

小样本条件下的带钢表面缺陷检测^①

宋文琦¹, 吴龙^{1,2}, 黎尧²

¹(福州大学 机械工程及自动化学院, 福州 350116)

²(三明学院 机电工程学院, 三明 365001)

通信作者: 宋文琦, E-mail: 1057929430@qq.com



摘要: 针对工业场景下带钢表面缺陷样本少、缺陷尺寸大小不一等问题, 提出一种适用于小样本条件下的带钢表面缺陷检测网络. 首先, 算法以 YOLOv5s 框架为基础, 设计一种融合注意力机制的多尺度路径聚合网络作为模型的颈部, 增强模型对缺陷目标的多尺度预测能力; 其次, 提出一种自适应解耦检测结构, 缓解小样本情况下分类和定位任务之间的矛盾; 最后, 提出一种融合 Wasserstein 距离的边界框回归损失函数, 提升模型对小目标缺陷的检测精度. 实验表明, 在构建的小样本带钢表面缺陷数据集上, 本文模型的检测性能优于其他小样本检测模型, 更适用于工业环境下的小样本缺陷检测任务.

关键词: 钢材表面缺陷检测; 小样本; 注意力机制; 多尺度路径聚合网络; 解耦检测结构

引用格式: 宋文琦, 吴龙, 黎尧. 小样本条件下的带钢表面缺陷检测. 计算机系统应用, 2024, 33(5): 85-93. <http://www.c-s-a.org.cn/1003-3254/9506.html>

Surface Defect Detection of Strip Steel with Few Shots

SONG Wen-Qi¹, WU Long^{1,2}, LI Yao²

¹(School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350116, China)

²(School of Mechanical and Electrical Engineering, Sanming University, Sanming 365001, China)

Abstract: To address the problems of few shots and varying sizes in the surface defects on steel strips in industrial scenarios, this study proposes a detection network for surface defects on steel strips readily applicable to few-shot situations. Specifically, the algorithm is based on the you only look once version 5 small (YOLOv5s) framework and a multi-scale path aggregation network with an attention mechanism is designed to serve as the neck of the model and thereby enhance the ability of the model to predict the defect objects on multiple scales. Then, an adaptive coord-decoupled head is proposed to alleviate the contradiction among classification and positioning tasks in few-shot scenarios. Finally, a bounding box regression loss function fused with the Wasserstein distance is presented to improve the accuracy of the model in detecting small defect objects. Experiments show that the proposed model outperforms other few-shot object detection models on the few-shot dataset of surface defects on steel strips, indicating that it is more suitable for few-shot defect detection tasks in industrial environments.

Key words: surface defect detection of steel; few shots; attention mechanism; multi-scale path aggregation network; coord-decoupled head

热轧钢作为一种重要工业材料, 在机械制造、航空工业等领域被广泛应用. 在热轧过程中, 由于生产工艺和生产设备限制造成的钢材表面缺陷, 如裂纹、

辊印、氧化铁皮等, 对于钢材的耐磨性、抗腐蚀性和疲劳强度等都会产生负面影响^[1]. 目前仍有不少企业采用人工缺陷检测方式, 这种检测方式成本高、效率

① 基金项目: 福建省科技重大专项 (2022HZ026025); 福建省自然科学基金 (2022J011182)

收稿时间: 2023-12-03; 修改时间: 2023-12-29; 采用时间: 2024-01-12; csa 在线出版时间: 2024-03-22

CNKI 网络首发时间: 2024-03-27

低^[2]。随着深度学习算法的发展,基于卷积神经网络的缺陷检测方法逐渐替代了人工加内存和传统的机器视觉检测方法。为增强网络表征能力,提高特征重用,Kou等人^[3]引入密集卷积块提取特征信息,提出一种基于YOLOv3的钢带缺陷检测模型。余松森等人^[4]针对现有模型对瓷砖表面微小缺陷识别能力不足的问题,提出一种轻量型瓷砖表面缺陷检测模型,极大改善原有模型对小目标缺陷的错检、漏检问题。上述研究证明了基于深度学习的模型在缺陷检测任务上的优越性。然而深度学习模型的良好性能往往依赖大量带标签的训练数据。而在工业领域,产品制造出现缺陷的比率很低,数据的稀缺性导致模型泛化性弱,容易过拟合。

为缓解样本稀缺问题,Kang等人^[5]第1个开发出目标检测的小样本学习模型(FSRW),在YOLOv2模型框架中引入元特征学习器以实现小样本学习下的端到端的训练模型。Wang等人^[6]认为只在现有模型基础上附加元学习器,模型记忆效率会随支持集类别数量增加而降低,因此基于双阶段检测模型提出广义的小样本目标检测(G-FSOD)微调方式,在提高新类检测精度的同时能保持基类数据的检测精度不下降。然而,这些通用型算法无法很好地应对工业场景中不同形状和尺寸的缺陷。针对这一挑战,于重重等人^[7]引入动态卷积,提出一种可变形的特征金字塔网络,以更好地提取形态各异的缺陷特征。陈朝等人^[8]基于一个两阶段检测网络,通过一个缺陷突显模块来更好地利用无缺陷样本增强缺陷区域特征,获得良好的检测性能。Wang等人^[9]基于Faster R-CNN提出一个小样本模型训练框架,提升了模型检测的鲁棒性。上述几种方法针对工业环

境下的缺陷检测研究表明小样本学习技术同样适用于工业场景中,但它们都属于两阶段检测模型,虽然在检测精度上表现良好,在检测速率上却难以满足实际工业自动化生产的需求。同时,由于带钢表面缺陷存在与背景对比度低、目标尺寸跨度大、小目标较多等问题,使得缺陷检测更为困难。为克服上述挑战,本文采用小样本学习技术,并基于YOLOv5s单阶段检测模型构建一个小样本钢材表面缺陷检测网络FODNet。本文主要工作如下:(1)设计一种融合注意力机制的多尺度路径聚合网络作为模型的颈部,以增强模型对缺陷目标的多尺度预测能力和特征提取能力。(2)提出一种自适应解耦检测结构(coord-decoupled head, Coord-DH),以更有效地缓解小样本条件下模型分类任务和定位任务之间的矛盾。(3)提出一种融合Wasserstein距离的边界框回归损失函数(Wasserstein distance-IoU, *WD-IoU*),以进一步提升模型对小目标缺陷的检测精度。

1 YOLOv5 模型结构

如图1所示,YOLOv5模型可分为主干网络Backbone、特征融合网络Neck和检测网络Head这3部分。其中,Backbone采用CSPDarknet53提取特征,利用CSP结构可以有效减少网络参数和计算量,提高特征提取效率;空间金字塔池化结构SPPF则可以将不同尺度的特征图固定在一个尺度进行特征融合。Neck采用路径聚合网络PANet(path aggregation network, PANet)^[10]将不同层次的特征图通过自顶向下和自底向上的方式融合,以获得更丰富的特征信息。最后,通过Head输出3种不同尺度的特征图,并在特征图上预测各个目标的边界框位置、大小和类别。

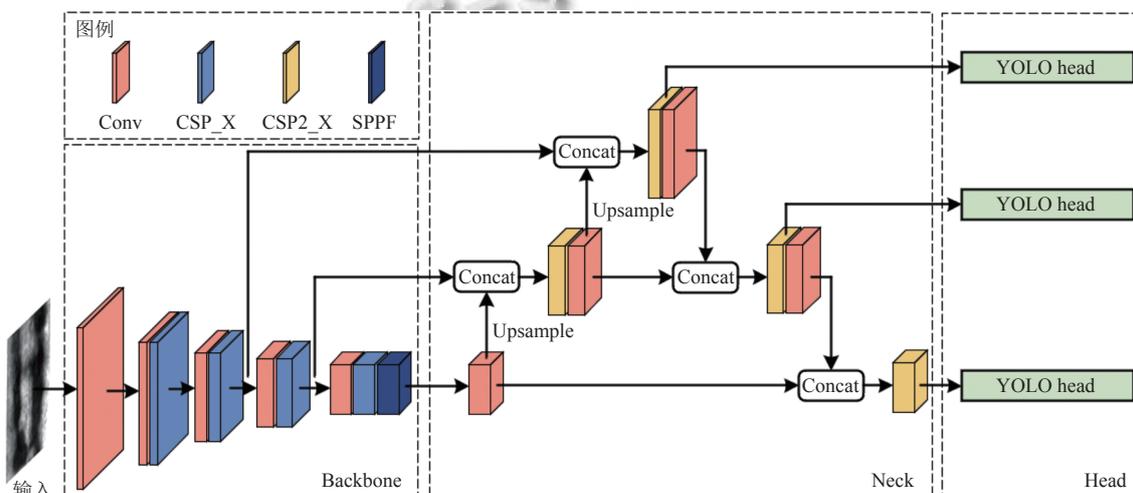


图1 YOLOv5 模型结构图

2 小样本带钢表面缺陷检测模型

本文提出的 FODNet 模型结构如图 2 所示. 与 YOLOv5 不同, FODNet 的 Neck 部分采用一种融合注意力机制的多尺度路径聚合网络, Head 部分采用一种自适应解耦检测结构 Coord-DH. 首先, 主干网络对输入图像进行特征提取, 得到 20×20 、 40×40 、 80×80 这 3 种不同尺度大小的特征图, 再一并传入特征融合部分. 特征融合结构将主干网络提取到的深层特征通过 CARAFE 模块进行上采样, 减少细节特征丢失; 接着与浅层特征拼接, 实现多尺度特征融合; 再将得到的特征图通过 C3SA 模块进行全局特征提取, 获得更加丰富

的特征信息. 最后, 检测器预测大、中、小型目标, 并通过损失函数计算得到检测结果.

2.1 融合注意力机制的多尺度路径聚合网络

由于带钢表面缺陷与背景对比度低, 且缺陷尺寸大小不一, 因此对缺陷进行有效提取是关键. 为满足需求, 本文在 PANet 多尺度特征表达结构基础上, 融入 Shuffle Attention 双路注意力机制和内容感知特征重组的上采样方式 (content-aware reassembly of features, CARAFE) 以增强网络的特征提取能力, 从而让模型能够更好地区分缺陷和背景, 进一步提升对不同尺度目标的识别能力.

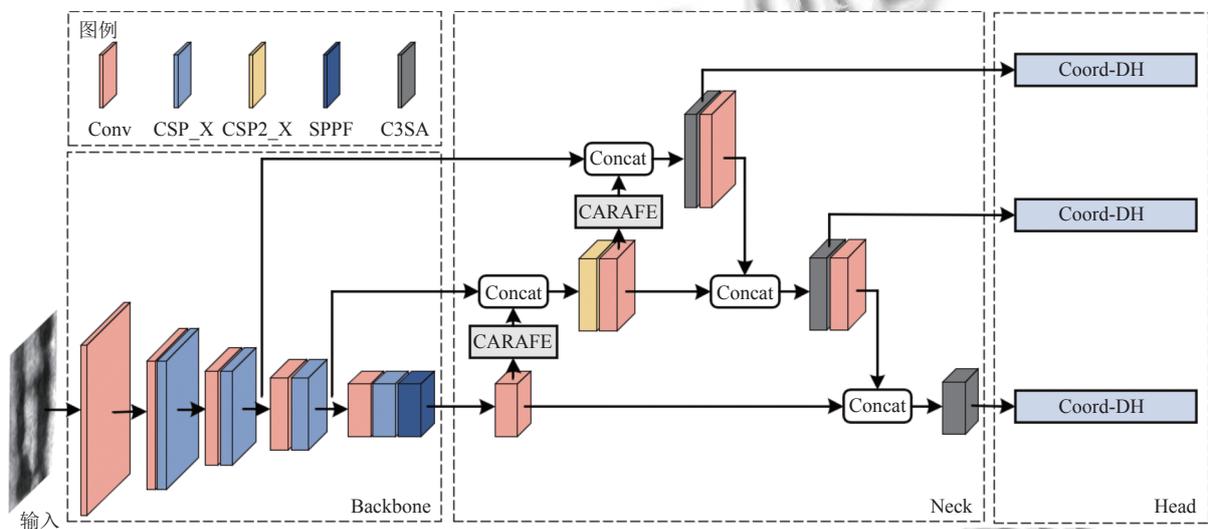


图 2 FODNet 模型结构图

2.1.1 融合 Shuffle Attention 双路注意力机制

注意力机制作用于特征图的不同通道和空间位置来抑制不太显著的特征, 帮助网络在样本数据稀疏的情况下更高效地找到应该关注的重点. 为减少算力消耗, 本文引入轻量且高效的 Shuffle Attention 机制^[11]. 该双路注意力机制可充分利用空间和通道注意力的关系, 使模型损失值更快收敛.

图 3 中, 本文先将 Shuffle Attention 与 BottleNeck2 模块结合为 SA_BottleNeck 模块, 再整合到 CSP2 模块中形成 C3SA 模块, 应用于每个检测器的输入之前. 经过特征融合操作的特征图在传入检测器前, 都会经过 C3SA 模块进行全局特征提取. 作为 C3SA 主要的梯度流分支, 融入注意力机制的 SA_BottleNeck 模块在减少计算量的同时可以获得比原先更加丰富的梯度流信息, 既保证特征图中各尺度特征信息的丰富性, 也抑制冗余信息对检测器的干扰.

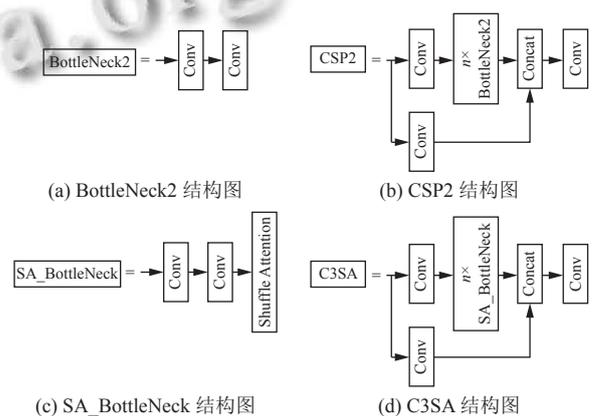


图 3 C3SA 结构示意图

如图 4 所示, Shuffle Attention 先将特征图 F 沿其通道维度划分为 g 个组, 即 $F = [F_1, \dots, F_g]$, $F_k \in R^{c/g \times h \times w}$. 随后 F_k 沿通道维度被分成 F_{k1} 和 F_{k2} , 分别传入通道注意力分支和空间注意力分支. 如式 (1), 在通道注意力

分支中, F_{k1} 经过全局平均池化操作得到通道统计量 s . 接着按照式 (2), 将 s 经过权重卷积和 Sigmoid 函数操作得到通道注意力权重矩阵, 再与 F_{k1} 相乘得到 F'_{k1} . 其中 W_1 和 b_1 是通道注意力中的权重矩阵和权重向量.

$$s = F_{gp}(F_{k1}) = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w F_{k1}(i, j) \quad (1)$$

$$F'_{k1} = \sigma(F_c(s)) \cdot F_{k1} = \sigma(W_1 s + b_1) \cdot F_{k1} \quad (2)$$

在空间注意力分支中, 通过对 F_{k2} 使用 GroupNorm (GN)^[12] 获取空间维度信息, 再经过式 (3) 对不同空间位置进行加权得到特征 F'_{k2} .

$$F'_{k2} = \sigma(W_2 \cdot GN(F_{k2}) + b_2) \cdot F_{k2} \quad (3)$$

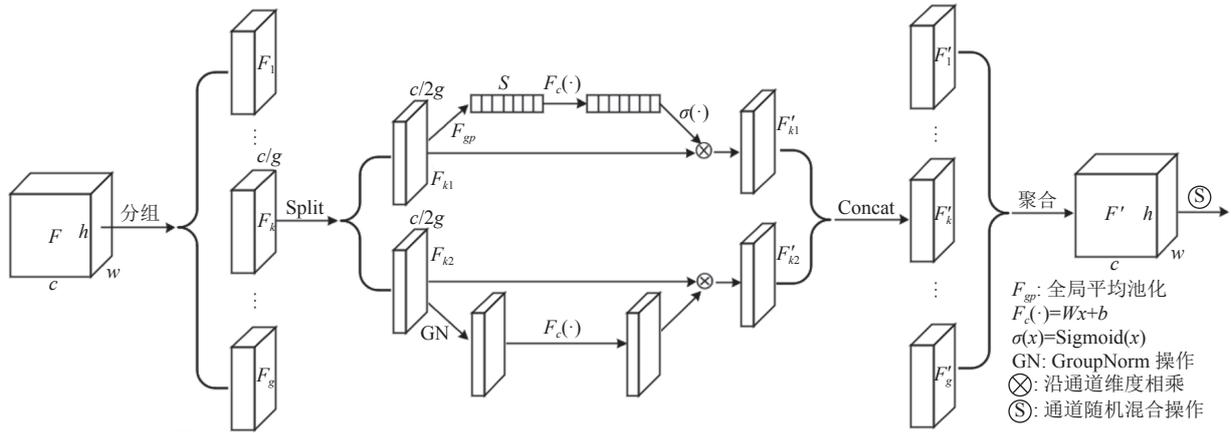


图4 Shuffle Attention 模块

两个分支结果按通道拼接后, 特征聚合成 F' , 再通过通道随机混合实现跨组信息交流.

2.1.2 CARAFE 模块

为解决 YOLOv5 中最邻近插值上采样方式易出现棋盘现象, 丢失部分图像信息的问题^[13], 本文引入 CARAFE 模块^[14]. 该模块能在更大的感知域中聚合上下文信息, 并在不同位置使用自适应和优化的重组内核, 得到更细腻的高分辨率特征图, 以减少细节特征丢失对多尺度特征融合过程的影响.

CARAFE 包含内核预测模块和特征重组模块, 两个模块的处理过程分别如式 (4) 和式 (5) 所示. 内核预测模块利用一个 $k_{\text{encoder}} \times k_{\text{encoder}}$ 大小的卷积核在输入特征图 X 的每个位置 l 都预测一个上采样核 W_l , 即 $N(X_l, k_{\text{encoder}})$ 操作. 然后特征重组模块将输出特征图 X' 中的每个位置 l' 都映射回特征图 X , 假设上采样核的尺寸为 $k_{\text{up}} \times k_{\text{up}}$, 则从 X 中取出以位置 l 为中心的 $k_{\text{up}} \times k_{\text{up}}$ 的区域, 即 $N(X_l, k_{\text{up}})$ 操作, 再与该点预测出的上采样核 W_l 做点积, 即 $\Phi(N(X_l, k_{\text{up}}), W_l)$ 操作, 最后得到输出 X' .

$$W_l = \Psi(N(X_l, k_{\text{encoder}})) \quad (4)$$

$$X'_{l'} = \Phi(N(X_l, k_{\text{up}}), W_l) \quad (5)$$

其中, Ψ 为内核预测模块, Φ 为特征重组模块. l 为原目标位置, l' 为上采样后的目标位置. $N(X_l, k)$ 表示以位置 l 为中心的 X 的 $k \times k$ 子区域.

2.2 自适应解耦检测结构 Coord-DH

特征融合后的特征图包含更丰富的细节特征和语义特征, 检测头基于这些特征信息进行目标分类和定位. 分类任务依据语义特征学习平移不变性, 定位任务依据细节特征学习平移可变性, 即分类任务希望目标的分类结果不受其所在位置的影响, 而定位任务希望检测网络对目标的位置信息有良好的感知能力, 两个任务相互冲突, 而注释样本稀缺加剧了这一点^[15]. 因此, 本文提出自适应解耦检测结构 Coord-DH 对目标进行预测. 不同于耦合检测结构将多种信息融合在一个特征图上, 本文的检测结构将分类和定位任务解耦为两个单独的分支, 并自适应地学习两个分支不同的任务需求, 缓解两个任务的矛盾, 提高检测精度.

如图 5 所示, 输入特征图首先经过 1×1 卷积进行降维, 而后分为两路分支. 分类分支经过坐标卷积保持平移不变性, 再经过 1×1 卷积进行分类操作; 回归分支经过坐标卷积学习平移可变性, 再分解成两个平行分支分别进行定位和置信度检测.

在坐标卷积的作用下, 自适应解耦结构可根据任

务的不同,学习不同程度的平移不变性和平移可变性.如图6所示,传统卷积因具备平移不变性,适用于分类任务;但在处理定位任务时,传统卷积因无法感知位置信息而表现不佳.坐标卷积在特征图 F 后新增原始输入的 x 和 y 坐标通道,使得卷积过程可以感知特征图的空间信息.因此,当坐标卷积的坐标通道未学习到任何信息时,其等同于传统卷积,具备平移不变性;当坐标卷积学习到一定信息,就具备一定的平移可变性,从而实现解耦检测结构中不同分支自适应地学习不同任务的目的.

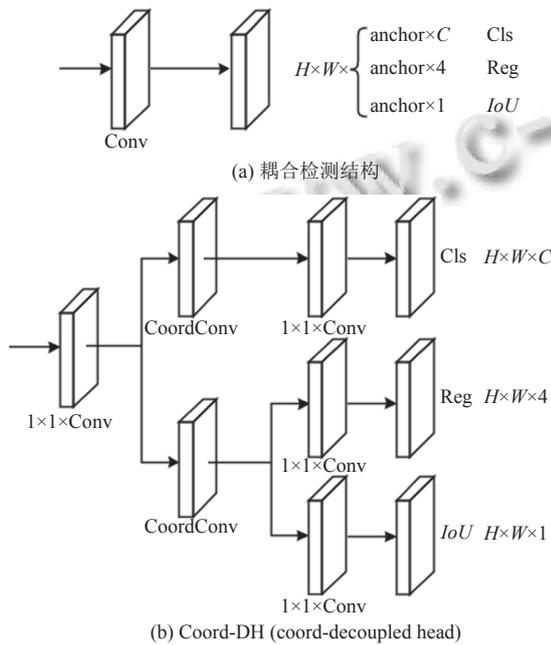


图5 耦合检测结构与 Coord-DH

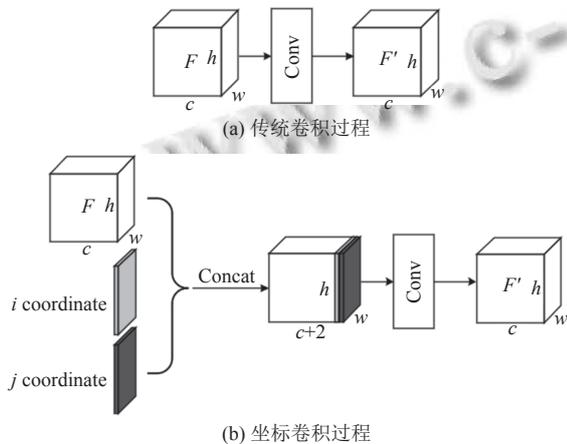


图6 传统卷积过程与坐标卷积过程

2.3 边界框回归损失函数 $WD-IoU$

表1是本文实验使用的 NEU-DET 数据集^[16]的目的

标面积大小统计表.由表1可知,缺陷面积占比 $\leq 5\%$ 的目标缺陷数约占缺陷总数的24.2%,缺陷面积占比 $\leq 10\%$ 的约占缺陷总数的44.8%,可见钢材数据中表面小缺陷、中小缺陷的比重不可忽视.

表1 NEU-DET 数据集的缺陷面积大小统计

缺陷面积与图片面积占比	个数
$\leq 5\%$	1014
$5\% < \text{area} \leq 10\%$	863
$> 10\%$	2312

在YOLOv5中,损失函数 $CIoU$ 采用交并比 (IoU) 计算位置边界框损失.而Wang等人^[17]指出, IoU 对小目标边界框的偏移非常敏感,对小目标不友好.因此本文针对不同尺寸大小的目标,提出一种更公平地计算目标边界框偏移的损失函数 $WD-IoU$, 以提高回归精度. $WD-IoU$ 由3部分构成.

(1) 将实际标签边框和预测框建模为二维高斯分布,并利用归一化的高斯 Wasserstein 距离公式计算两个分布间的距离,以度量它们的相似性^[17].由于小目标的前景多集中在边界框中心,背景多分布于边界框边缘,因此采用二维高斯分布表示目标边界框,使得分布权重从中心向边缘逐渐减小,更符合小目标特性.公式为:

$$W_2^2(N_{gt}, N_{pred}) = \left\| \left[\begin{matrix} cx_{gt}, cy_{gt}, \frac{w_{gt}}{2}, \frac{h_{gt}}{2} \end{matrix} \right]^T, \left[\begin{matrix} cx_{pred}, cy_{pred}, \frac{w_{pred}}{2}, \frac{h_{pred}}{2} \end{matrix} \right]^T \right\|_2^2 \quad (6)$$

并利用指数将 Wasserstein 距离归一化到 (0, 1), 得到式 (7):

$$NWD(N_{gt}, N_{pred}) = \exp \left(-\frac{\sqrt{W_2^2(N_{gt}, N_{pred})}}{C} \right) \quad (7)$$

其中, (cx_{gt}, cy_{gt}) 是实际标签边框的中心点坐标, w_{gt} 和 h_{gt} 分别表示其宽和高; (cx_{pred}, cy_{pred}) 是预测边框的中心点坐标, w_{pred} 和 h_{pred} 分别表示其宽和高. C 为常数,由数据集中目标的平均绝对尺寸决定.

(2) 实际标签边框与预测框之间的长宽比差异.良好的边界框回归损失应考虑预测框与实际标签框的重叠面积、中心点距离和长宽比^[18].上述第(1)部分中的 NWD 相当于同时处理重叠面积和中心点距离问题,因此再引入 $EIoU$ ^[19]中的边界框长宽比公式:

$$asp = \frac{\rho^2(w_{pred}, w_{gt})}{C_w^2} + \frac{\rho^2(h_{pred}, h_{gt})}{C_h^2} \quad (8)$$

其中, w_{gt} 和 h_{gt} 是实际标签边框的宽和高, C_w 和 C_h 是预测框和实际标签边框最小外接矩形的宽和高, ρ 为欧氏距离.

(3) 实际标签边框与预测框的重叠面积及中心点距离. 考虑到大目标不一定服从二维高斯分布规律, 使用 IoU 来衡量边界框之间的贴合度更适合评估大目标边界框的偏移程度. 由表 1 可知, 数据集中大目标的比例较高, 因此两个边界框的重叠面积和中心点距离要素仍需考虑, 公式为:

$$IoU_{dis} = IoU - \frac{\rho^2(A_{pred}, A_{gt})}{c^2} \quad (9)$$

$$IoU = \frac{|A_{pred} \cap A_{gt}|}{|A_{pred} \cup A_{gt}|} \quad (10)$$

其中, IoU 为预测框与实际标签边框的交并比, A_{pred} 是预测框面积, A_{gt} 是实际标签边框面积, ρ 为欧氏距离, c 为两个边界框的最小外接矩形对角线长度.

最后, $WD-IoU$ 计算公式及 $WD-IoU$ 损失计算公式如下:

$$WD-IoU = \lambda_1 \cdot NWD(N_{gt}, N_{pred}) + \lambda_2 \cdot IoU_{dis} - asp \quad (11)$$

$$Loss_{WD-IoU} = 1 - WD-IoU \quad (12)$$

其中, λ_1 和 λ_2 是与数据集相关的比例系数, 取 $\lambda_1, \lambda_2 \in [0, 1]$, 且限制 $WD-IoU$ 的值在 $[0, 1]$ 范围内. λ_1 与 λ_2 的具体设置将在第 3.3 节中进一步说明.

3 实验设计与结果分析

为验证本文方法的有效性, 使用东北大学提供的公开数据集 NEU-DET 对模型的检测性能进行评估. 实验在操作系统 Ubuntu 20.04 上进行, 硬件配置为 AMD Ryzen 5 5600 6-Core Processor, NVIDIA GeForce RTX 3060; 环境配置为 Python 3.7 + PyTorch 1.10.1 + Cuda 11.2. 模型网络参数配置如下: 输入图片尺寸为 640×640 , 采用 SGD 优化器, 动量因子为 0.937, 初始学习率为 0.01, 权重衰减因子为 0.0005, batch-size 为 16.

3.1 实验数据集

NEU-DET 数据集包含 6 类典型缺陷, 如图 7 所示, 每类缺陷包含 300 张 200×200 分辨率的图片. 现有的小样本钢材表面缺陷检测研究^[9]将 NEU-DET 数据

集划分为基类 C_{base} (包含夹杂、压入氧化铁皮和划痕) 和新类 C_{novel} (包含裂纹、斑块和麻点). 基类数据集 D_{base} 包括训练集 D_{base}^t 720 张和验证集 D_{base}^v 180 张, 用于模型预训练和验证; 新类数据集 D_{novel} 包括训练集 D_{novel}^t 720 张和验证集 D_{novel}^v 180 张, 微调阶段从 D_{novel}^t 中随机抽取每种类别 k 张图像进行训练, 并在 D_{novel}^v 上进行验证.

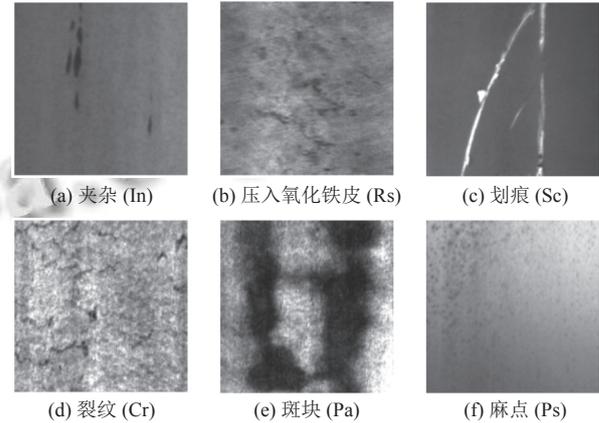


图 7 NEU-DET 各类缺陷示例

与上述方法不同, 本文在获取预训练模型后, 使用 $D_{base+novel}^t$ 进行微调, 即从 D_{base}^t 和 D_{novel}^t 中随机采样每种类别 k 个样本, 并在 $D_{base+novel}^v$ 上进行验证, 其中 k 值可取 5, 10, 30. 图 8 是本文的小样本带钢表面缺陷数据集构建方式.

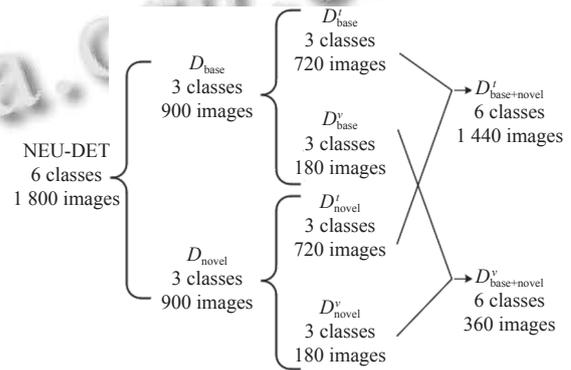


图 8 NEU-DET 数据集的基类新类划分图

模型训练流程为: (1) 基训练阶段采用 D_{base}^t 进行训练, D_{base}^v 进行验证, 获取预训练模型; (2) 随机初始化检测器中的分类器的类别个数为微调阶段的类别个数; (3) 冻结预训练模型的主干网络, 并在 $D_{base+novel}^t$ 上进行微调, 在 $D_{base+novel}^v$ 上进行验证. 为增强模型的泛化能力, 在基训练阶段和微调阶段每次训练迭代过程中, 输

入图片有 80% 的概率进行随机尺度变化,以增强模型的尺度不变性;有 50% 的概率进行旋转,以增强模型

的旋转不变性;有 100% 的概率进行 Mosaic 数据增强,以丰富训练数据.图 9 是模型微调阶段训练流程图.

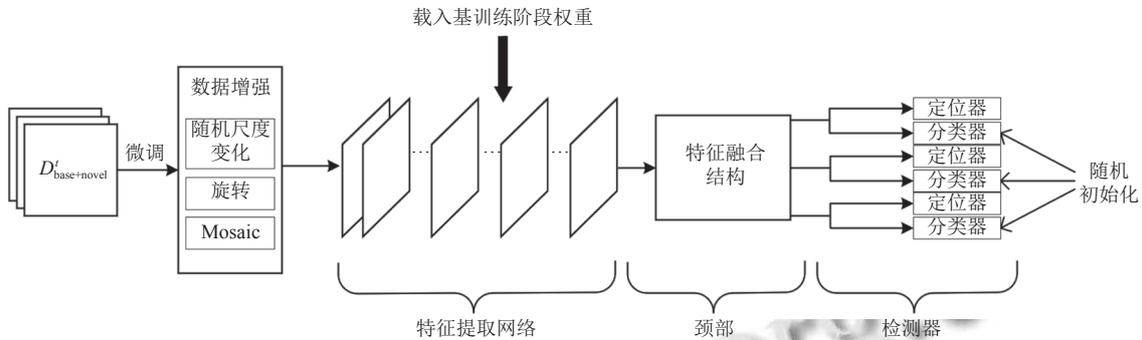


图 9 微调细节图

3.2 对比实验

为评估本文模型的检测性能,使用 YOLOv5s 模型框架作为基线,记作 YOLO+ft;此外,还将本文模型分别与自然场景下的小样本研究典型 TFA^[6]和工业场景下的小样本研究典型 Faster R-CNN (IFDD)^[9]进行比较,在 NEU-DET 小样本数据集上进行实验.其中,*表示模型在微调阶段每种类别随机采样 k 张图片而非 k 个样本,且仅在新类样本上进行训练,而非使用同等数量的基类和新类样本一起训练^[9],其他训练设定与本文相同.实验结果见表 2.

表 2 在 NEU-DET 新类数据集上的性能评估
(mAP=AP₅₀)(%)

模型	5-shot	10-shot	30-shot
YOLO+ft*	55.6	61.2	64.8
TFA ^[6] *	37.1	39.4	41.1
Faster R-CNN (IFDD) ^[9] *	52.64	58.36	64.29
FODNet (ours)*	62.3	64.3	69.5
YOLO+ft	53.9	59.6	66.5
TFA ^[6]	36.7	40.1	45.4
FODNet (ours)	60.8	65.1	71.0

从表 2 可以看出,无论是采用*训练方式还是非*训练方式,本文模型在各种条件下(5-shot, 10-shot, 30-shot)都表现出最佳检测性能.具体而言,在*训练方式下,相比基线 YOLO+ft*,FODNet*的 mAP 在 3 种 shot 情况下分别增长了 6.7%、3.1% 和 4.7%;与 Faster R-CNN (IFDD)*相比,FODNet*的 mAP 分别提升了 9.66%、5.96% 和 5.21%;与 TFA*相比,FODNet*的 mAP 分别提升了 25.2%、24.9% 和 28.4%.类似地,在非*训练方式下,相比 YOLO+ft,FODNet 的 mAP 分别提升了 6.9%、5.5% 和 4.5%;相比 TFA,FODNet 的

mAP 分别提升了 24.1%、25.0% 和 25.6%.实验结果充分证明了 FODNet 在工业环境下的卓越检测性能.不过,TFA 和 TFA*的 mAP 远低于其他模型,这说明通用型小样本目标检测模型在处理尺度多变、形态各异的缺陷检测任务时并不适用,不能直接应用于工业场景中.

此外,FODNet 从 5-shot 到 30-shot 的 mAP 增长速度(从 60.8% 到 71.0%)明显快于 FODNet* (从 62.3% 到 69.5%),这种情况也表现在 YOLO+ft 和 TFA 上.这表明本文采用的模型训练方式更适用于小样本模型,更有助于提升模型的检测性能.

图 10 展示了在本文数据集划分条件下,YOLO+ft、TFA 和 FODNet 对 NEU-DET 新类的测试结果.图中 FODNet 表现最佳,而其他算法的检测效果相对不理想.YOLO+ft 将裂纹(Cr)样本的上半部分的缺陷识别为多个缺陷,而改进后的 FODNet 成功识别为一个缺陷,显示了 FODNet 全局信息提取能力的优势.同时,YOLO+ft 将与麻点(Ps)缺陷相似的伪缺陷误识别为真实缺陷,而 FODNet 未受伪缺陷干扰,准确识别出真实缺陷位置.此外,TFA 在斑块(Pa)样本检测中遗漏部分缺陷,在裂纹(Cr)和麻点(Ps)样本检测中未能发现任何缺陷,这在工业应用中是不允许发生的.因此,FODNet 拥有更稳定的检测性能,更适用于工业场景下的带钢表面缺陷检测任务.

3.3 边界框回归损失函数实验分析

λ_1 和 λ_2 是与训练样本的尺寸分布比例相关的系数.当样本中小目标占比较多时,就增大 λ_1 与 λ_2 的比值以提高高斯 Wasserstein 距离公式的使用概率;反之

则减小 λ_1 与 λ_2 的比值以提高边界框交并比的使用概率. 考虑到微调阶段需要多次随机采样样本进行训练, 即需要设定多组 λ_1 和 λ_2 , 因此设置 $\lambda_1 + \lambda_2 = 1$ 来简化取值情况的讨论. 设 n 为缺陷面积与图像总面积的比值, 缺陷总目标数为 N , 面积占比小于 n 的目标数量为 N_s , 取 $\lambda_1/\lambda_2 = N_s/(N - N_s)$, 记损失值为 $WD-IoU_n$. 表 3 在新类设置 5-shot 条件下对比了 $CIoU$ 、 NWD 以及本文提出的 $WD-IoU$ 这 3 种边界框回归损失函数对模型检测性能的影响. 其中, 根据表 1 对 NEU-DET 数据集的样本数量统计, 进一步对 $WD-IoU_n$ ($n=0.05, 0.1, 0.2$) 分为 3 种情况讨论.

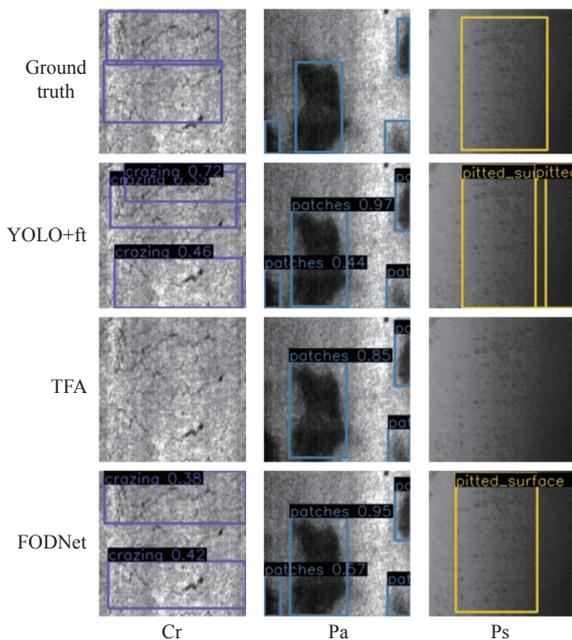


图 10 小样本缺陷检测模型的检测结果

表 3 5-shot 条件下不同损失函数在 NEU-DET 新类数据集上的检测结果 (%)

损失函数	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
$CIoU$	26.7	53.9	23.1	10.6	20.5	36.0
NWD	27.4	55.1	25.3	14.2	21.3	34.9
$WD-IoU_{0.05}$	27.6	55.4	25.4	13.7	20.8	36.2
$WD-IoU_{0.1}$	28.5	56.2	26.0	15.9	21.1	36.4
$WD-IoU_{0.2}$	27.9	55.7	25.7	15.2	20.6	35.7

由表 3 可知, 相较于原算法的 $CIoU$ 损失函数, 采用 NWD 损失函数可在 AP_{50} 上提升 1.2%, 且对小型和中型目标有更好的检测性能, 在 AP_S 和 AP_M 上分别提升 3.6% 和 0.8%; 而 $CIoU$ 损失函数对大目标有更好的检测效果, 其 AP_L 值比 NWD 损失函数高出 1.1%.

此外, 根据表 3 数据, 采用 $WD-IoU_{0.1}$ 损失函数得

到的检测效果最优, 即选取 λ_1 与 λ_2 的比值为面积占比 $\leq 10\%$ 的目标数量与剩余目标数量之比, AP_{50} 达到 56.2%, 相比原算法提升了 2.3%; 同时, 小型和大型目标的检测精度也达到最佳水平, AP_S 和 AP_L 分别为 15.9% 和 36.4%; 并且 $WD-IoU_{0.1}$ 损失函数得到的 AP_S 和 AP_L 之间的差距也最小, 为 20.5%. 实验结果表明, $WD-IoU$ 损失函数能够精准回归目标边界框, 提升模型对不同尺寸大小的目标的检测性能.

3.4 消融实验

为评估本文改进的部分对模型性能的影响, 针对新类样本的不同 shot 条件进行消融实验. non-SA & CARAFE 表示模型采用 YOLOv5 的 PANet 结构进行特征融合, 而不使用本文提出的特征融合结构, 其他设置与 FODNet 一致; non-DH 表示采用 YOLOv5 的检测头进行检测, 而不使用 Coord-DH 检测结构, 其他设置与 FODNet 一致; non-WD 表示模型采用 YOLOv5 默认的边界框回归损失函数进行损失计算, 而不使用 $WD-IoU$ 损失函数, 其他设置与 FODNet 一致. 消融实验结果如表 4 所示.

表 4 NEU-DET 新类数据集进行消融实验的结果 (%)

方法	5-shot	10-shot	30-shot
non-SA & CARAFE	57.6	63.0	68.7
non-DH	58.4	61.8	66.5
non-WD	60.4	64.5	70.3
FODNet (ours)	60.8	65.1	71.0

从表 4 结果可以看出, 如果在特征融合结构中不使用 C3SA 和 CARAFE 模块, 模型在 5-shot、10-shot 和 30-shot 条件下的检测精度将分别下降 3.2%、2.1% 和 2.3%, 这表明上述两个模块对于增强网络的特征提取能力有重要作用, 能够有效地学习关键特征. 如果在检测结构中不使用 Coord-DH, 模型在 3 种 shot 条件下的检测精度将分别下降 2.4%、3.3% 和 4.5%, 这表明 Coord-DH 检测结构可以有效缓解分类任务与定位任务之间的矛盾, 有助于提升模型的检测性能. 此外, 若模型不采用 $WD-IoU$ 损失函数进行预测框回归优化, 模型的检测精度在 3 种 shot 条件下会分别下降 0.4%、0.6% 和 0.7%, 可见 $WD-IoU$ 的使用对模型识别小目标也有一定帮助.

4 结论与展望

为应对工业场景中缺乏足够缺陷样本的挑战, 本

文提出一种适用于小样本条件下的带钢表面缺陷检测模型 FODNet。具体地,通过设计融合注意力机制的多尺度路径聚合网络,使模型专注于缺陷特征,增强模型对缺陷目标的多尺度预测能力;通过提出自适应解耦检测结构 Coord-DH,将模型检测部分解耦为可自适应学习对应任务的分类器和定位器,缓解了分类和定位任务的矛盾;通过提出融合 Wasserstein 距离的边界框回归损失函数 $WD-IoU$,提升了模型对小目标缺陷的检测精度。在 NEU-DET 新类数据集上的对比实验和消融实验结果验证了 FODNet 的先进性和模型各个模块的有效性。下一步将会采集工作现场的缺陷图像,建立一个自己的缺陷数据集进行训练,并将本文研究的小样本模型部署到工厂中进行运作。

参考文献

- 郭联金, 罗炳军. PNN 与 BP 神经网络在钢板表面缺陷分类中的应用研究. 机电工程, 2015, 32(3): 352–357. [doi: 10.3969/j.issn.1001-4551.2015.03.011]
- 刘磊, 王冲, 赵树旺, 等. 基于机器视觉的太阳能电池片缺陷检测技术的研究. 电子测量与仪器学报, 2018, 32(10): 47–52. [doi: 10.13382/j.jemi.2018.10.007]
- Kou XP, Liu SJ, Cheng KQ, *et al.* Development of a YOLO-V3-based model for detecting defects on steel strip surface. Measurement, 2021, 182: 109454. [doi: 10.1016/j.measurement.2021.109454]
- 余松森, 张明威, 杨欢. 改进 YOLOv5 的瓷砖表面缺陷检测. 计算机系统应用, 2023, 32(8): 151–161. [doi: 10.15888/j.cnki.csa.009185]
- Kang BY, Liu Z, Wang X, *et al.* Few-shot object detection via feature reweighting. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 8419–8428.
- Wang X, Huang TE, Darrell T, *et al.* Frustratingly simple few-shot object detection. Proceedings of the 37th International Conference on Machine Learning. JMLR.org, 2020. 920.
- 于重重, 萨良兵, 马先钦, 等. 基于度量学习的小样本零部件表面缺陷检测. 仪器仪表学报, 2020, 41(7): 214–223. [doi: 10.19650/j.cnki.cjsi.J2006443]
- 陈朝, 刘志, 李恭杨, 等. 适于少样本缺陷检测的两阶段缺陷增强网络. 计算机工程与应用, 2022, 58(20): 108–116. [doi: 10.3778/j.issn.1002-8331.2111-0470]
- Wang HH, Li ZL, Wang HQ. Few-shot steel surface defect detection. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 5003912. [doi: 10.1109/tim.2021.3128208]
- Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.
- Zhang QL, Yang YB. SA-Net: Shuffle attention for deep convolutional neural networks. Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto: IEEE, 2021. 2235–2239.
- Wu YX, He KM. Group normalization. Proceeding of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19. [doi: 10.1007/978-3-030-01261-8_1]
- 林金朝, 蒋媚秋, 庞宇, 等. 基于改进 U-Net 网络的内窥镜图像烟雾净化算法. 中国生物医学工程学报, 2021, 40(3): 291–300. [doi: 10.3969/j.issn.0258-8021.2021.03.05]
- Wang JQ, Chen K, Xu R, *et al.* CARAFE: Content-aware reassembly of features. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 3007–3016.
- Qiao LM, Zhao YX, Li ZY, *et al.* DeFRCN: Decoupled faster R-CNN for few-shot object detection. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 8661–8670.
- He Y, Song KC, Meng QG, *et al.* An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493–1504. [doi: 10.1109/tim.2019.2915404]
- Wang JW, Xu C, Yang W, *et al.* A normalized Gaussian Wasserstein distance for tiny object detection. arXiv: 2110.13389, 2022.
- Zheng ZH, Wang P, Liu W, *et al.* Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 12993–13000.
- Zhang YF, Ren WQ, Zhang Z, *et al.* Focal and efficient IoU loss for accurate bounding box regression. Neurocomputing, 2022, 506: 146–157. [doi: 10.1016/j.neucom.2022.07.042]

(校对责编: 孙君艳)