激光与光电子学进展 先进成像

一种面向钢材表面缺陷检测的改进型YOLOv5算法

李少雄[1]，史再峰[1,3*]，孔凡宁[1]，王若琪[1]，罗韬[2]
[1]天津大学微电子学院，天津 300072；
[2]天津大学智能与计算学部，天津 300072；
[3]天津市成像与感知微电子技术重点实验室，天津 300072

摘要 针对钢材表面缺陷尺度不一，现有检测算法多尺度特征处理能力较差、精度有待提高的问题，提出了一种面向钢材表面缺陷检测的改进型YOLOv5算法。首先，在骨干网络的特征输出层后添加感受野模块以增强特征的判别性与鲁棒性，可以更好地感知不同尺度的特征信息；然后，利用对齐的特征聚合模块替换传统的特征融合结构，解决了高低分辨率特征图在融合过程中存在的特征错位问题；最后，采用带有高效通道注意力机制的解耦头输出检测结果，注意力机制可以自适应地校准通道响应，解耦头使得分类与回归任务可以独立执行。在NEU-DET数据集上的实验结果显示，所提出方法的平均精度均值为80.51%，相比基准模型提升了4.48%，检测速度为31.96 frame/s。相比其他主流的目标检测算法，在保持一定检测速度的前提下，所提算法具有更高的精度，能够实现高效的钢材表面缺陷检测。

关键词 表面缺陷检测；感受野；特征对齐；解耦头；注意力机制
中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP230711

An Improved YOLOv5 Algorithm for Steel Surface Defect Detection

Li Shaoxiong[1], Shi Zaifeng[1,3*], Kong Fanning[1], Wang Ruoqi[1], Luo Tao[2]
[1]*School of Microelectronics, Tianjin University, Tianjin* 300072, *China*;
[2]*College of Intelligence and Computing, Tianjin University, Tianjin* 300072, *China*;
[3]*Tianjin Key Laboratory of Imaging and Sensing Microelectronic Technology, Tianjin* 300072, *China*

Abstract Scale of steel surface defects is different, but existing detection algorithms have poor multi-scale feature processing ability and low accuracy. Therefore, an improved YOLOv5 algorithm for steel surface defect detection is proposed. First, receptive field modules are added after the feature output layer of the backbone to enhance the discrimination and robustness of the features which can better perceive the feature information of different scales. Then, aligned feature aggregation modules are used to replace the traditional feature fusion structure to solve the feature misalignment problem in the fusion process of high and low resolution feature maps. Finally, decoupled heads with efficient channel attention mechanisms are used to output the detection results. The attention mechanism can adaptively calibrate the channel response, and the decoupled heads enable classification and regression tasks to be performed independently. The experimental results on NEU-DET dataset show that the mean average precision of the proposed method is 80.51%, which is 4.48% higher than that of the benchmark model, and the detection speed is 31.96 frame/s. Compared with other mainstream object detection algorithms, the proposed algorithm has higher accuracy while maintaining certain detection speed, enabling efficient steel surface defect detection.

Key words surface defect detection; receptive field; feature alignment; decoupled head; attention mechanism

1 引 言

工业生产的钢材表面往往会出现裂纹、划痕等缺陷，这些缺陷不仅影响美观，还会降低钢材的抗腐蚀性，损害产品质量[1]。采用人工目视的检测方法来筛选良品是工厂的常用手段，但人工检测效率低下，工人的主观判断也使得检测标准无法统一，进而导致检测结果不稳定。

收稿日期：2023-02-27；**修回日期**：2023-03-18；**录用日期**：2023-04-07；**网络首发日期**：2023-04-17
基金项目：天津市科技计划项目(22JCYBJC00140)
通信作者：*shizaifeng@tju.edu.cn

随着机器视觉技术的快速发展，人们开始将其运用到缺陷检测领域。光学设备用于采集图像，图像处理算法用于提取图像信息，根据获得的特征信息即可识别缺陷[2]。然而人工设计的特征提取算法在鲁棒性和泛化能力上表现不佳，这会降低检测的准确率。近年来，凭借着强大的特征提取能力，基于深度学习的目标检测算法逐渐成为解决问题的新选择。具体来说，主要是对两阶段检测器 Faster R-CNN[3]以及单阶段检测器 SSD[4]、RetinaNet[5]、YOLO 系列算法[6-7]等进行改进以更好地适应缺陷检测任务。Liu 等[8]提出了一种基于 Faster R-CNN 的多尺度上下文钢材缺陷检测网络，由空洞卷积构成的并行卷积架构用于捕获多尺度上下文信息，特征增强与选择模块既可以增强特征的判别性也可以减少信息混淆。罗晖等[9]利用图像处理算法提高缺陷目标与背景的对比度，并改进损失函数，改进后的 Cascade R-CNN 算法的精度相较于基准模型更高。Cheng 等[10]在 RetinaNet 中添加通道注意力机制以尽可能地保留特征信息，添加空间特征融合模块来增强深层和浅层特征信息的交互，最终达到更好的缺陷检测性能。孙连山等[11]在 YOLOv3 中添加了另一级特征尺度来获取小目标的特征信息，设计出轻量级注意力引导模块以增强网络的特征提取能力，使用 K-medians 算法更准确地描述了锚框尺寸的分布规律。Yu 等[12]提出了一种基于 YOLOv4 的高效尺度感知缺陷检测网络，该模型侧重于强化包含丰富几何信息的浅层特征以减少微小目标的信息损失，此外，还提出一种具有动态感受野的检测头以缓解检测头中感受野与目标尺度不匹配的问题。程松等[13]提出了一种适用于嵌入式设备的轻量型 YOLOv5 缺陷检测算法，模型中大量引入深度可分离卷积，最终在降低参数的同时提升了检测精度。总的来说，使用深度学习的方法进行缺陷检测的具体步骤一般是先利用骨干网络提取特征信息，再进行特征融合，最后使用相应的检测算法实现缺陷检测。

工业生产中的缺陷往往存在两个特点，一是同类型缺陷之间形状差异较大，二是不同类型缺陷之间特征信息相似[14]。上述模型的设计初衷大多是缓解因此而引发的检测精度低的问题，它们习惯于在特征融合网络中添加相关模块以促进特征融合。本模型的设计思路更偏向于改进基准模型中存在的固有缺陷，并增强模型的多尺度表征能力，进而提升检测精度。为此，本文提出了一种基于改进型 YOLOv5 的缺陷检测模型。首先，在骨干网络的输出特征层后添加感受野模块（RFB）[15]以拓宽感受野范围，获取丰富的多尺度信息，进而增强模型的表征能力。然后，对特征融合网络中简单上采样后的特征图的合并操作进行改进，利用对齐的特征聚合模块（AFAM）[16]来学习像素的变换偏移量，从而解决相邻级别特征图间的特征错位问题，便于上下文特征信息融合。最后，在预测网络中使用带有高效通道注意力（ECA）机制[17]的解耦头（Decoupled head）[18]分离出分类任务与回归任务，进一步提升检测的准确率。实验结果表明，改进后模型的精度有所提升。

2 YOLOv5 算法

受跨阶段局部网络（CSPNet）[19]的启发，YOLOv5 以 C3 模块为核心构建骨干网络。C3 模块将基础的特征层划分为两部分，这两部分特征图各自经过不同的操作之后再合并，这样就能避免学习过多重复的梯度信息，最终在减少计算量的同时增强模型的学习能力。在骨干网络的最后一级输出特征后往往还会添加 SPPF（Spatial Pyramid Pooling-Fast）模块以扩大感受野，它将一个卷积层与三个内核大小相同的池化层串联起来并整合特征映射。为融合来自不同层次的特征信息，YOLOv5 结合特征金字塔网络（FPN）[20]和路径聚合网络（PANet）[21]的特点构建特征融合网络。高层特征图经上采样操作后会沿着自上而下的路径与低层特征图融合，这可以使获得的特征同时拥有较多的几何和语义信息。低层特征图经下采样操作后会沿着自下而上的路径与高层特征图融合，由于特征信息的交互行程被缩短，特征金字塔的优势被进一步放大。为应对检测任务中不同大小的目标，YOLOv5 采用多尺度预测策略，预测网络将特征图划分成一定数量的网格，当目标中心落在某网格内时，则由该网格来预测目标。最终输出特征图的通道数为 $3\times(5+N)$，其中：3 为锚框数量；5 包含边界框的中心坐标 (x,y)、框的宽和高以及目标的置信度；N 为类别数。

然而，通用目标检测算法的应用场景通常是常规自然图像。工业生产环境下的钢材表面缺陷往往存在尺度多变的特点，因此单纯运用基础的目标检测技术并不能良好地适配于缺陷检测，需要对模型结构进行针对性优化。

3 改进后的 YOLOv5 算法

3.1 整体网络结构

选择 YOLOv5-m 作为基准模型并对其进行改进。如图 1 所示，改进后的 YOLOv5 整体结构主要由特征提取网络、特征融合网络和预测网络三部分组成。图像首先被输入到由普通卷积层、C3 模块、RFB 以及 SPPF 模块构成的特征提取网络中。普通卷积层用于压缩图像尺寸，C3 模块将特征图分别输入到两路分支中，一路分支仅含有单个卷积层，另一路分支由单个卷积层及多个残差块[22]堆叠而成，再将经过两路分支操作后得到的特征映射合并即可。一个下采样卷积层之后往往会堆叠一个 C3 模块，多次下采样操作之后即可形成初步的骨干网络。RFB 和 SPPF 模块的作用均是扩大骨干网络输出特征的感受野。特征融合网络接收来自前级网络的多尺度特征图，通过双向路径实现特

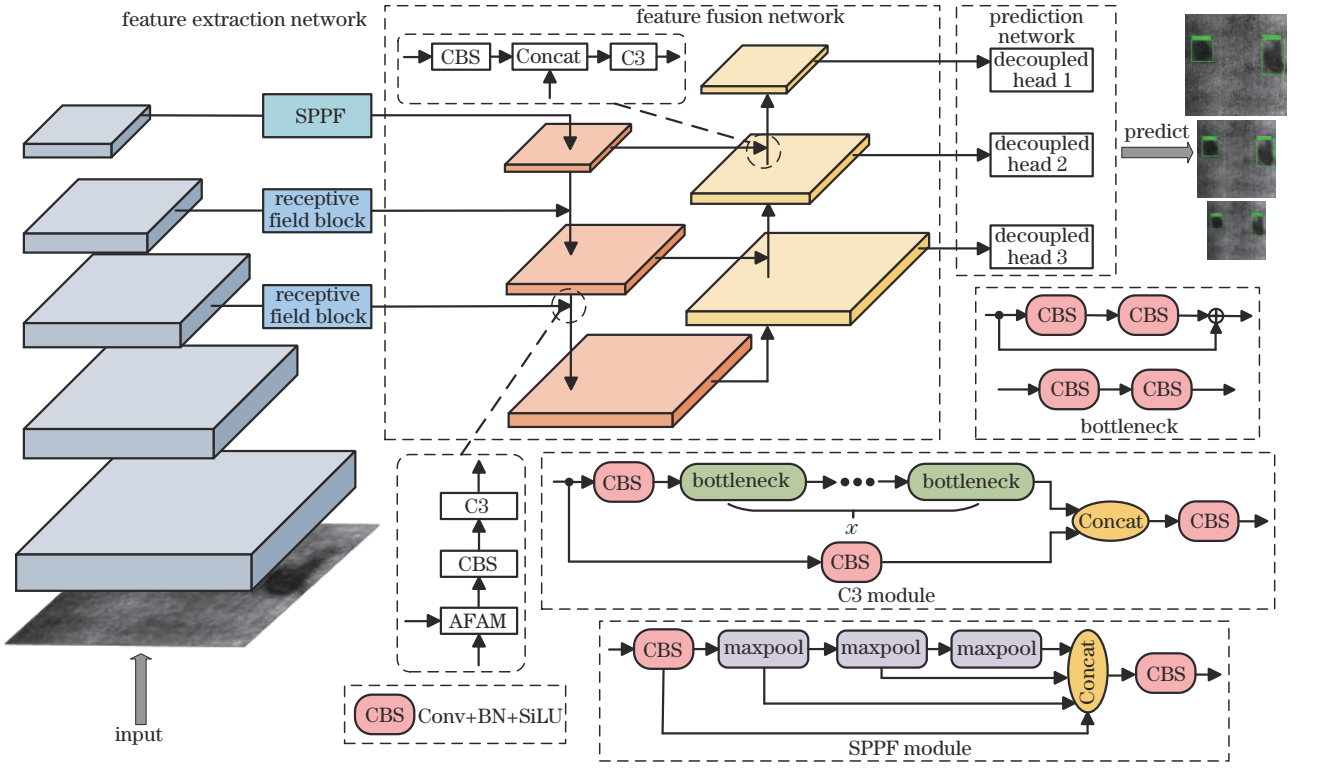

图 1　YOLOv5 整体网络结构

Fig. 1　Overall network architecture of YOLOv5

征融合。其中，自上而下的路径主要由对齐的特征聚合模块以及 C3 模块构成，自下而上的路径主要由执行下采样操作的卷积层和 C3 模块构成。值得注意的是，这里 C3 模块中的残差块取消了捷径连接，仅由多个堆叠的卷积层构成。最后预测网络接收到来自特征融合网络的特征输出，利用带有 ECA 的解耦头在不同尺度的特征图上实现最终的分类与回归。

3.2　RFB

考虑到缺陷之间尺度和形状差异过大的问题，需要进一步提升网络的多尺度特征提取能力。之前的工作大多都是通过构建更深层次的骨干网络来获取更多特征，提升模型的表征能力，但这样做会引入较多参数。另一方面，以下采样的方式来扩大感受野往往伴随着图像尺寸的缩小，这样就会丢失特征信息。好在人类视觉系统中的感受野结构为解决此类问题提供了灵感。研究人员发现感受野大小与偏心率之间存在某种函数关系，便基于 Inception 结构[23]和空洞卷积[24]设计出了 RFB。该模块在多个分支上执行卷积操作以更好地捕获多尺度信息，但由于这些操作都在同一个中心采样，使得获得的特征信息缺少判别性，这时再采用具有不同扩张率的空洞卷积来模拟不同偏心率对感受野的影响，将得到的特征映射进行整合，最终产生类似于人类视觉系统的感受野阵列。

RFB 的细节如图 2 所示，前级输入首先会经过类似 Inception 的多分支卷积结构，通过内核大小分别为 1、3、5 的卷积来形成不同大小的感受野，捷径连接也

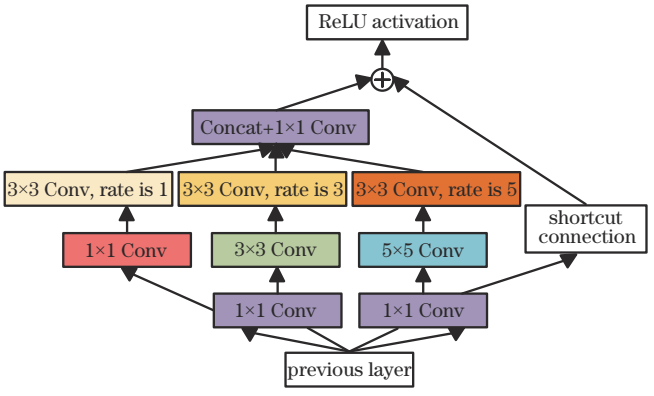

图 2　RFB 结构

Fig. 2　Structure of RFB

被引入到模块当中。由于卷积核的大小与扩张率之间也具有类似于感受野大小与偏心率之间的正相关关系，接收到来自前层的特征信息后，通过在卷积核中填充相应个数的空格形成带有不同扩张率的空洞卷积，利用这些空洞卷积为每个分支分配一个特定的偏心率，就可以重现感受野的偏心率对人类视觉系统的影响。最后，在通道维度将特征图拼接起来并利用 1×1 卷积融合特征信息，再将整合后的特征映射与捷径连接相加并激活响应。YOLOv5 的特征提取网络中有三个输出特征层，而最后一个输出特征层会经过 SPPF 模块来扩大感受野，因此将 RFB 仅添加到前两个输出特征层的末尾以更好地描述不同尺度的对象。

3.3　AFAM

反复执行下采样和特征提取操作导致各级特征图中表达出的特征信息有差异。浅层特征中蕴含的语义信息较少，但位置信息明确；深层特征的语义信息表达充分，但细节有所缺失。因此，在模型中进行多尺度特征融合是提升检测精度的常用手段，其中具有代表性的结构就是FPN和PANet，二者通过上下采样操作对相邻级别的特征图在通道维度进行合并，然后再利用卷积操作融合信息。然而，常用上采样操作的不可学习性以及上下采样操作的反复使用会使得特征间出现错位问题，最终降低检测性能[16]。AFAM采用一种简单可学习的插值策略学习像素之间的变换偏移量，通过卷积操作学习到的偏移量既可以指导特征对齐，也可以减少高分辨率特征中过多的细节特征，进而方便特征聚合。因此采用该AFAM来取代传统上采样后的特征融合步骤以提升检测精度。

以经过SPPF结构处理后的最高层特征图与相邻级别次高层特征图的融合过程为例，传统的融合步骤是先利用1×1卷积对最高层特征图执行通道降维与信息整合操作获得$\boldsymbol{F}_n$，对$\boldsymbol{F}_n$进行上采样操作后再与邻级特征图$\boldsymbol{F}_{n-1}$在通道维度进行合并，并将合并后的特征映射输入到C3模块中。而AFAM实现特征融合的方式如图3所示，其前序步骤与传统的融合方式相同，区别在于合并之后的特征图会通过几个卷积层获得两个二维偏移量映射$\Delta\boldsymbol{F}_n$和$\Delta\boldsymbol{F}_{n-1}$，高低层次的特征图分别学习对应的二维偏移量从而指导自身特征图的相应空间点位进行动态调整，使得不同分辨率的特征图上的像素点建立起位置对应关系，这样就可以精确对齐特征，最后再对两级特征图执行逐元素相加操作，即可高效完成特征聚合。相关公式为

$$\tilde{\boldsymbol{F}}_n = u\left[U_S(\boldsymbol{F}_n), \Delta\boldsymbol{F}_n\right] + u(\boldsymbol{F}_n - 1, \Delta\boldsymbol{F}_n - 1), \quad (1)$$

$$\boldsymbol{U}_{hw} = \sum_{h'=1}^{H}\sum_{w'=1}^{W}\boldsymbol{F}\,h'w'\max\left(0, 1 - \left|h + \Delta_{1hw} - h'\right|\right)\max\left(0, 1 - \left|w + \Delta_{2hw} - w'\right|\right), \quad (2)$$

式中：$U_S(\cdot)$表示对最高层特征图进行双线性插值上采样操作；$u(\cdot,\cdot)$为根据偏移量映射来指导特征对齐的函数。式(2)为对齐函数的具体实现过程，其中：$\boldsymbol{U}_{hw}$是对齐函数$u(\cdot,\cdot)$的输出；H和W为特征图的高和宽；h'和w'为空间坐标的具体取值；最高层特征图上采样点的位置为$(h+\Delta_{1hw}, w+\Delta_{2hw})$；$\Delta_{1hw}$和$\Delta_{2hw}$为点位$(h, w)$学习到的二维偏移量。

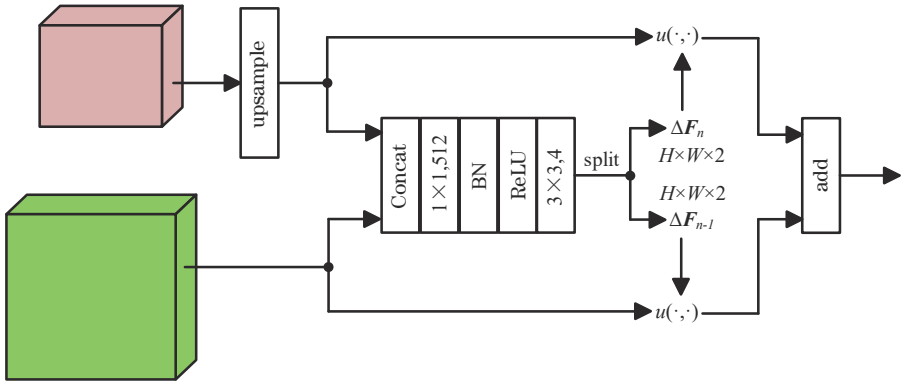

图3　AFAM结构

Fig. 3　Structure of AFAM

3.4　解耦头与注意力机制

目前，YOLO系列检测器大多采用的是耦合头(sibling head)输出检测结果，它通过普通卷积操作在同一个分支上执行分类与回归任务。然而，越来越多的研究表明，分类任务与回归任务之间存在冲突。Song等[25]发现这两种任务对特征空间的敏感区域不同，分类任务用来学习判别性的特征，它对目标的特定区域敏感，某些显著区域的特征可能包含丰富的分类信息；回归任务则用来精确定位，它对目标边界区域敏感，边界附近的特征更有利于位置回归。最近，YOLOX[18]也证明了使用解耦头可以进一步提升精度。因此，将原有的耦合头更换为解耦头，如图4(a)所示。解耦头接收特征融合网络输出的特征图，这些特征映射首先经过1×1卷积层，然后使用两个并行分支将分类任务与回归任务分离，并在回归分支上添加了置信度分支。为进一步减少计算成本，每个分支上仅执行两次卷积操作。

另外，在每个并行分支的3×3卷积层之前添加ECA机制，它基于一种通道交互策略对通道特征响应进行自适应校准，从而有选择地保留重要的特征映射，丢弃无用的特征映射。如图4(b)所示，经过全局平均池化之后，ECA对每个通道及其邻近通道的特征信息进行分析，进而捕获局部跨通道的交互信息。为了提高效率，ECA使用快速一维卷积生成通道权重，使得所有通道共享相同的学习参数。最后，将权重附加到特征映射上，得到最终的输出。局部跨通道之间信息交互的覆盖范围可以用内核大小来表示，换句话说，对某个通道进行注意力预测时，与此相关的相邻通道的

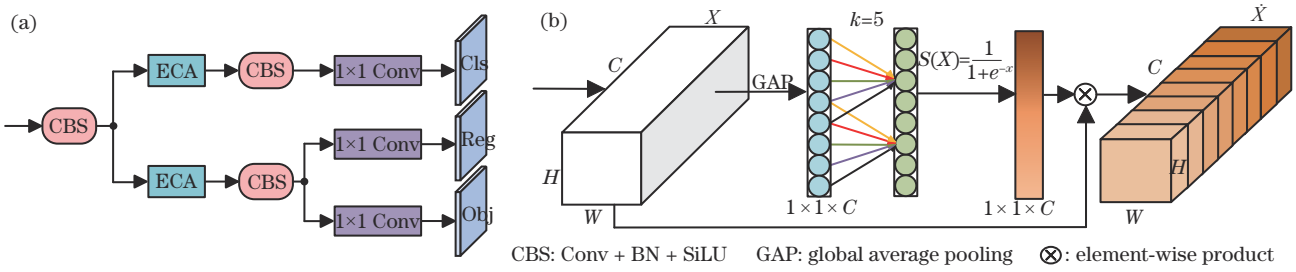

图 4　解耦头与注意力机制。(a)解耦头;(b) ECA

Fig. 4　Decoupled head and attention mechanism. (a) Decoupled head; (b) ECA

个数就等价于覆盖范围的大小。相关公式为

$$\boldsymbol{W}_k=\begin{bmatrix} w^1 & \cdots & w^k & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w^1 & \cdots & w^k & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & w^1 & \cdots & w^k \end{bmatrix}, \tag{3}$$

$$\omega_i=\sigma\left(\sum_{j=1}^{k} w^j y_i^j\right), y_i^j\in\Omega_i^k, \tag{4}$$

式中:$\boldsymbol{W}_k$为权重矩阵,k为内核大小;ω_i为通道y_i的权重大小,此权重只需考虑通道y_i与其相邻通道的相互作用;w^j为第j个通道对应的权重大小;y_i^j为与通道y_i相邻的第j个通道;Ω_i^k为与通道y_i相邻的k个通道的集合。

4　分析与讨论

4.1　实验细节、数据集及评价指标

实验过程中模型训练和测试是在 Windows 10 操作系统、PyTorch 框架下完成的。硬件配置为 Intel Xeon W-2245 CPU @ 3.90 GHz、NVIDIA GeForce RTX 3090 GPU(24 GB)。开发环境为 PyTorch 1.7.0 版本、CUDA 11.0 版本、Python 3.8 版本。实验中采用随机梯度下降(SGD)对网络进行训练,动量设置为 0.937,权重衰减设置为 0.0005,初始学习率设置为 0.01,并使用余弦退火学习率来调整该学习率。批量输入 32 张图片,输入图片尺寸设置为 320 pixel × 320 pixel。此外,在训练过程中对图片进行缩放、翻转、调整对比度等操作以增强模型的鲁棒性。

选择 NEU-DET[26]数据集来证明所提出方法的有效性。NEU-DET 数据集是用于钢材表面缺陷检测的公开数据集,它包含六种类型的缺陷,分别是裂纹(Cr)、夹杂物(In)、斑块(Pa)、点蚀面(Ps)、轧入氧化皮(Rs)和划痕(Sc),如图 5 所示,图片总数量为 1800 张。对数据集进行随机划分,使得训练集与测试集比例为 9∶1,其中训练集会被进一步划分为训练集与验证集,比例也为 9∶1。最终,训练集图片为 1458 张,验证集图片为 162 张,测试集图片为 180 张。

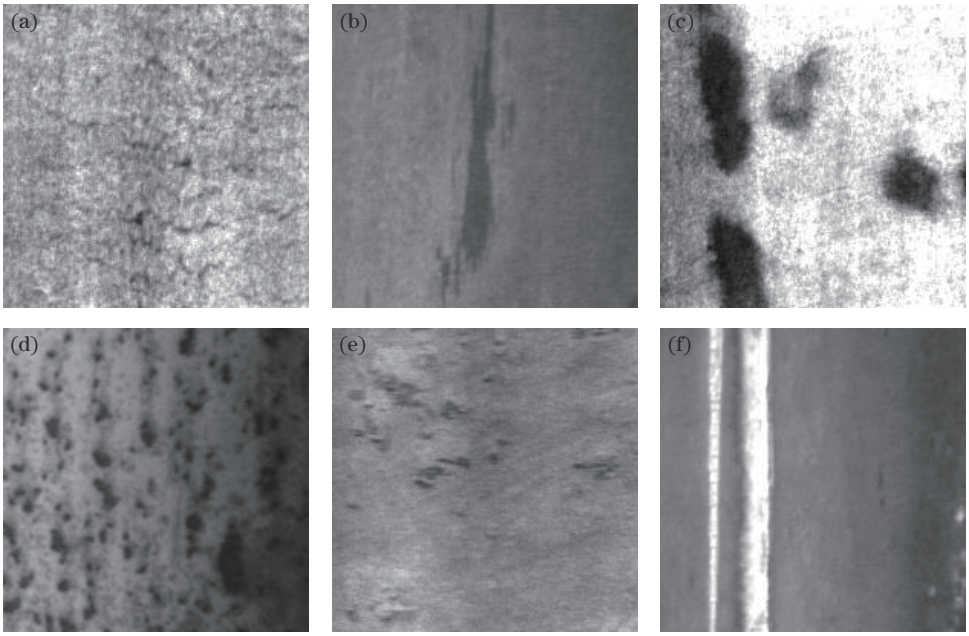

图 5　NEU-DET 数据集。(a)裂纹;(b)夹杂物;(c)斑块;(d)点蚀面;(e)轧入氧化皮;(f)划痕

Fig. 5　NEU-DET dataset. (a) Crazing; (b) inclusion; (c) patches; (d) pitted surface; (e) rolled-in scale; (f) scratches

评价指标方面，采用目标检测领域常用的Average Precision（AP）和mean Average Precision（mAP）来评价模型的精度。AP值表示的是单个类别中所有物体检测精度的平均值，其值等于准确率P和召回率R形成的P-R曲线与坐标轴围成的几何图形的面积。mAP值是取所有类别AP值的平均值。除此之外，利用帧每秒（FPS）来评价模型的检测速度。FPS表示每秒可以检测到的图片数量，它反映了模型的实时检测速度。相关公式为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}},\tag{5}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}},\tag{6}$$

式中，N_{TP}、N_{FP}和N_{FN}分别为真阳性、假阳性和假阴性样本的数量。

4.2　对比实验

为评估模型性能，将主流的目标检测器、同类工作的改进模型以及所提模型在NEU-DET数据集上进行相关实验，并分析结果。图6以损失曲线的下降趋势展示了所提模型在NEU-DET数据集上的训练过程，可以看出，在训练了400个迭代周期后损失值趋于稳定。由表1可以看出，Faster R-CNN的mAP值达到72.29%，但是FPS只有7.56 frame/s，虽然可以保持一定的检测精度，但是由于两阶段检测器中存在较多计算冗余，在速度方面表现不佳。SSD的mAP值为72.66%，FPS高达43.24 frame/s，展现出多尺度检测分支的优势，同时也因取消区域提议（region proposal）阶段提升了检测速度。RetinaNet达到73.37%的mAP值和30.26 frame/s的检测速度，检测精度和速度方面较为平衡。YOLOv3的整体效果与SSD相近。YOLOv4则利用跨阶段局部网络增强了特征提取能力，进一步提升了检测精度，mAP值达到74.43%，同时也保持了一定的检测速度。YOLOv5-s的精度与YOLOv4相近，但整体模型结构较为轻便，检测速度方面展现出很大优势，FPS高达62.75 frame/s。YOLOv5-m相比YOLOv5-s而言，网络结构更深，精度方面表现更好，mAP值达到76.03%，速度方面有所下降，FPS为48.07 frame/s。无锚框检测器YOLOX在基准模型中表现不俗，mAP值达到77.28%，FPS为47.25 frame/s。YOLOv7的mAP值为73.74%，FPS为44.72 frame/s。YOLOv8的整体性能相较于YOLOX更胜一筹，在所有的基准模型中精度表现最好，检测速度仅次于YOLOv5-s。

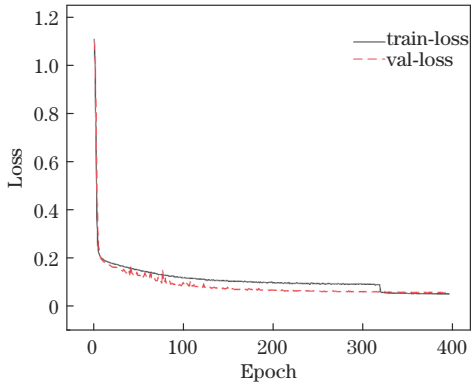

图6　在NEU-DET数据集上的损失曲线

Fig. 6　Loss curves on NEU-DET dataset

表1　不同模型在NEU-DET数据集上的实验结果

Table 1　Experimental results of different models on NEU-DET dataset

Method	AP(Cr) /%	AP(In) /%	AP(Pa) /%	AP(Ps) /%	AP(Rs) /%	AP(Sc) /%	mAP /%	FPS /(frame/s)
Faster R-CNN[3]	35.96	81.80	88.34	81.82	59.31	86.53	72.29	7.56
SSD[4]	38.50	80.78	93.94	82.62	67.96	72.13	72.66	43.24
RetinaNet[5]	36.38	81.62	91.56	81.74	60.81	88.10	73.37	30.26
YOLOv3[6]	34.26	83.88	91.52	81.74	60.91	88.32	73.44	43.67
YOLOv4[7]	31.42	84.61	94.72	81.60	61.36	92.86	74.43	31.15
YOLOv5-s	39.13	79.21	96.46	87.12	60.19	84.19	74.38	**62.75**
YOLOv5-m	38.96	84.50	94.81	87.59	61.31	89.03	76.03	48.07
YOLOX[18]	34.19	84.67	**97.59**	88.72	67.36	90.41	77.28	47.25
YOLOv7[27]	38.97	81.60	93.74	81.61	62.73	83.80	73.74	44.72
YOLOv8	41.21	84.66	93.97	88.43	63.17	**96.66**	78.01	61.33
MSC-DNet[8]	42.40	84.50	94.30	91.50	71.60	92.00	79.38	14.10
DEA_RetinaNet[9]	**60.93**	82.49	94.27	**95.79**	67.16	74.05	79.11	12.20
ES-Net[11]	56.00	**87.60**	88.30	87.40	60.40	94.90	79.10	—
This Research	43.31	87.24	96.53	88.33	**72.80**	94.84	**80.51**	31.96

对比通用目标检测算法，改进后模型在精度方面相比其基准模型均有所提升：MSC-DNet[8]利用上下文增强模块丰富了多尺度信息，mAP值提升显著，高达79.38%；DEA_RetinaNet[9]中添加的通道注意力机制与自适应特征融合模块也增强了特征融合，mAP值达到79.11%；ES-Net[11]则通过加强微小目标的低层

特征有效提升了检测性能，mAP值达到79.10%。最后，相比实验中的其他模型，所提改进后模型在不过分牺牲速度的前提下提升了检测精度，mAP值高达80.51%，FPS为31.96 frame/s。可以看出，改进后的模型在性能上更具优势。

4.3 消融实验

为了进一步探究单个改进点对基准模型的影响，设计了如下的消融实验：将所采用的几种优化方法分别添加到YOLOv5-m中，在各个改进点之间进行组合搭配以探究它们对基准模型的影响。从表2中可以看出，若仅在YOLOv5的骨干网络的特征输出层后增加两个RFB，改进后的YOLOv5相比之前，mAP值提升约2.04个百分点，这说明多尺度模块能够通过扩大感受野捕获更多的上下文信息，更适用于尺度多变的缺陷检测任务。添加RFB后FPS下降为41.08 frame/s，可以看出，并行的多分支卷积结构虽然在训练阶段有助于提升模型的表征能力，但在推理阶段的表现的确逊色于直筒式的网络结构。若仅使用AFAM替换传统的特征融合结构，mAP值相比之前提升约1.02个百分点，这说明该模块有效地缓解了多分辨率特征融合中特征不对齐的问题。改进后模型的FPS降低为43.27 frame/s，这说明相较于简单的上采样融合步骤，学习二维偏移量指导特征图对齐的过程增加了计算成本，进而降低了检测速度。若仅使用解耦头替换耦合头，mAP值提升约1.43个百分点。另外，在此基础上添加注意力机制，发现带有注意力机制的解耦头相比单纯的解耦头效果更好，相比YOLOv5-m方法，其mAP值可提升1.91个百分点。这些结果表明，解耦头很好地解决了分类任务与回归任务的冲突，ECA也利用其特征选择能力进一步增强了性能。检测速度方面，解耦头对基准模型的影响略大于另外两者，这是并行解耦的输出方式需要付出的代价，省去的部分卷积操作就是为了减少速度损失，换取精度与速度的平衡。另外，注意力机制对检测速度的影响可以忽略不计，这与其自身紧凑高效的结构特点是分不开的。除此之外，将这些改进点进行组合并添加到YOLOv5-m中，从表3的结果中可以看出，相比单个改进点而言，组合后的改进点带来的性能提升更大，这说明所提出的优化方法对基准模型的性能影响大致是独立的，它们之间并不会有明显的抑制作用。

表2 消融实验1结果

Table 2 Results of ablation study 1

Method	mAP /%	Improvement /百分点	FPS /(frame/s)
YOLOv5-m	76.03	—	48.07
YOLOv5-m + RFB	78.07	2.04	41.08
YOLOv5-m+AFAM	77.05	1.02	43.27
YOLOv5-m+decoupled head without ECA	77.46	1.43	39.81
YOLOv5-m+decoupled head with ECA	77.94	1.91	39.35

表3 消融实验2结果

Table 3 Results of ablation study 2

RFB	AFAM	Decoupled head with ECA	mAP /%	Improvement /百分点	FPS /(frame/s)
—	—	—	76.03	—	**48.07**
√	√	—	78.81	2.78	35.72
√	—	√	79.62	3.59	33.19
—	√	√	78.73	2.70	36.28
√	√	√	**80.51**	**4.48**	31.96

4.4 问题讨论与分析

从上述实验结果以及图7的检测结果中可以看出，使用深度学习的方法来进行缺陷检测的确取得了不错的效果，但仍然存在一些问题。裂纹类缺陷存在的噪声干扰以及背景与目标之间的低对比度导致实验中所有模型在面对此类缺陷时均表现不佳。轧入氧化皮类缺陷的检测精度相较于其他类型缺陷同样偏低，究其原因是此类缺陷的位置较为分散，难以确定边界范围，模糊的缺陷边界也会导致单个缺陷被检测为两个相邻的缺陷。对于上述情况：一方面，可以先使用特定的图像预处理算法来减少噪声，提高对比度，然后再进行检测；另一方面，也可以进一步优化标签信息，更加准确地描述有关缺陷的信息。另外，针对数据集样本数量较少的问题，可以利用生成对抗网络产生更多复杂样本，从而实现对缺陷目标的精准检测。

5 结 论

提出了一种面向钢材表面缺陷检测的改进型YOLOv5算法。该模型在骨干网络的输出特征层后增加以Inception结构结合空洞卷积的RFB，有利于应对检测中缺陷之间尺度不一的问题；AFAM替换原有特征融合网络中简单的特征映射融合操作，促进了相

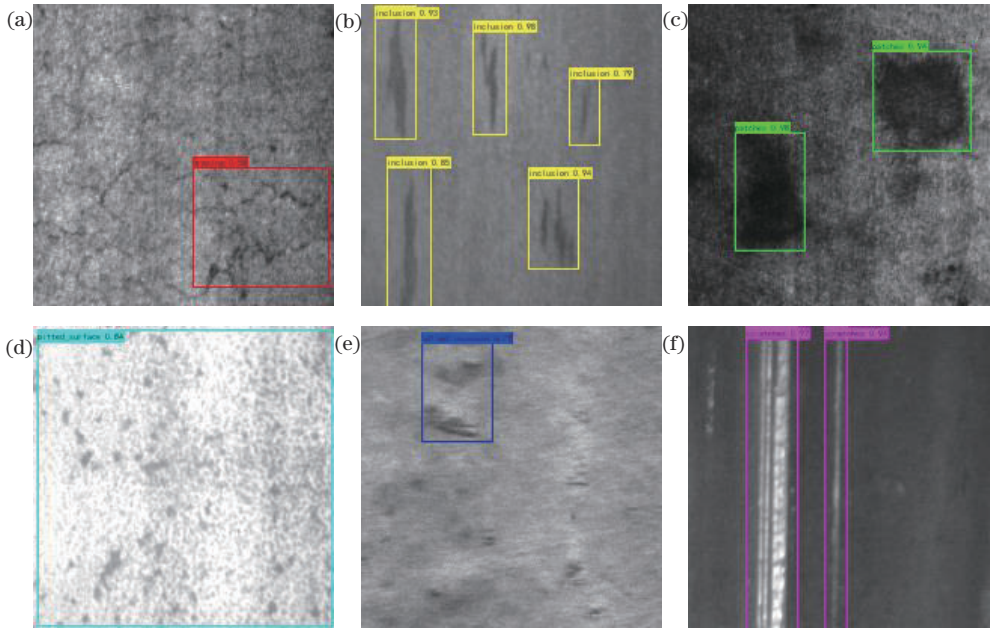

图 7　所提模型在 NEU-DET 数据集上的检测结果。(a)裂纹;(b)夹杂物;(c)斑块;(d)点蚀面;(e)轧入氧化皮;(f)划痕

Fig. 7　Detection results of the proposed model on NEU-DET dataset. (a) Crazing; (b) inclusion; (c) patches; (d) pitted surface; (e) rolled-in scale; (f) scratches

邻级别特征图间精确的信息融合;具有 ECA 机制的解耦头替换原来的耦合头,在突出强调重要特征信息的同时抑制无用的特征信息,解耦输出的结构可缓解分类与回归任务之间特征空间的差异性问题。在 NEU-DET 数据集上的实验结果表明,与其他主流算法相比,所提改进后模型在精度与速度方面均表现良好,可以更好地应用于钢材表面缺陷检测任务。

参　考　文　献

[1] 李维创, 尹柏强. 工业金属板带材表面缺陷自动视觉检测研究进展[J]. 电子测量与仪器学报, 2021, 35(6): 1-16.
Li W C, Yin B Q. Research progress of automated visual surface defect detection for industrial metal planar materials[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6): 1-16.

[2] 汤勃, 孔建益, 伍世虔. 机器视觉表面缺陷检测综述[J]. 中国图象图形学报, 2017, 22(12): 1640-1663.
Tang B, Kong J Y, Wu S Q. Review of surface defect detection based on machine vision[J]. Journal of Image and Graphics, 2017, 22(12): 1640-1663.

[3] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[4] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[5] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceeding of the 2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.

[6] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2022-11-20]. https://arxiv.org/abs/1804.02767.

[7] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2022-11-25]. https://arxiv.org/abs/2004.10934.

[8] Liu R Q, Huang M, Gao Z M, et al. MSC-DNet: an efficient detector with multi-scale context for defect detection on strip steel surface[J]. Measurement, 2023, 209: 112467.

[9] 罗晖, 李健, 贾晨. 基于图像增强与改进 Cascade R-CNN 的钢轨表面缺陷检测[J]. 激光与光电子学进展, 2021, 58(22): 2212001.
Luo H, Li J, Jia C. Rail surface defect detection based on image enhancement and improved cascade R-CNN[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2212001.

[10] Cheng X, Yu J B. RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 2503911.

[11] 孙连山, 魏婧雪, 朱登明, 等. 基于 AM-YOLOv3 模型的铝型材表面缺陷检测算法[J]. 激光与光电子学进展, 2021, 58(24): 2415007.
Sun L S, Wei J X, Zhu D M, et al. Surface defect detection algorithm of aluminum profile based on AM-YOLOv3 model[J]. Laser & Optoelectronics Progress, 2021, 58(24): 2415007.

[12] Yu X Y, Lü W T, Zhou D, et al. ES-net: efficient scale-aware network for tiny defect detection[J]. IEEE Transactions on Instrumentation and Measurement,

2022, 71: 3511314.

[13] 程松, 杨洪刚, 徐学谦, 等. 基于YOLOv5的改进轻量型X射线铝合金焊缝缺陷检测算法[J]. 中国激光, 2022, 49(21): 2104005.
Cheng S, Yang H G, Xu X Q, et al. Improved lightweight X-ray aluminum alloy weld defects detection algorithm based on YOLOv5[J]. Chinese Journal of Lasers, 2022, 49(21): 2104005.

[14] Dong H W, Song K C, He Y, et al. PGA-net: pyramid feature fusion and global context attention network for automated surface defect detection[J]. IEEE Transactions on Industrial Informatics, 2020, 16(12): 7448-7458.

[15] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 404-419.

[16] Huang Z L, Wei Y C, Wang X G, et al. AlignSeg: feature-aligned segmentation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 550-557.

[17] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[18] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2022-11-26]. https://arxiv.org/abs/2107.08430.

[19] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.

[20] 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), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[21] 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, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.

[22] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[23] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[EB/OL]. (2016-08-23)[2022-11-26]. https://arxiv.org/abs/1602.07261.

[24] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[C]//International Conference on Learning Representations, May 2-4, 2016, San Juan, Puerto Rico, USA. [S.l.: s.n.], 2016.

[25] Song G L, Liu Y, Wang X G. Revisiting the sibling head in object detector[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11560-11569.

[26] He Y, Song K C, Meng Q G, 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.

[27] 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[EB/OL]. (2022-07-06)[2023-03-10]. https://arxiv.org/abs/2207.02696.