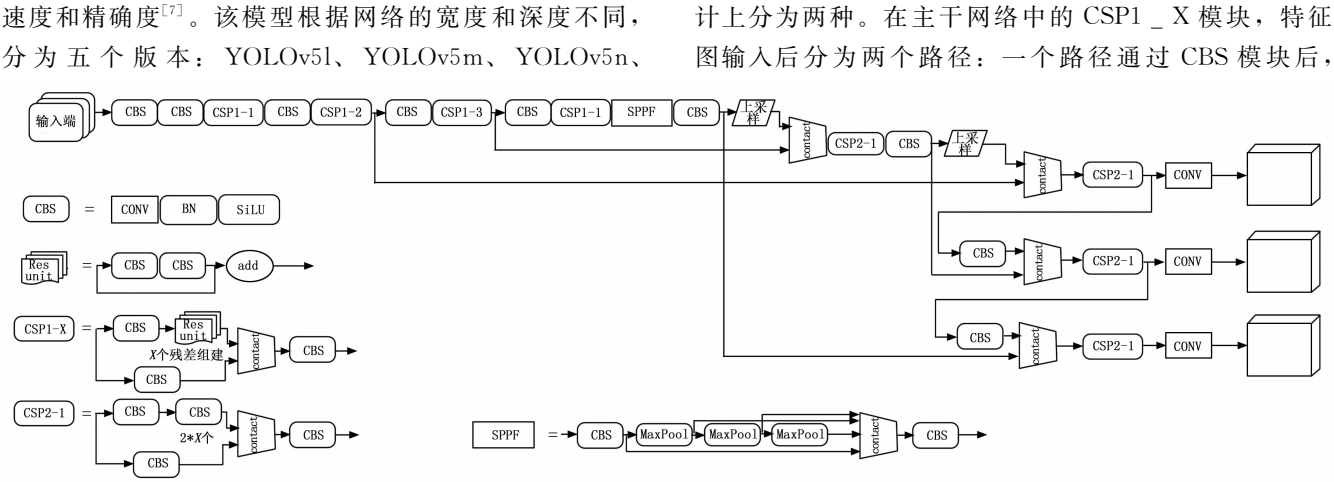

图 1　YOLOv5 6.0 版本网络结构图

继续通过 X 个残差结构的卷积；另一个路径仅通过 CBS 模块。最终，这两个路径在连接层汇合，并再次经过 CBS 模块处理。另一种是 CPS2 _ X 存在于 Neck 网络中，它与 CSP1 _ X 的唯一不同之处在于将 CSP1 _ X 的 X 个残差结构替换成了 CBS，强化了特征提取能力[13]。在 6.0 版本中，引入了 SPPF（Spatial Pyramid Pooling Fast）结构，全称空间金字塔池化，旨在提升卷积神经网络在提取重复图像特征时的效率。该结构通过生成快速的候选框来加速处理过程，同时显著减少计算资源的消耗。此外，SPP 能有效避免由于图像区域的裁剪和缩放所可能引起的图像失真。SPPF 在 SPP 的基础上进行了进一步的优化，显著提升了处理速度。在 SPPF 中，特征图首先通过 CBS（Convolution-BatchNorm-Swish）结构进行处理，以增强特征激活和正规化效果。之后，特征图被送入 3 个并行的最大池化层，每个层针对不同的尺度捕获空间特征，从而进一步提高了模型在处理不同尺度输入时的适应性和运行速度。这种结构的引入，使得 6.0 版本在保持图像特征精度的同时，能够以更快的速度进行图像分析处理。

Neck 部分的设计采用结合 FPN（Feature Pyramid Network）和 PAN（Path Aggregation Network）的设计，这一高级结构旨在优化和增强特征处理。FPN 作为一种经典的特征融合架构，其设计核心在于利用多层次的特征金字塔来优化网络对不同尺寸对象的处理能力。FPN 将顶层的高语义特征通过一系列上采样和横向连接操作，有效地传递到低层特征图中，这些低层特征图在经历上采样后，虽然保持了高分辨率，但通常语义较弱。FPN 的这种结构，通过融合不同层次的信息，弥补了单一尺度特征在表达能力上的不足。进一步地，PAN 作为对 FPN 的优化和扩展，引入了自底向上的信息反馈路径，这种创新设计使得低层的细节信息也能够反馈到高层特征中，从而进一步丰富高层特征的细节表达。PAN 通过增设的下采样路径，使得低层的高分辨率特征能够通过逐级的集成，被有效地传递到顶层。这样的信息流不仅是自顶向下的，也包括了自底向上的反馈机制，为网络提供了一个全方位的特征信息流动结构。此外，PAN 通过这种双向信息流的设置，有效地提升了特征图的语义一致性和空间精确度，增强了模型对于复杂场景下各种尺度物体的检测能力。这两种网络结构的融合不仅优化了特征的层次性分布，还增强了特征的表达能力，这在处理多尺度、高复杂度的视觉任务时显得尤为重要。FPN 和 PAN 的结合有效解决了传统深度学习模型在处理复杂视觉任务时遇到的尺度变化和信息丢失问题。这种先进的网络结构设计，为深度学习技术在实际应用中的广泛部署开辟了新的可能，特别是在那些对实时性和精确度要求极高的场合，展示了卓越的性能。

在 YOLOv5 模型的架构中，输出层首先初步定位目标，计算并输出边界框的具体坐标。接着，该层对检测到的对象进行类别分类，并计算每个对象的置信度，即预测的边界框确实包含某类对象的概率。置信度打分机制考虑了类别的概率和边界框的精确度。最后，为了优化输出结果，通过非极大值抑制[14] 算法处理，该算法筛选出一组最终的边界框，减少了检测过程中边界框的重复数量。NMS（Non-Maximum Suppression）通过比较不同边界框的重叠程度和置信度，保留最有可能代表独立实体的边界框，从而有效地降低了误检率和模型的计算负担。

2 改进策略

2.1 更换 Ghost 网络

在主流神经网络中，特征图的丰富冗余不仅有助于提升 CNN 的性能，更在增强模型对复杂环境的适应能力、提高其鲁棒性以及加深对上下文信息的理解方面发挥着关键作用。针对这一现象，本文提出了一种创新的 Ghost 模块，该模块不仅不避免冗余的特征映射，反而采纳并扩展这些映射，通过成本效益高的策略增加特征映射的数量。该模块利用一组基础特征映射，并应用一系列低成本的线性变换，生成大量的 Ghost 特征映射，以此来更全面地揭示和利用潜在的特征信息，从而在保持计算效率的同时，极大地丰富了模型的表达能力[15]。

深度卷积神经网络通常包含大量的卷积层，这会导致计算量大增。例如，MobileNetV3[16] 采用较小的卷积核以构建高效的 CNN，尽管如此，网络中剩余的 1×1 卷积层仍占用大量内存和计算资源。此外，网络中常用的激活函数如 sigmoid[17]、ReLU（Rectified Linear Unit）和 Leaky ReLU（Leaky Rectified Linear Unit）也在各层中起到关键作用，通过引入非线性来帮助网络捕捉复杂的输入模式。如图 2 所示为常规卷积，运算公式如下（1），它将一个卷积核滤波应用于输入特征图以生成输出特征图。其中，X 为特征图，它的形状为 $h*w*c$，h 和 w 为输入特征图的高度和宽度，c 为输入特征图的通道数，n 为输出特征图的通道数，f 是过滤器，在二维卷积操作中，过滤器 f 通常是一个小的、深度等于输入特征图深度的三维数组。它的尺寸通常是 $k*k*c$，这里 k 是过滤器在空间维度上的大小，而 c 同样是输入通道数，h'和 w'为输出特征图的高度和宽度[18]。b 为偏置项，对于每个过滤器都有一个对应的偏置值。常规卷积虽然有效，但是计算成本高，因为它涉及到的每个输出通道都要执行一次完整的卷积操作。

$$Y = X * f + b \tag{1}$$

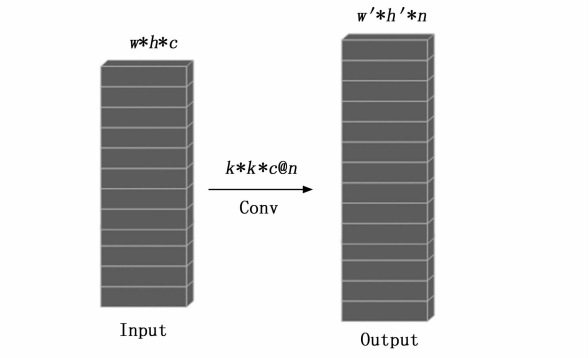

图 2　常规卷积

如图 3 所示，Ghost 卷积的核心思想是减少执行实际卷积操作的通道数量，以此降低计算成本。在 Ghost 模块的设计中，首先采用一个小核尺寸的卷积操作来生成一个初步的特征图。随后，不通过完整的卷积核处理，而是利用一系列简单的线性变换，如平移和旋转来生成额外的特征图。这种方法有效减少了传统卷积操作所需的计算资源。由于这些衍生特征图仅涉及轻量级的操作，相较于标准卷积，它们显著降低了处理过程中的计算负担。此外，这种创新的设计不仅提高了计算效率，还保持了模型的表征能力。通过精心设计的线性操作来扩展特征集，Ghost 卷积能够在不牺牲性能的前提下，有效地压缩模型规模和加速推理过程。

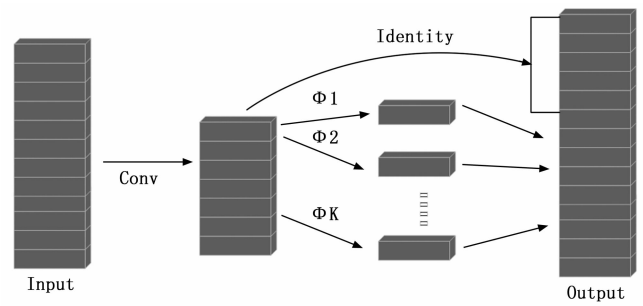

图 3　Ghost 卷积

图 4 展现了 Ghost 模块架构的细节。在图 4（a），通过两个 Ghost 模块串接形成主路径，第一个模块增加通道数，第二个恢复至初始大小，与 ResNet（Residual Network）的跳跃连接方式相似，实现了输入输出尺寸的一致，便于集成到其他 CNN 中。图 4（b）的区别在于两 Ghost 模块间加入步长为 2 的深度卷积，实现特征图尺寸减半，跳跃连接也相应降采样，保证操作对齐。该模块能替换 CNN 中的降采样层。Ghost 模块的理论加速比用以下公式（2）表示。r_s 基于输入特征图尺寸 $h'*w'$、卷积核大小 $k*k$ 和 s，s 是每个特征图生成的 *ghost* 特征图数量。公式考虑了替换常规卷积核的操作和额外的（$s-1$）个线性操作的计算成本。在 $d*d$ 接近于 $k*k$ 且 $s<<c$ 的条件下，理论加速比近似等于 s。压缩比 r_c 是参数数量减少的独度量，也近似等于 s，与速度提升比相等，意味着利用 Ghost 模块可以在参数量上也大约减少 s 倍。

$$r_s=\frac{n*h'*w'*c*k*k}{\frac{n}{s}*h'*w'*c*k*k+(s-1)*\frac{n}{s}*h'*w'*d*d} \tag{2}$$

简化后得到：

$$r_s=\frac{c*k*k}{\frac{1}{s}*c*k*k+\frac{s-1}{s}*d*d} \tag{3}$$

假设 $d*d$ 与 $k*k$ 的大小相近，并且 s 远小于 c，则加速比 r_s 近似等于：

$$r_s\approx\frac{s*c}{s+c-1} \tag{4}$$

而压缩比 r_c 也类似表示为：

$$r_c=\frac{n*c*k*k}{\frac{n}{s}*c*k*k+(s-1)*\frac{n}{s}*d*d} \tag{5}$$

简化后得到：

$$r_c\approx\frac{s*c}{s+c-1} \tag{6}$$

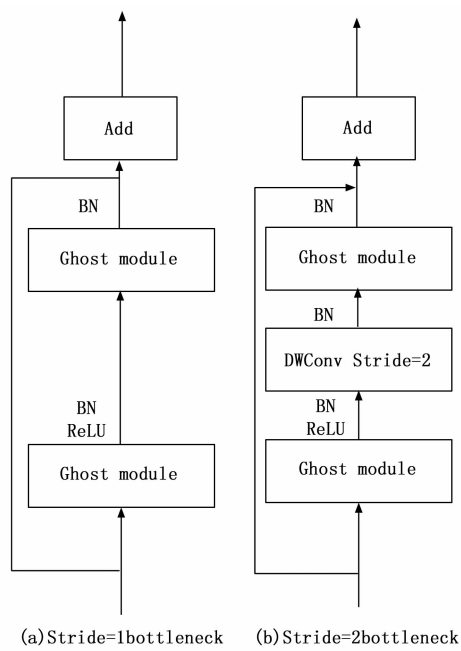

图 4　步长为 1 和 2 的 bottleneck

这表明通过使用 Ghost 模块，理论上能够达到与 s 相当的加速和压缩比，即在使用较少的计算资源的同时仍能达到相当的特征提取能力。

2.2　添加 MHSA 注意力机制

Aravind Srinivas、Tsung-Yi Lin[19] 等人提出一种概念简单、功能强大的主干架构，名为 BoT Net（Bottleneck Transformer Network）。如图 5（a）、（b）所示，

图（a）为 ResNet Bottleneck，图（b）为 Bottleneck Transformer，其原理是用 MHSA 替换 Rest Net 中的 3 * 3 卷积层。

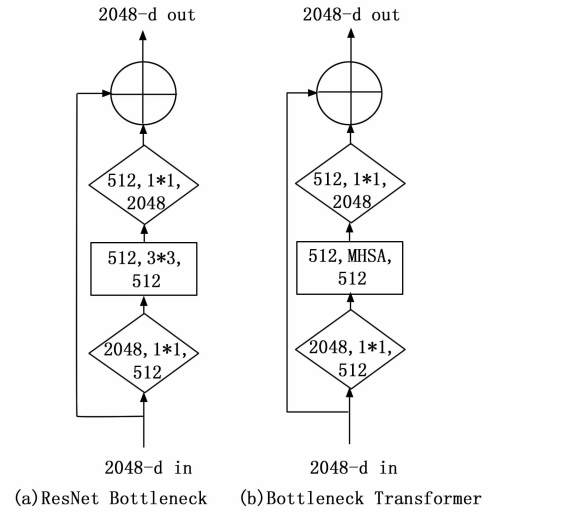

图 5 ResNet Bottleneck 与 BoT Net 网络结构

由图 6 可知，MHSA 输入一个维度 $H * W * d$ 的特征图 X，其中 H 和 W 分别代表特征图的高度和宽度，d 是每个特征点特征向量的维度。特征图 X 经过 3 个不同的 $1 * 1$ 的卷积分别生成 q、k、v。对于自注意力中的每个元素，还需要通过 R_h 和 R_w 实现一个相对位置编码。这些编码使模型能够考虑元素在空间上的位置。接下来计算两个注意力得分，一个 qk^T 另一个是 qr^T。这两个得分相加，为每个位置提供一个得分，这个得分表明在生成输出时应该给予输入特征图中的其他位置多少关注。再通过 softmax 函数对注意力得分进行归一化，归一化后的注意力权重对值 v 进行加权和聚合，最终得到输出特征图 Z。

因此多头自注意力机制的引入，主要是为了提升模型的表达能力和泛化能力。该机制通过并行运用多个注意力头，独立地计算不同视角的注意力权重。每个注意力头聚焦于输入数据的不同部分，通过这种方式，模型能够捕捉到更多层次和维度的信息。在得到各注意力头的输出后，这些结果会被综合起来，通过拼接或加权求和的方式，以形成一个综合性更强、信息丰富度更高的表示。如图 7 所展示的，本文研究中，将此 MHSA 模块整合至 Neck 网络架构的 CSP 层输出位置，这一设计使得网络在处理复杂特征时更加有效，进一步增强了模型对于各种输入数据的适应性和响应能力。

2.3 修改损失函数

在目标检测技术中，边界框回归是一个关键步骤，它直接影响到检测的准确度和效率。传统的目标检测算法常用的损失函数，如 SmoothL1、IoU（Intersection over Union）及 GIoU（Generalized Intersection over U-

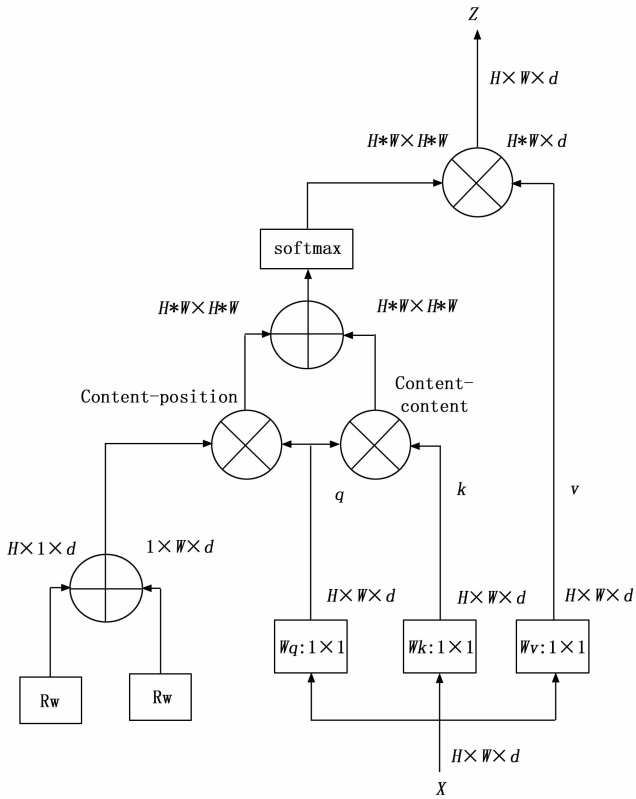

图 6 MHSA 层网络结构

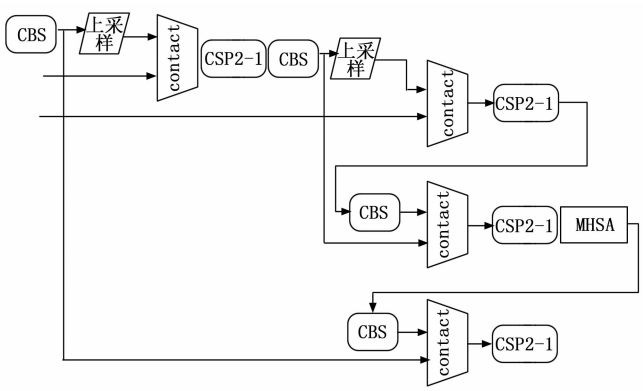

图 7 添加 MHSA 模块

nion），主要通过增加检测框与目标框的重叠区域来优化模型性能，并解决当检测框与目标框完全不重叠时的问题。然而，GIoU 在预测框与目标框的差集大小相同时，效果与 IoU 相同，这限制了其在某些情况下的应用效果。为了克服这一局限，Zhao Hui Zheng [20] 等提出了 DIoU 损失函数。DIoU 损失不仅考虑了传统的交并比，还引入了预测框与目标框中心点之间的归一化距离，从而不仅仅是增加两个框的重叠区域，而是更加注重使预测框的中心点靠近真实框的中心点。这种方法极大地提升了目标的精确定位能力，尤其是在需要高精度大小和位置确定的应用场景中，如机器人导航、自动驾

驶车辆的环境感知以及高级视频监控系统。DIoU 通过这种细致的中心对齐机制，显著提高了模型在复杂环境中的表现和可靠性。

在传统的交并比计算基础上引入了一个新的正则项，即预测框与真实框中心点之间的欧几里得距离 $\rho(bp, bg)$。此距离定义为包含两个框的最小闭包框的对角线长度，其计算公式如式（7）：

$$DIoU = IoU - \frac{\rho(b^p, b^g)}{c^2} \tag{7}$$

其中：减去的部分即 $\frac{\rho(b^p, b^g)}{c^2}$ 是中心点距离的归一化项，通过将中心点距离的平方除以对角线长度的平方，我们得到一个范围在 0 到 1 之间的值，它衡量了预测框和真实框的中心对齐程度。中心对齐越好，这个值越小，$DIoU$ 值越大，损失函数 L 的值就越小，损失函数 L 的定义为式（8）：

$$L = 1 - DIoU = 1 - IoU + \frac{\rho(b^p, b^g)}{c^2} \tag{8}$$

3　实验与分析

3.1　数据集的构建

本次实验的数据集一共包含 1 986 张图片，数据集来源于 ImageNet，它是一个大型的图像数据库，包含数百万的标注图像，这些图像分布在数千个类别中。ImageNet 的目标是为了各种形状、尺寸和颜色的物体提供大量的视觉样本。数据集中包含了七类标签，分别是：月牙形缺口（Crescent Gap）、丝状斑点（Silk Spot）、水渍斑点（Water Spot）、焊缝线（Weld Line）、油渍斑点（Oil Spot）、冲孔（Punching）、腰折（Waist Folding）。本实验使用的数据集图片如图 8 所示，图中包含了全部待检测的缺陷类别。

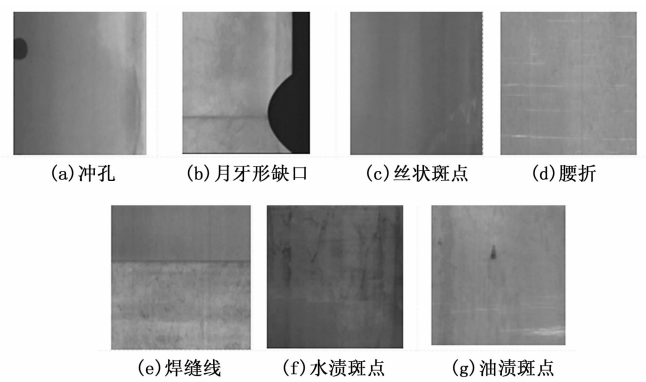

图 8　全部表面缺陷示意图

3.2　数据集的处理

首先使用 Labelimg 软件进行图像标注，并将生成的 XML 文件存储在 labels 目录中。接着，利用 Python 脚本将这些 XML 文件转换成 YOLO 格式可识别的 TXT 文件。完成转换后，根据 8：1：1 的比例，将图片和对应的 TXT 文件分配至训练集、验证集和测试集，数量分别为 1 588、199 和 199。

3.3　实验平台搭建及流程

表 1　实验环境

名称	配置
操作平台	Ubuntu18.04.6
CPU	IntelCorei7-7800XCPU
GPU	NVIDIAGeForceRTX1080Ti 显卡 * 2
显存	22 G
内存	31 G
编译环境	Python

服务器相关配置如下：实验操作平台为 Ubuntu18.04.6、CPU 为 IntelCorei7-7800XCPU、GPU 为 NVIDIAGeForceRTX1080Ti 显卡 * 2，显存 22 G，内存 31 G、软件框架使用 Pytorch、编程环境为 Python 语言。

3.4　模型参数指标及参数设置

实验所使用的参数指标：查准率 P（Precision）、查全率 R（Recall）、mAP（Mean Average Percision）、权重文件大小（MB）、检测帧数（FPS）。其中 P、R、mAP 参数指标计算公式如下所示：

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} AP = \int_0^1 P(R)\,\mathrm{d}R \tag{11}$$

在目标识别系统的性能评估中，TP（True Positives）定义为正确识别的目标数量；FP（False Positives）表示误标为目标的实例数量；FN（False Negatives）指未被检测到的实际目标。查准率 P 计算了模型正确识别目标的比例相对于所有被识别为目标的实例。与此同时，平均精度值 mAP 作为一个更全面的评价指标，能够在不同的判定阈值下评估模型的整体性能。通常，较高的 mAP 值表明模型在整体上表现更加出色。输入图片尺寸为 640 * 640，初始学习率设置为 0.01，交并比阈值（IoU）设置为 0.7，本文中所有实验均是在不加载相应模型的预训练权重下进行。

3.5　实验结果评估分析

1）对比实验：

本实验参考了已发表有关金属表面缺陷检测的数据，选取了一些主流算法进行对比，例如 Faster-RCNN、SSD、YOLOv2、YOLOv3、YOLOv5。表 2 为主流算法的对比，直观的展现出 YOLOv5 算法不论是在

查准率、查全率、*mAP*、权重文件大小方面都具有一定优势。

表 2 主流目标检测算法对比

算法	查准率/%	查全率/%	*mAP*/%
Faster-RCNN[21]	60.98	70.78	64.69
SDD[22]	54.34	60.44	58.45
YOLOv2[22]	48.68	53.09	52.02
YOLOv3[22]	59.60	63.78	60.15
YOLOv4[23]	-	52.00	58.76
YOLO5s[23]	—	78.70	82.40
YOLOv5m[23]	—	73.40	80.10
YOLOv5l[23]	—	77.00	78.40
YOLOv5x[23]	—	72.60	80.00
本文算法	85.30	79.70	85.10

2）注意力机制实验对比：

为了证明更换 MHSA 注意力机制的有效性，本实验将 MHSA 注意力模块增加到 Neck 网络，在整个原网络的第 21 层，C3 模块后。并与加入 SE、ECA、CBAM 模块进行对比实验，实验结果如表 3。实验结果表明，加入 MHSA 注意力模块，模型能得到较高的 *mAP*，但在检测速度方面会小幅度降低。加入 SE 模块后，模型的总体精度低于原网络 0.8%，但检测速度却有大幅提升。SimAM、CBAM 模块在 *mAP* 方面也各有提升，综合对比实验结果后，将 MHSA 加入原网络。

表 3 注意力机制对比实验结果

算法	查准率/%	查全率/%	*mAP*@0.5/%	FPS
YOLOv5s	84.4	80.4	81.9	277
YOLOv5s+MHSA	82.9	82.1	84.6	238
YOLOv5s+SE	85.8	78.1	81.1	333
YOLOv5s+SimAM	84.0	80.7	84.0	277
YOLOv5s+CBAM	85.7	79.7	83.2	208

3）Ghost 网络实验对比：

本文为了验证 Ghost 网络的有效性，将 Backbone 网络更换为 Ghost 网络，实验结果如表 4 所示。通过实验结果表明，替换 Ghost 网络后，模型的检测精度提升了 1.9%。

表 4 Ghost 网络实验对比

算法	查准率/%	查全率/%	*mAP*@0.5/%	FPS
YOLOv5s	84.4	80.4	81.9	277
YOLOv5s+Ghost	79.0	84.4	83.8	131

4）损失函数实验对比：

本实验提出使用 DIoU、GIoU、SIoU、EIoU 来代替原网络中的损失函数 CIoU 分别对比实验，网络的其余部分均不发生任何改变。实验结果如表 5 所示，使用 EIoU 代替 CIoU，模型 *mAP* 下降了 0.1%。SIoU、GIoU、DIoU 较原网络在 *mAP* 上也有一定提升，DIoU 的 *mAP* 提升明显，提升了 4.3%。综合对比实验结果后，最终选择 DIoU 来替代 CIoU。

表 5 损失函数实验对比

算法	查准率/%	查全率/%	*mAP*@0.5/%	FPS
YOLOv5s	84.4	80.4	81.9	277
YOLOv5s+EIoU	80.6	80.5	81.8	244
YOLOv5s+SIoU	80.7	82.7	82.7	322
YOLOv5s+GIoU	83.7	80.1	83.7	263
YOLOv5s+DIoU	86.7	81.0	86.2	277

5）消融实验：

对比 YOLOv5s 原网络，本实验进行了三点改进：将 GhostNet 加入到 Backbone 网络中、将 MHSA 注意力模块加入 Neck 网络，最后修改损失函数 CIoU 为 DIoU。为了观察各个模块组合后对模型的影响，现将 Ghost、MHSA、DIoU 分别组合，表 6 展示了消融实验结果。

表 6 消融实验结果

算法	查准率/%	查全率/%	*mAP*@0.5/%	FPS
YOLOv5s	84.4	80.4	81.9	277
YOLOv5s+Ghost	79.0	84.4	83.8	131
YOLOv5s+MHSA	82.9	82.1	84.6	238
YOLOv5s+DIoU	86.7	81.0	86.2	277
YOLOv5s+DIoU+Ghost	80.6	85.9	84.9	130
YOLOv5s+DIoU+MHSA	82.2	80.7	83.2	233
YOLOv5s+Ghost+MHSA+DIoU	85.3	70.7	85.1	118

消融实验结果表明，更换 Ghost 网络后 *mAP* 提升了 1.9%，加入 MHSA 注意力机制 *mAP* 提升了 2.7%。更换损失函数 DIoU 后 *mAP* 提升了 4.3%。将模块全部加入网络 *mAP* 提升了 3.2%。通过消融实验，验证了该网络在目标检测算法上的有效性。

为了清楚展示改进后网络相对于原网络的性能提升，本文分别对两者在测试集上的表现进行了对比。如图 9 展示，图 9（a）和图 9（b）分别展示了两种网络的测试结果。从对比分析结果可以明显观察到，原网络在识别丝状斑点和水渍斑点时存在漏检现象。这是由于环境光线条件不稳定、表面反射强烈以及金属表面处理不均匀等因素，导致图像捕捉到的内容复杂多变，从而对缺陷的准确检测造成了明显的干扰。在更改 Ghost 网络后，生成更多特征图，从而加强了网络的特征提取能力。

4 结束语

本文探讨了在金属表面缺陷检测过程中，由于工件表面光洁度不一、光线反射不均和环境光照变化等因

(a)原YOLOv5s检测效果　　(b)改进后检测效果

图 9　YOLOv5s 网络与改进后网络效果对比图

素，导致检测图像中的视觉信息复杂化，进而对缺陷识别精度造成显著影响。这些因素会使得基于传统图像处理技术的检测系统在精度上不够高，同时也可能导致漏检和误检问题。如在光线反射强烈的情况下，轻微的丝状斑点或水渍斑点可能会被忽略，而在表面处理不均匀的金属上，非缺陷区域可能被错误识别为缺陷。这些问题直接影响了检测系统的可靠性和实用性。因此本文将 Ghost 网络加入 Backbone 中，增加了网络特征提取能力；添加 MHSA 自注意力机制，增强了模型对数据的理解和特征提取能力，尤其是在处理图像细节时，它允许模型捕捉到图像的不同部分之间的关系，这些关系是通过传统的卷积层难以捕捉的；将原网络中的 CIoU 替换为 DIoU，有助于减少目标中心偏差、最小化边界框之间的距离，从而提高定位精度，加速模型收敛。实验结果表明，3 种模块均加入到原网络中，网络的 mAP 精度提升了 3.2%，达到了 85.1%，证明了该算法的优越性。由于金属表面的不同反射率、加工过程中的磨损以及操作环境中的尘埃积聚等因素，误检和漏检问题经常发生，进而影响整体检测性能。本文提出的算法在应对这些挑战方面仍有提升空间。后续研究将深入探讨在数据集不足、光照不足、图像分辨率低、表面磨损严重以及环境尘埃影响等不同情形下，进一步优化该检测系统的精确度。

参考文献：

[1] 吕秀丽，卢海滨，侯春光，等. 改进 YOLOv5s 的钢材表面缺陷检测算法 [J]. 化工自动化及仪表，2024，51 (2)：301－309.

[2] 李　伟，姚小敏，张鹏超，等. 改进的 YOLO-V7 钢材表面缺陷检测算法研究 [J/OL]. 机械科学与技术：2024. 1－10.

[3] 王　涵，刘海明，邵雨虹. 一种改进 YOLOv5 算法的金属表面缺陷检测研究 [J/OL]. 机械科学与技术：2024. 1－6.

[4] 王亚珅，葛悦涛，鞠卓亚，等. 2023 年深度学习技术主要发展动向分析 [J]. 无人系统技术，2024，7 (1)：50－58.

[5] 潘烨新，黄启鹏，韦　超，等. 基于注意力机制的 YOLOv5 优化模型 [J]. 计算机技术与发展，2023，33 (12)：163－170.

[6] 渠　逸，汪　诚，余嘉博，等. 基于 YOLOv5 的表面缺陷检测优化算法 [J]. 空军工程大学学报，2023，24 (5)：80－87.

[7] YAN P C，SUN Q S，YIN N N，et al. Detection of coal and gangue based on improved YOLOv5.1 which embedded ScSE module [J]. Measurement，2022，188：110530.

[8] 李跃华，仲　新，姚章燕，等. 基于改进 YOLOv5s 的着装不规范检测算法研究 [J/OL]. 图学学报，2024－03－14：1－12.

[9] YING Z P，LIN Z T，WU Z Y，et al. A Modified-YOLOv5s model for detection of wire braided hose defects [J]. Measurement，2022，190：110683.

[10] BOCHKOVSKIY ALEXEY，WANG C Y，LIAO H Y. YOLOv4：optimal speed and accuracy of object detection [J]. In：Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，2020，Online virtual hosting，126－138.

[11] FU G，SUN P，ZHU W，et al. A deep-learning-based approach for fast and robust steel surface defects classification [J]. Optics and Lasers in Engineering，2019，121：397－405.

[12] 王浩然，李廷会，曹玉军，等. 改进 YOLOv5s 网络在缺陷检测中的应用 [J]. 网络新媒体技术，2022，11 (2)：58－65.

[13] WANG C Y，LIAO H Y M，Wu Y H，et al. CSPNet：A new backbone that can enhance learning capability of CNN [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops，2020：390－391.

[14] 杨有为，周　刚. 面向自然场景文本检测的改进 NMS 算法 [J]. 计算机工程与应用，2022，58 (1)：204－208.

[15] HAN KAI，WANG YUNHE，TIAN QI，et al. GhostNet：More features from cheap operations [C] // In：Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，2020，1577－1586.

[16] 王海云. 基于深度卷积神经网络的金属板带材表面缺陷检测研究与应用 [D]. 昆明：昆明理工大学，2021.

[17] 张　旭. 基于深度卷积神经网络的铝材表面缺陷检测 [D]. 上海：华东理工大学，2020.

(下转第 63 页)