

基于 YOLOX- α SMV 的带钢材料表面缺陷检测算法

曹义亲¹, 刘文才¹, 徐露²

(1. 华东交通大学软件学院, 江西 南昌 330013; 2. 江西交通职业技术学院机电工程学院, 江西 南昌 330013)

摘要: 针对 YOLOX 算法在钢材表面缺陷检测中特征提取不充分、多目标缺陷检测能力较弱等问题, 提出改进损失函数的多维度特征融合带钢材料表面缺陷检测算法。首先, 在 Backbone 部分应用 SPP_SF 保留多尺度特征信息, 提高分类精度。其次, 在 Neck 部分加入多维度特征融合模块 MDFFM, 将通道、空间、位置信息融入特征向量中, 加强算法的特征提取能力。最后, 引入 Varifocal Loss 和 α -CIoU 加权正负样本, 提高预测框的回归精度。实验结果表明, YOLOX- α SMV 在 NEU-DET 数据集中的 mAP@0.5:0.95 达到了 47.54%, 较 YOLOX 算法提高了 3.43%。在保持检测速度基本不变的情况下, 对模糊缺陷和小目标缺陷的识别、定位能力明显提升。

关键词: YOLOX; 缺陷检测; α -CIoU; 坐标注意力; Varifocal Loss; SoftPool

中图分类号: TP391

文献标志码: A

YOLOX- α SMV algorithm for surface defect detection of strip steel material

Cao Yiqin¹, Liu Wencai¹, Xu Lu²

(1. School of Software, East China Jiaotong University, Nanchang 330013, China; 2. School of Mechanical and Electrical Engineering, Jiangxi V&T College of Communications, Nanchang 330013, China)

Abstract: In order to solve the problems of insufficient feature extraction and weak ability of multi-target defect detection of YOLOX algorithm in steel surface defect detection, a multi-dimensional feature fusion strip material surface defect detection algorithm based on improved loss function is proposed. First of all, apply SPP_SF to the Backbone part to retain multi-scale feature information and improve classification accuracy. Secondly, the multi-dimensional feature fusion module MDFFM is added in the Neck part to integrate the channel, space and position information into the feature vector to strengthen the feature extraction ability of the algorithm. Finally, the introduction of Varifocal Loss and α -CIoU weighted positive and negative samples to improve the regression accuracy of the prediction box. The experimental results show that YOLOX- α SMV in NEU-DET data set mAP@0.5:0.95 reaches 47.54%, which is 3.43% higher than YOLOX algorithm. While keeping the detection speed basically unchanged, the ability to identify and locate fuzzy defects and small target defects is significantly improved.

Key words: YOLOX; defect detection; α -CIoU; Coordinate Attention; Varifocal Loss; SoftPool

收稿日期: 2023-06-05

基金项目: 国家自然科学基金项目(61861016); 江西省科技支撑计划重点项目(20161BBE50081)

受生产工艺、环境、使用年限等因素的影响，带钢材料表面会出现表面点蚀、划痕、裂纹、轧制氧化皮等缺陷，导致钢材质量的下降，对工业生产及产品的可靠性和安全性将造成严重影响。因此，如何对带钢材料产品表面的缺陷进行及时有效的检测，一直是相关企业和学者的重要研究内容。

常用的带钢材料表面缺陷检测方法主要有 CCD 摄像扫描器检测、激光扫描检测、漏磁检测、人工检测等，虽然可以对带钢表面的缺陷进行检测，但存在难度大、成本高、易受外部干扰、精度低、耗时长等一系列问题，很难符合实际检测需求。

机器视觉检测技术凭借成本低、速度快、精度高等优势逐渐成为许多学者的研究热点，并取得了良好效果^[1]。深度学习作为机器学习中的一种特殊范式，应用于越来越多的领域，基于该方法的缺陷检测技术也逐渐取代传统方法，得到了系统研究。Hu B^[2]提出一种改进的两阶段网络 Faster R-CNN 对 PCB 缺陷图像进行检测，使用 shuffle 残差单元代替基本残差单元，减少了整个网络的计算量，准确率基本保持不变，参数量减少了 2 倍。这种方法进一步简化了缺陷检测操作，降低了测试成本，但使用不同缺陷数据集时精度下降，泛化能力差。Liu X 等人^[3]使用 SSD 模型结合 ResNet50 特征提取网络，将不同等级的语义信息进行融合和互补，引入通道注意力减少计算量，检测速度提高了 12.9%。文献[4]提出 E-YOLOX 算法采用边缘 Cutout 数据增强对缺陷图片进行预处理，使用新的特征提取网络 ECMNet 提取梯度流特征信息，为特征融合和缺陷预测提供了良好基础，较好地平衡了检测精度和速度。Jin X 等^[5]采用双阶段网络对概率图模型 FRGMM 生成的分割方案进行目标定位，具有更高的准确率和召回率，也更具鲁棒性，适用于弱照明、外部噪音、油污等外部因素较多的情况下。

上述深度学习算法虽然在一定程度上达到了传统缺陷检测方法无法达到的检测速度，但在精度方面仍有所欠缺。Urbanas A 等^[6]使用迁移学习和数据增强的方法，采用更快的 R-CNN 网络，将特征提取网络替换为使用预训练权重的 AlexNet、VGG、GoogLeNet 和 ResNet 来检测木板的表面缺陷，检测精度达到了 96.1%。该方法的局限性在

于缺陷的特征模糊，手动标记的数据容易出现错误和误差，从而影响网络性能。李维刚等人^[7]使用改进的加权 K-means 算法聚类先验框参数，使用大尺度检测层融合浅层和深层特征，将 YOLOv3 的缺陷检测精度提升到了 80%，但 anchor 机制带来了更大参数量从而降低了算法速度。Jiaqiao Z 等^[8]基于 YOLO (You Only Live Once) 网络，使用两个密集块对图像进行优先级分类，让网络提前接收多层卷积特征，从而增强特征融合，检测精度得到进一步提高，同时缩短了检测时间。这种优先级分类依赖于大量的训练数据，对于小样本缺陷的检测很不友好。

目前的缺陷检测算法虽在检测中有良好效果，但在提高算法的检测精度的同时，降低网络参数量、提升检测速度是亟待解决的问题。MT-YOLOv5 算法^[9]的检测精度虽然达到了 82.4%，但检测速度只有 65.4 f/s，参数量也高达 29.7M，且在轧制氧化皮等多目标、小缺陷的检测精度只有大概 70%。翁玉尚等人^[10]改进的 Mask R-CNN 算法准确率达到到了 96%，检测速度却只有 5.9f/s。

针对 YOLOX 算法在钢材表面缺陷检测中特征提取不充分、多目标缺陷检测能力较弱等问题，本文以 YOLOX^[11]算法为基准模型进行改进，提出 YOLOX- α SMV 模型，旨在取得较高检测精度的同时，降低网络参数量，达到较快的检测速度，实现检测速度与检测精度之间更好的平衡。论文主要贡献为：① 基于 SoftPool，提出 SPP_SF 结构，最大程度保留池化过程中的多尺度特征信息，提高分类准确率及算法的训练效率。② 提出了多维度特征融合模块 (Multi-dimensional Feature Fusion Module, MDFFM)，将通道、空间、位置信息融入特征图中，增强算法的特征提取能力。③ 利用 Varifocal Loss 和 α -CIoU 加权不同训练样本，提高回归精度，对离群数据有一定的调整能力，增强训练稳定性。④ 在带钢表面缺陷 NEU-DET、GC10-DET 数据集和公共数据集 PASCAL VOC2012 上均取得良好的实验效果，验证了算法具有一定的泛化能力和鲁棒性。

1 YOLOX- α SMV 目标检测模型

YOLOX 是旷视科技在 2021 年推出的单阶段目标检测算法，由特征提取网络、特征融合网络和检测头三个部分组成。特征提取网络部分使用

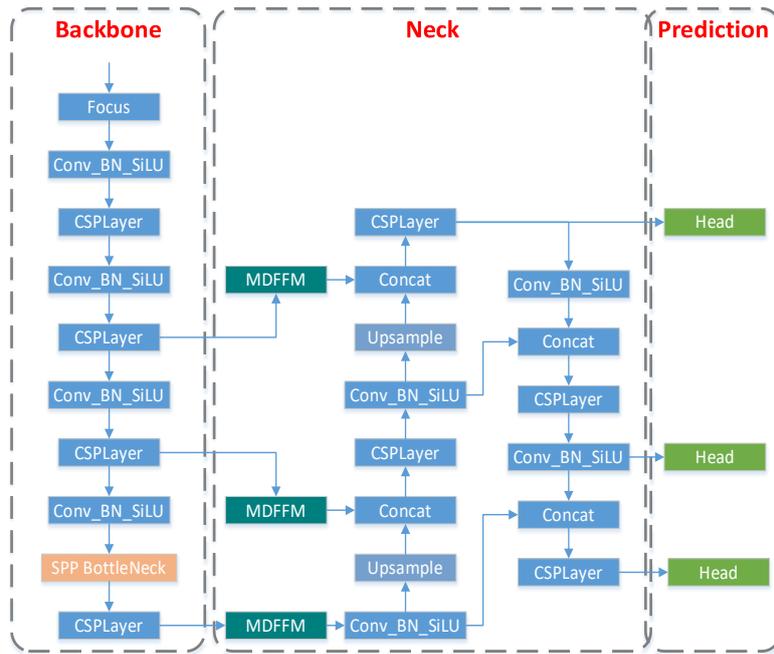


图1 模型网络结构

Fig.1 Network structure of model

CSPDarkNet 网络结构，特征融合网络使用 PAFPN，检测头使用 Decoupled head 结构。

YOLOX 算法采用了 Anchor-Free 的思想，不需要预先设定锚框大小，由网络自主学习锚框的大小，并通过损失计算，在训练过程中调整锚框中心坐标和长宽比，可以更好地适应不同目标尺寸和形状，让模型可以适用各种目标检测任务。同时，YOLOX 算法使用了解耦的检测头，即使用 3 个不同的 1×1 卷积层分别预测类别分数、边界框回归参数和置信度，大大加快了网络的收敛速度，增强了模型的泛化能力和鲁棒性，准确率也有所提升。算法在选择正负样本时采用 SimOTA 正负样本匹配策略，通过计算每个特征点生成的候选框与真实框的 IoU 值，选择 10 个最优特征点，之后结合损失计算得到最终的预测框，其余候选框即为负样本。这种方法运算速度快且不需要额外超参数，能够在一定程度上避免出现正负样本不平衡的问题。

本文提出的目标检测模型 YOLOX- α SMV 对 YOLOX 算法进行针对性改进，由特征提取网络 Backbone、特征融合网络 Neck 以及检测头 Prediction 三部分组成，模型的网络结构如图 1 所示。

特征提取网络 Backbone：缺陷图片经过 Focus 进行初步下采样，使用不同大小的卷积运算和 SPP_SF 进行特征提取和融合，生成大、中、小三

种不同尺度的细粒度特征向量，用于检测不同大小的缺陷区域。

特征融合网络 Neck：三种不同尺度的特征向量先通过 MDFFM 融入缺陷的通道、空间和位置信息，再经过一系列上采样、下采样、特征拼接等卷积操作，加强对缺陷类别的特征提取能力，进行浅层和深层的高级语义信息交互，最终输出 3 个更加丰富的特征层。

检测头 Prediction： 80×80 、 40×40 、 20×20 三种预测不同缺陷大小的特征层经过一个 1×1 卷积调整通道数后，分成两个特征向量分别进行回归和分类。用于分类的特征向量通过 Varifocal Loss 加权正负样本，倾斜训练重心，使训练更高效；用于回归的特征向量通过 α -CIoU 提高预测框的回归精度，加速训练，最终和置信度预测拼接成每个特征层的预测结果。

1.1 SPP_SF 模块

YOLOX 网络在 Backbone 中引入 SPP 模块扩大感受野，融合不同尺度特征图的信息。如图 2 所示，SPP 主要由 3 个并联的最大池化层和残差连接构成，最大池化层选取特定区域的最大值为输出，保留主要特征的同时，降低模型计算量，防止过拟合。

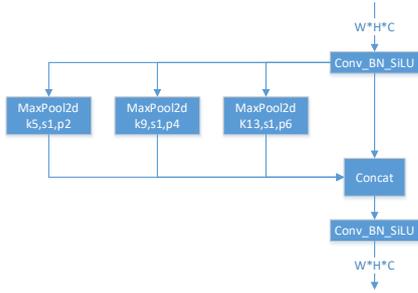


图2 SPP网络结构
Fig.2 SPP network structure

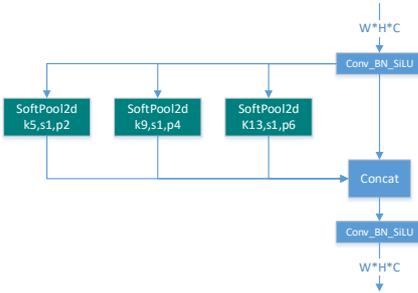


图3 SPP_SF网络结构
Fig.3 SPP_SF network structure

最大池化虽然降低了特征冗余度，但也丢失了图像中的大多数信息，降低了网络性能。Alexandros Stergiou 等^[12]受早期手工编码特征的池化实验启发，提出了软池化（SoftPool），既保持了基本的池化功能，让区域内的所有特征值都对最终输出有贡献，又减少了信息损失。

SoftPool 基于自然指数 e ，利用激活的 softmax 加权求和。整个过程如式（1），式（2）所示，首先计算核大小内的权重，再将权重与每个特征值相乘并求和，得到最终输出，从而保留输入的基本属性，并放大更强的特征映射。同时，整个过程是可微的，这表示在反向传播期间，核大小区域内的所有特征至少会被分配一个最小梯度值进行更新，提高了训练期间的神经连通性，有利于提高训练效果。

$$w_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}} \quad (1)$$

$$\tilde{a} = \sum_{i \in R} w_i * a_i \quad (2)$$

其中， R 为池化区域， w_i 为 R 内权重， a_i 为 R 内每个特征值。

本文保持融合特征图尺度不变，提出了 SPP_SF，如图 3 所示。相比于 YOLOX 中的 SPP，SPP_SF 将最大池化层替换为 SoftPool，在不增加

参数和计算量的情况下，保留原图中更多细节，提高分类准确率，加快推理时间。

1.2 MDFFM 多维度特征融合模块

带钢材料表面缺陷数据集中的图片分辨率较低，不同缺陷类型颜色背景、纹理特征较为相似，这给缺陷检测带来了难度。为此，本文引入了卷积注意力模块（Convolutional Block Attention Module, CBAM)^[13]和坐标注意力（Coordinate Attention, CA)^[14]，将通道、空间和位置信息融入特征图中，加强模型对缺陷目标的定位和识别能力。

CBAM 网络结构如图 4 所示，整个过程可表示为：

$$\begin{aligned} F_c &= M_c(F) \otimes F \\ F_s &= M_s(F_c) \otimes F_c \end{aligned} \quad (3)$$

式中， \otimes 为元素相乘。

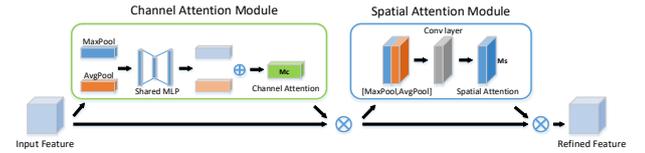


图4 CBAM 网络结构

Fig. 4 CBAM network structure

首先，通道注意力模块利用输入特征通道间关系，生成通道注意力特征图。为提高网络的表达能力，对压缩维度后的特征图同时使用平均池化和最大池化，聚合特征图的空间信息。将生成的两个特征向量通过一个多层感知器映射后，求和合并，得到通道注意力模块的特征向量 F_c 。整个过程 M_c 可表示为：

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (4)$$

式中， σ 为 Sigmoid 激活函数，AvgPool 为平均池化操作，MaxPool 为最大池化操作。

其次，空间注意力模块生成的特征图，与通道注意力互补。首先，沿输入特征向量的通道依次应用最大池化和平均池化操作，突出显示信息区域，并将它们连接起来，生成一个更有效的特征向量，并通过卷积运算得到空间注意力模块的特征向量 F_s 。整个过程 M_s 可表示为：

$$M_s(F_c) = \sigma(f^{(7 \times 7)}([\text{AvgPool}(F_c); \text{MaxPool}(F_c)])) \quad (5)$$

式中， $f^{(7 \times 7)}$ 为卷积核为 7×7 的卷积运算。

通过以上两步计算，网络的各个分支可以分别了解在通道和空间维度上，要“参与什么”和

“到哪里去”，帮助特征信息在网络中流动。

为弥补 CBAM 不关注或只关注局部位置信息的问题，并让网络更精确地定位缺陷位置，本文同时引入 CA 模块。CA 模块在两个空间方向上分别编码方向和位置映射，互补地增强目标区域的表现，整个过程如图 5 所示。

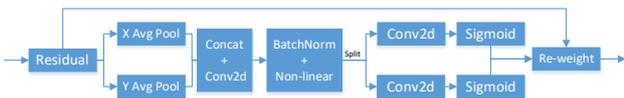


图 5 CA 注意力网络结构

Fig.5 CA attention network structure

首先，对坐标信息进行嵌入。CA 注意力将对通道的压缩转换为对两个一维特征的编码操作。即，对于输入特征，使用平均池化对水平和垂直方向的通道信息进行编码，分别沿着两个空间方向聚合特征，各生成一个对方向敏感的特征向量。

然后，将生成的两个特征向量连接起来，将它们输入到共享的卷积核大小为 1×1 的卷积运算中，以适当减少通道数，降低后续操作的开销。最后，利用另外两个卷积核大小为 1×1 的卷积变换，使产生的坐标注意力恢复到原始通道数，再与输入的特征向量融合。最终输出如公式 (6) 所示。

$$y_c(i, j) = x_c(i, j) * g_c^h(i) * g_c^v(j) \quad (6)$$

式中， g_c^h ， g_c^v 分别为两个空间方向的特征映射。

CA 坐标注意力沿输入特征向量的水平方向和垂直方向的编码过程都反映了感兴趣的对象是否存在于相应的行和列中，可以充分利用捕获的位置信息，准确地突出显示感兴趣区域，从而更好地识别缺陷位置和类别，提高模型的适用性和检测能力。

本文同时应用 CBAM 和 CA 模块，提出多维度特征融合模块，将通道、空间与位置 3 个维度的特征信息融入特征图中，丰富 Backbone 中输出的多尺度特征向量，为后续融合低层细节信息和高层语义信息提供帮助，增强模型检测能力。

1.3 Varifocal Loss 模块

在带钢材料表面缺陷图片中，存在的缺陷区域只占整个图片较少的部分，使用目标检测模型进行缺陷检测时，不可避免地会产生正负样本数不匹配的问题。因此，本文模型使用了 Varifocal Loss^[15]，可以有效地解决此问题。Varifocal Loss 计算方式如公式 (7) 所示：

$$\text{VFL}(p, q) = \begin{cases} -q(q \log(p) + (1-q) \log(1-p)) & q > 0 \\ -\alpha p^\gamma \log(1-p) & q = 0 \end{cases} \quad (7)$$

其中， p 为预测的 IACS 值；对于 q ，若是负样本，则为 0，若是正样本，则为预测的边界框和真实框的 IoU； γ 为比例因子。

Varifocal Loss 只对负样本进行了衰减，降低负样本对损失的贡献度。对正样本使用 q 加权处理，提高其对损失的贡献度，从而倾斜训练重点至高质量的正样本。实验证明，这样的处理可以大大增加不同缺陷类别的检测精度。

1.4 α -CIoU 模块

为解决预测框和真实框不重叠时，IoU 损失出现的梯度消失问题，出现了几种基于 IoU 改进的损失设计，如 CIoU、GIoU 等。He J 等^[16]在现有的 IoU 损失中引入了 power 变换，提出了一种新的损失函数 α -IoU Losses，进一步化简可得其一般形式为：

$$L_{\alpha\text{-IoU}} = 1 - \text{IoU}^\alpha \quad (8)$$

通过增加幂指数 α ($\alpha=3$)，使目标检测模型更灵活的实现不同 Bounding box (bbox) 的回归精度，而且，对小数据集和噪声框更健壮。根据 α -IoU Losses，可概括 CIoU 损失函数新的形式为：

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \beta v \quad (9)$$

$$L_{\alpha\text{-CIoU}} = 1 - \text{IoU}^\alpha + \frac{\rho^{2\alpha}(b, b^{gt})}{c^{2\alpha}} + (\beta v)^\alpha$$

其中， c 为同时包含预测框和真实框的最小包围区域的对角线距离， $\rho^2(b, b^{gt})$ 为预测框和真实框中心点的欧式距离， βv 用于度量长宽比的一致性。

CIoU 相比 IoU，增加了重叠面积、中心点距离、长宽比、检测框尺度损失四种几何参数，而 α -CIoU 损失函数优于原本形式，让 α -CIoU 损失在不引入额外参数、不增加推理时间的情况下，通过自适应增加高 IoU 对象的损失和梯度权重，使模型更关注高 IoU 对象，从而提高了 bbox 的回归精度，增强了模型对部分小目标缺陷和含有噪声缺陷的定位、识别能力。

2 实验结果与分析

2.1 实验环境与参数设定

实验运行环境如表 1 所示。

表1 实验环境
Tab.1 Experimental environment

Lab environment	Platform parameters
System	Windows 10
GPU	NVIDIA GeForce RTX 3070 8G
CPU	AMD Ryzen 7 5800H
DL framework	pytorch 1.10.0 + cuda 11.3
IDE	PyCharm

模型训练迭代次数 (epoch) 设置为 300 次, batch_size 设置为 8。优化器采用随机梯度下降法 (SGD), 初始学习率为 0.01, 使用余弦退火学习策略, 权重衰减为 0.0005。最后 15 轮取消 Mosaic 和 Mix-up 数据增强, 只保留随机水平翻转和颜色扰动两个较为简单的数据增强。

2.2 数据集

实验所用的 NEU-DET 数据集包含 6 类缺陷, 如图 6 所示, 分别为裂纹 (Cr)、夹杂 (In)、斑块 (Pa)、表面点蚀 (Ps)、轧制氧化皮 (RS)、划痕 (Sc), 每种缺陷有 300 张, 共 1800 张图片。按照 7:2:1 随机划分数据集, 其中 1260 张训练集, 360 张验证集, 180 张测试集。

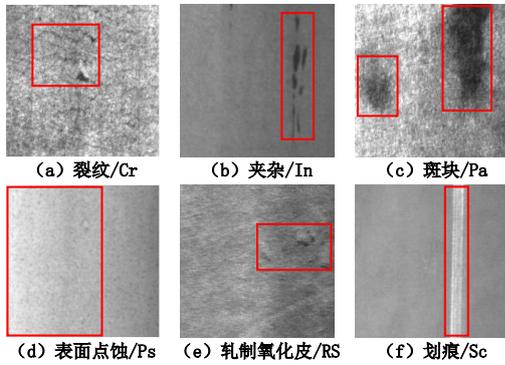


图6 6种缺陷示意图

Fig. 6 Schematic diagram of 6 kinds of defects

2.3 模型训练

本文算法的训练过程如图 7 所示。由图中曲线的变换趋势可知, 随着训练次数的增加, 曲线下降趋势逐步减缓并趋于平稳, 最后 15 轮时, 由于数据增强的关闭, 去除了外在因素引起的训练误差, 损失又大幅下降, 并完成最终收敛。相较于 YOLOX, 本文算法损失下降更快、更多, 最先趋于平缓, 在关闭数据增强后, 损失下降幅度也较小, 充分证明了本文算法的训练效果突出, 训练过程稳定。

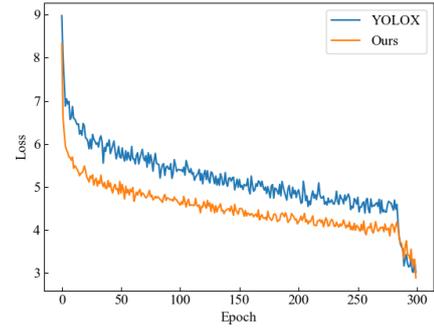


图7 损失函数对比

Fig. 7 Comparison of loss function

2.4 评价指标

实验采用平均精度均值 (mean average precision, mAP), 检测刷新率 (frame per second, FPS) 和参数量 Params 作为模型性能的评估指标。

1) 本文中 IoU 为真实框和预测框的交并比, 如公式 (10) 所示。

$$IoU = \frac{A \cup B}{A \cap B} \quad (10)$$

其中, A 为真实框区域, B 为预测框区域。

2) FPS 为每秒钟检测的图片数量, 如式 (11) 所示。

$$FPS = \frac{FrameNum}{ElapsedTime} \quad (11)$$

其中, $FrameNum$ 为检测图片数量, $ElapsedTime$ 为模型检测运行的总时间。

3) mAP 即 $mAP@0.5:0.95$, IoU 从 0.5 取到 0.95, 步长为 0.05, 计算所有 AP 的均值; mAP_{50} 为 IoU 取 0.5, 计算所有 AP 均值 (本文主要采用 mAP 作为评价指标, 力求更严格地评估算法性能)。计算方式如式 (12), 式 (13) 所示。

$$AP = \int_0^1 p(r) dr \quad (12)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (13)$$

其中, $p(r)$ 为以准确率 P 为纵坐标, 以召回率 R 为横坐标的 PR 曲线。

2.5 对比实验及结果分析

为验证本文算法在 NEU-DET 数据集上的检测效果, 在相同实验环境和数据集划分条件下, 选取了两阶段目标检测网络 Faster RCNN^[17]、单阶段目标检测网络 CenterNet^[18]、SSD^[19]、

表2 不同模型的实验对比
Tab.2 Experimental comparison of different models

Model	Backbone	Input_size	AP ₅₀ /%						mAP ₅₀ /%	mAP /%	Params /M	FPS /(f/s)
			Cr	In	Pa	Ps	RS	Sc				
Faster RCNN	ResNet50	1333×800	37.50	77.00	86.50	84.50	64.90	88.00	73.10	40.1	41.15	20.3
CenterNet	ResNet18	512×512	40.10	61.50	83.90	71.60	60.30	59.50	62.80	28.9	14.21	224.4
SSD	VGG16	300×300	42.80	61.20	89.10	75.10	59.40	55.00	63.80	31.5	24.41	67.9
FCOS	ResNet50	1333×800	42.50	59.20	87.80	76.90	61.40	60.10	64.70	27.8	31.85	22.0
YOLOv3 ^[21]	DarkNet53	416×416	44.00	72.10	61.00	72.90	71.30	85.90	67.90	34.0	61.55	63.5
YOLOv5	CSPDarkNet	640×640	27.10	83.50	86.10	77.90	60.50	86.00	70.20	38.6	7.2	99.8
YOLOX	CSPDarkNet	640×640	44.82	82.50	86.48	82.78	70.64	92.95	76.70	44.11	8.94	177.2
YOLOv7 ^[22]	E-ELAN	640×640	46.10	83.50	83.20	74.40	64.50	91.10	73.80	45.9	36.9	138.9
文献[23]	CSPDarkNet	640×640	34.30	84.10	87.60	82.00	61.90	87.00	72.80	36.40	23.9	75.8
YOLOv8	CSPDarkNet	640×640	45.90	84.50	90.10	85.10	71.20	94.60	78.60	47.3	11.2	115.8
YOLOX- α SMV	CSPDarkNet	640×640	46.48	84.61	90.77	85.16	71.86	94.16	78.84	47.54	9.02	175.5

FCOS^[20]、YOLO 系列以及钢材缺陷目标检测领域的算法进行对比实验分析，统计不同主流模型的评价指标，实验结果如表 2 所示。

由表 2 实验结果可知，YOLOX- α SMV 的 mAP 为 47.54%，相比于其他目标检测算法均有不同程度的提高，比 YOLOv3 提高了 13.54%，比 YOLOv5 提高了 8.94%，比 YOLOv7 提高了 1.64%，比文献[23]提高了 11.14%，比基准模型 YOLOX 提高了 3.43%。并且，除在 Sc 缺陷类别上检测精度略低于 YOLOv8 外，YOLOX- α SMV 对每种缺陷类别的检测精度均高于其他对比算法，mAP₅₀ 达到了最优值 78.84%。

YOLOX- α SMV 模型的参数量为 9.02 M，比大多数模型参数量低，仅比 YOLOv5 高 1.82 M，而检测速度达到了 175.5 f/s，比 YOLOv5 提高了 75.7 f/s，比 YOLOv7 提高了 36.6 f/s，与基准模型 YOLOX 相比基本保持不变。

与目前目标检测领域的 SOTA 算法 YOLOv8 相比，本文模型 mAP 提高了 0.24%，参数量只有 YOLOv8 的 80%，检测速度提高了 51.55%。

综合以上实验结果可知，在 NEU-DET 数据集上，YOLOX- α SMV 以较小的代价，达到了较优的检测结果，说明本文算法适用于带钢材料表面缺陷检测，有较强的缺陷检测能力，能够检测出其他目标检测模型难以识别的模糊缺陷、小目

标等缺陷。

为验证本文算法的泛化能力，进一步在钢材表面缺陷数据集 GC10-DET 和公共数据集 PASCAL VOC2012 上对 YOLOX 算法和 YOLOX- α SMV 算法进行对比实验。GC10-DET 包括冲孔、焊缝、夹杂物、新月形缝隙等 10 种缺陷类别；VOC2012 包括人、动物、交通工具、家具四大类，共 20 种常见目标类别，实验结果如表 3 所示。

表3 在两个数据集上的对比实验结果
Tab.3 Comparative experimental results on two datasets

Datasets	Model	mAP/%	FPS/(f/s)
GC10-DET	YOLOX	34.90	163.67
	YOLOX- α SMV	37.67	161.98
VOC2012	YOLOX	55.66	143.10
	YOLOX- α SMV	57.68	141.30

由表 3 数据可知，本文算法在 GC10-DET 上，mAP 达到了 37.67%，提高了 2.77%，在 PASCAL VOC2012 数据集上，mAP 达到了 57.68%，相对于 YOLOX 网络，提高了 2.02%，FPS 均基本保持不变，表现良好。充分说明了本文提出的 YOLOX- α SMV 算法在同类型数据集和公共数据集上依旧具有良好的泛化能力和检测性能。

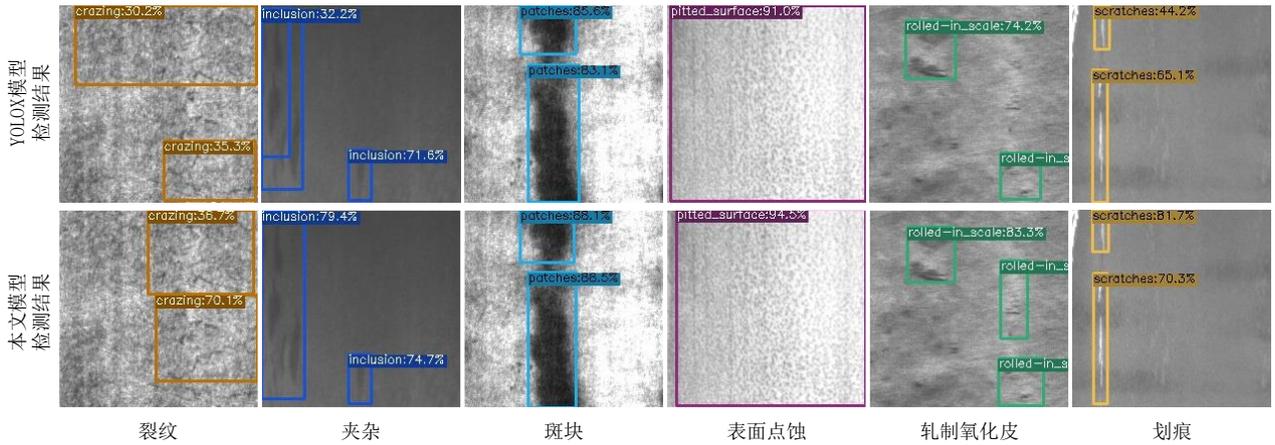


图 8 改进前后缺陷检测对比

Fig. 8 Comparison of defect detection before and after improvement

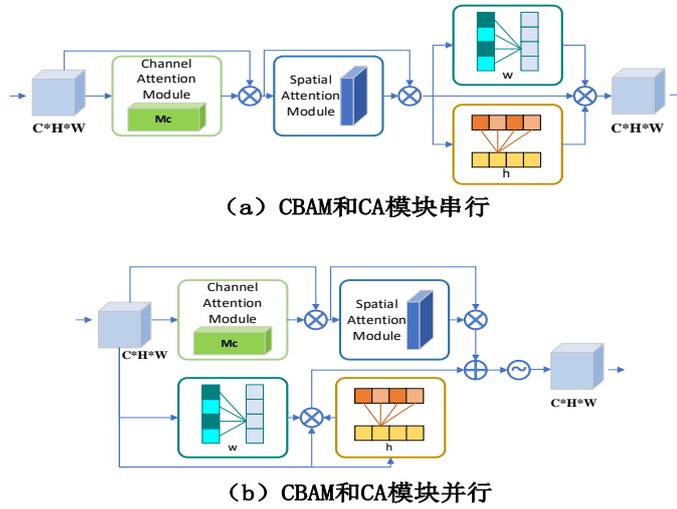


图9 两种不同结构的多维度特征融合模块对比图

Fig. 9 Comparison diagram of two multi-dimensional feature fusion modules with different structures

图 8 直观地展示了 YOLOX 算法和本文算法在实际检测中对 6 种缺陷的检测效果。不同缺陷类别使用不同颜色的检测框，检测框左上角即为缺陷的检测类别和置信度，置信度越高，说明检测框中为该缺陷的可能性越高。两种算法对比可知：本文算法能检测出 YOLOX 算法检测不到的小目标缺陷；检测框更贴近真实缺陷区域；相同区域检测框的置信度更高。这充分证明了本文对 YOLOX 算法改进的有效性，改进后的算法更适用于带钢材料表面缺陷检测，具有良好的实用价值。

2.6 消融实验及分析

2.6.1 CBAM 和 CA 模块融合方式

为对比 CBAM 和 CA 模块融合多维度信息不同顺序和方式对模型精度的影响，本文设计了如图 9 所示的两种多维度特征融合模块。

图 9 (a) 为 CBAM 和 CA 模块串行，特征向量先经过 CBAM 融合通道和空间信息，接着通过 CA 模块融入缺陷的位置信息，最后输出融合后的特征向量。图 9 (b) 为 CBAM 和 CA 模块并行，输入的特征向量同时经过 CBAM 和 CA 模块融入相应维度的特征信息，再相加并通过一个 SiLU 激活函数后输出。不同融合结构的消融实验如表 4 所示。

表4 不同融合结构消融实验

Tab.4 Ablation experiment of different fusion structures

Structure	Params/M	mAP/%	FPS/(f/s)
(a)	9.02	47.54	175.5
(b)	9.02	45.64	176.3

由表 4 数据可知, CBAM 和 CA 模块采用串行结构虽然在速度上比并行结构略慢, 但精度提高了 1.9%。因此, 本文模型采用 CBAM 和 CA 模块串行融合多维度特征信息。

2.6.2 消融实验

为进一步验证本文模型的检测能力, 分析 4 种改进方案对算法的作用, 在 NEU-DET 数据集上, 设置相同的运行环境和实验超参数, 对各改进方案

进行消融实验, 表中, “√”表示使用该改进方案。由表 5 实验数据可知, 单独采用 4 种改进方案对模型的参数量、检测能力和检测速度都有不同程度的影响。由实验 2, 3, 4, 5 可知, 4 种改进方案对模型性能都有增益, MDFFM 的引入加强了模型的特征提取能力, mAP 提升了 2.84%, SPP_SF 保留了更多的特征信息, 提高了模型的训练效率, FPS 达到了 182.3 f/s。

表5 不同改进方案的消融实验
Tab.5 Ablation experiment of different improved schemes

No.	MDFFM	SPP_SF	VFL	α -CIoU	AP ₅₀ /%						mAP ₅₀ /%	mAP /%	Params /M	FPS /(f/s)
					Cr	In	Pa	Ps	RS	Sc				
1					44.82	82.50	86.48	82.78	70.64	92.95	76.70	44.11	8.94	177.2
2	√				41.95	83.01	88.63	85.00	74.59	93.99	77.86	46.95	9.02	155.9
3		√			42.51	82.82	88.45	83.52	70.25	93.29	76.81	45.2	8.94	182.3
4			√		42.91	82.33	86.87	83.94	72.75	93.24	77.01	44.9	8.94	178.8
5				√	46.02	82.42	88.02	85.12	70.11	94.05	77.62	45.24	8.94	180.6
6	√	√			43.79	83.54	88.62	85.07	73.22	93.91	78.03	47.12	9.02	168.0
7			√	√	44.99	82.51	89.98	85.15	71.61	93.18	77.90	46.53	8.94	184.9
8	√	√	√	√	46.48	84.61	90.77	85.16	71.86	94.16	78.84	47.54	9.02	175.5

由实验 6 可知, 在 YOLOX 网络中同时应用 MDFFM 和 SPP_SF, 检测速度下降了 9.2 f/s, mAP 提高了 3.01%, 检测精度提升最为明显, 这说明了虽然 SPP_SF 对检测精度效果有限, 但可以弥补 MDFFM 在速度上的不足, 有效提高了模型的检测能力。由实验 7 可知, 同时引入分类损失 Varifocal Loss 和回归损失 α -CIoU, 在不增加模型参数量的前提下, mAP 提高了 2.42%, 检测速度也提高了 7.7 f/s, 说明将有限的训练资源倾斜到有目标的正样本上, 并优化 IoU 计算方式可以提高模型鲁棒性和检测速度。

由实验 8 可知, 同时采用这 4 种改进方案, 参数量提高了 0.08 M, 检测速度下降了 1.7 f/s, 在小幅度牺牲模型大小和检测速度的情况下, mAP 达到了 47.54%, mAP₅₀ 达到了 78.84%, 分别提高了 3.43%, 2.14%。RS 类缺陷由于部分样本偏暗, 导致特征提取过程中的信息丢失, 未达到最优, 不过仍优于 YOLOX 网络。除 RS 缺陷外, 其余缺陷类别精度均达到最优, 进一步证明了本文模型在带钢材料表面缺陷检测中的有效性。

3 结论

本文提出了一种改进 YOLOX 的带钢表面缺陷检测算法 YOLOX- α SMV。在 Backbone 部分应用 SPP_SF 保留池化过程中的特征信息, 降低冗余度, 防止过拟合。在 Neck 部分的开始位置加入 3 个不同特征尺度的多维度特征融合模块, 融入通道、空间、位置信息, 增强网络的识别和定位能力。在损失函数中引入分类损失 Varifocal Loss 和回归损失 α -CIoU, 在平衡正负样本的同时, 合理加权不同样本, 提高检测框的回归精度, 增强模型对不同缺陷类型的检测能力。实验结果证明了改进算法有效弥补了原网络的漏检、错检情况, 检测框的置信度也更高, 更接近真实框, 具有一定的鲁棒性。

本文算法在裂纹缺陷上的检测精度仍然较低, 后续将对此类缺陷的检测进行研究, 进一步提高算法的检测能力。

参考文献:

- [1] 曹义亲,何恬,刘龙标.基于改进 LSD 直线检测算法的钢轨表面边界提取[J].华东交通大学学报,2021,38(03):95-101.DOI:10.16749/j.cnki.jecjtu.20210706.006.
CAO Y Q, HE T, LIU L B. Rail Surface Boundary Extraction Based on Improved LSD Line Detection Algorithm [J]. Journal of East China Jiaotong University, 2021,38(03):95-101.DOI:10.16749/j.cnki.jecjtu.20210706.006.
- [2] HU B, WANG J. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network[J]. Ieee Access, 2020, 8: 108335-108345.
- [3] LIU X, GAO J. Surface Defect Detection Method of Hot Rolling Strip Based on Improved SSD Model[C]//International Conference on Database Systems for Advanced Applications. Springer, Cham, 2021: 209-222.
- [4] 曹义亲,周一纬,徐露.基于 E-YOLOX 的实时金属表面缺陷检测算法[J].图学学报,2023,44(04):677-690.
CAO Y Q, ZHOU Y W, XU L. A real-time metallic surface defect detection algorithm based on E-YOLOX [J] Journal of Graphics, 2023,44(04):677-690.
- [5] JIN X, WANG Y, ZHANG H, et al. DM-RIS: Deep multimodal rail inspection system with improved MRF-GMM and CNN[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1051-1065.
- [6] URBONAS A, RAUDONIS V, RYTIS MASKELIŪNAS, et al. Automated Identification of Wood Veneer Surface Defects Using Faster Region-Based Convolutional Neural Network with Data Augmentation and Transfer Learning[J]. Applied Sciences, 2019, 9(22): 4898.
- [7] 李维刚,叶欣,赵云涛,等.基于改进 YOLO v3 算法的带钢表面缺陷检测[J].电子学报,2020,48(07):1284-1292.
LI W G, YE X, ZHAO Y T, et al. Strip Steel Surface Defect Detection Based on Improved YOLOv3 Algorithm [J]. Acta Electronica Sinica, 2020, 48(07): 1284-1292.
- [8] ZHANG J, KANG X, NI H, et al. Surface defect detection of steel strips based on classification priority YOLOv3-dense network[J]. Ironmaking & Steelmaking, 2021, 48(5): 547-558.
- [9] 马燕婷,赵红东,阎超,等.改进 YOLOv5 网络的带钢表面缺陷检测方法 [J]. 电子测量与仪器学报,2022,36(08):150-157.DOI:10.13382/j.jemi.B2205354.
MA Y T, ZHAO H D, YAB C. Strip steel surface defect detection method by improved YOLOv5 network [J]. Journal of Electronic Measurement and Instrument, 2022,36(08):150-157.DOI:10.13382/j.jemi.B2205354.
- [10] 翁玉尚,肖金球,夏禹.改进 Mask R-CNN 算法的带钢表面缺陷检测[J].计算机工程与应用,2021,57(19):235-242.
WENG Y S, XIAO J Q, XIA Y. Strip Surface Defect Detection Based on Improved Mask R-CNN Algorithm [J]. Computer Engineering and Applications, 2021,57(19):235-242.
- [11] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021. arXiv: 2107.08430, 2021.
- [12] STERGIU A, POPPE R, KALLIATAKIS G. Refining activation downsampling with SoftPool[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10357-10366.
- [13] 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.
- [14] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.
- [15] ZHANG H, WANG Y, DAYOUB F, et al. Varifocalnet: An iou-aware dense object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8514-8523.
- [16] HE J, ERFANI S, MA X, et al. α -IoU: A Family of Power Intersection over Union Losses for Bounding Box Regression[J]. Advances in Neural Information Processing Systems, 2021, 34: 20230-20242.
- [17] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [18] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[J]. arXiv preprint arXiv:1904.07850, 2019.
- [19] LIU W, AMGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[c]//Computer Vision and Pattern Recognition 2015, Boston, MA, USA. 2015.
- [20] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9627-9636.
- [21] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [22] 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.
- [23] 曹义亲,伍铭林,徐露.基于改进 YOLOv5 算法的钢材

表面缺陷检测[J].图学学报,2023,44(02):335-345.

CAO Y Q, WU M L, XU L. Steel surface defect detection based on improved YOLOv5 algorithm [J].

Journal of Graphics,2023,44(02):335-345.

- [24] DING X, ZHANG X, HAN J, et al. Scaling up your kernels to 31×31 : Revisiting large kernel design in cnns[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11963-11975.
- [25] YANG R, MA H, WU J, et al. Scalablevit: Rethinking the context-oriented generalization of vision transformer[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 480-496.



通信作者: 曹义亲 (1964—), 男, 教授, 硕士, 硕士研究生导师, 研究方向为图像处理、模式识别。

E-mail: yqcao@ecjtu.edu.cn。