

基于改进 YOLOv5 的 O 型密封圈缺陷检测方法

朱文博¹, 夏林聪¹, 陈 龙¹, 吴晨睿¹, 陈红光²

(1. 上海理工大学 机械工程学院, 上海 200093; 2. 上海贝特威自动化科技有限公司, 上海 201109)

摘要: 针对 O 型密封圈缺陷难以人工识别的问题, 提出一种基于改进 YOLOv5 的表面缺陷自动检测方法。在数据预处理阶段, 采用半自动标注方法减少人工标注成本, 同时将拼接图片改为 9 张以实现 Mosaic 数据增强方法。在网络预测层引入标签平滑方法以减少模型过度依赖标签。在骨干网络中添加卷积注意力机制模块, 强化有效信息, 使骨干网络提取更加细致的局部特征信息。同时, 针对缺陷类型尺度变化大的特点, 引入剪枝的双向特征金字塔网络, 以解决大小缺陷在特征提取过程中的丢失问题。实验结果表明, 基于改进的 YOLOv5 与原 YOLOv5 相比, O 型圈表面缺陷检测平均精度均值提高了 4.26%, 并且检测速度在 25 ms 之内, 能够满足实际生产需要。

关键词: YOLOv5; O 型密封圈; 缺陷检测; 卷积注意力机制; 双向特征金字塔网络

中图分类号: TP 391 **文献标志码:** A

Defect detection method of O-ring based on improved YOLOv5

ZHU Wenbo¹, XIA Lincong¹, CHEN Long¹, WU Chenrui¹, CHEN Hongguang²

(1. School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. Shanghai Betterway for Automation Company, Shanghai 201109, China)

Abstract: Aiming at the difficulty of manual identification of O-ring defects, an automatic detection method of surface defects based on the improved YOLOv5 was proposed. In the data preprocessing stage, the semi-automatic labeling method was used to reduce the cost of manual labeling, and the number of mosaic images used was changed to nine to realize the mosaic data enhancement method. A label smoothing method was introduced in the network prediction layer to reduce the model's over-reliance on labels. The convolutional attention mechanism module was added to the backbone network to highlight the valid information, so that the backbone network could extract more detailed local feature information. At the same time, because of the characteristics of large scale changes of defect types, a pruned bidirectional feature pyramid network was introduced to tackle the loss of both large and small defects in the feature extraction process. The experimental results show that compared with the original YOLOv5, the average accuracy of O-ring surface defect detection in the improved YOLOv5 is increased by 4.26%, and the detection speed is within 25 ms, which can meet the actual production needs.

收稿日期: 2022-04-13

基金项目: 国家自然科学基金资助项目 (52075340, 52105525)

第一作者: 朱文博(1973-), 女, 副教授。研究方向: 数字化设计及制造。E-mail: teacherzwb@163.com

通信作者: 陈 龙(1978-), 男, 教授。研究方向: 产品智能计算设计与制造、机器视觉与机器人。E-mail: cl@usst.edu.cn

Keywords: YOLOv5; O-ring; defect detection; convolutional block attention; bidirectional feature pyramid network

O 型密封圈是目前应用最广泛、结构最简单的密封元件, 因此确保其质量尤为重要^[1]。O 型圈表面不仅存在尺度范围变化大的缺陷, 还存在特征相似的问题, 增加了实际检测的难度。目前的工业生产中, O 型圈的缺陷主要依靠人工检测发现, 这不仅大大增加了生产成本和工人劳动强度, 而且检测效率低、误差大, 难以保证检测结果的准确性^[2]。

机器视觉检测技术具有非接触、可在线、客观、自动化等突出优点^[3-4], 适合现代制造业的产品检测^[5]。但应用到 O 型圈缺陷检测时, 由于其缺陷像素占比不到整张图像像素的 0.5%, 造成模型可用的分辨率有限和特征不明显, 导致小缺陷检测困难^[6]。模板匹配法^[7]通过计算缺陷的周长或面积来判断缺陷是否在规定范围内, 从而确定其有无异常。文献 [8] 使用面积法检测表面缺陷, 计算各个连通区域的面积, 通过设定相应阈值来判断缺陷。文献 [9] 将获得的图像进行奇异值分解, 通过奇异区域检测缺陷。以上传统图像处理的方法主要适用于 O 型圈的边缘检测或者对特定类型的缺陷进行单独检测, 其缺点是难以设计一种或多种算法来全面满足各种缺陷同时检测的要求。

近年来, 基于卷积神经网络的深度学习算法在检测领域得到较多应用^[10], 例如双阶段检测算法中的 Faster R-CNN^[11] 和 R-FCN^[12], 单阶段检测算法中的 SSD^[13] 和 YOLO^[14]。但是, 对于具体实例, 尤其是小特征缺陷检测很难取得令人满意的效果^[15]。针对较小特征的检测, 文献 [16-17] 围绕图像特定区域进行处理, 聚焦于双阶段检测算法, 以牺牲推理时间来实现更好的性能, 因而难以满足实时检测要求。YOLOv5 是一种单阶段检测算法^[18], 结构清晰灵活, 计算量小, 识别速度快, 检测性能好, 但对于检测较小的物体时精度较低。

本文提出一种基于改进 YOLOv5 的 O 型圈缺陷检测算法, 相比上述传统图像处理算法, 该算法可同时全面地检测多种缺陷, 通过在网络的预测层嵌入标签平滑模块^[19]来解决模型对标签的过度依赖问题。在骨干网络中添加卷积注意力机制^[20]和在颈部结构中引入剪枝的双向特征金字塔网络^[21], 能使该算法对相似缺陷和小特征缺陷进行较高精度的检测, 满足工厂对不同类型缺陷进行分拣的要求。并且, 该算法可随时添加新的缺陷类型, 有助于后续工作的展开, 拥有较强的工艺柔性。

1 数据准备和预处理

1.1 数据集准备

通过工业相机采集全部 O 型圈图像, 数据采集装置如图 1 所示, 采集图像的原始分辨率为 1256×1080, 将筛选出存在缺陷的 5000 张图像中的 2000 张进行人工标注, 其中 1500 张作为初始训练集, 剩余 500 张作为测试集, 测试集中的缺陷类型和数量如表 1 所示。测试集虽是人工标出缺陷, 但在测试过程中只需将预测结果与人工标注的真实框对比计算即可。

半自动标注示例如图 2(a)所示, 利用初始权重自动推理出孔洞和刮蹭的标注信息, 即自动完

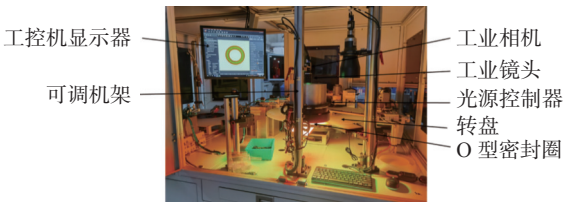


图 1 数据采集装置
Fig. 1 Data collection

表 1 测试集中的缺陷类型及数量
Tab.1 Number of defects in the test set

类型	断裂	褶皱	亮斑	孔洞	划痕	刮蹭	凸起	缺边
样图								
数量	93	68	179	72	43	151	246	79

成部分缺陷标注。其中，孔洞边界框不够紧致，需要人工微调，红色框内的凸起缺陷为漏检情况，需要人工补充。同理，将待标注的 3 000 张图片使用半自动标注方法进行粗标注。首先，通过 YOLOv5 将 3 000 张图片进行预测，得到预测框和类别，对预测模块的功能进行改进，将预测标签信息导入 Labelimg 模块中，即生成粗标注数据。最后，在 Labelimg 模块中通过人工微调及补充标签的方法得到标注数据集。数据标注流程如图 2(b)所示。

1.2 数据增强

在实际情况中，数据标注成本昂贵，创建虚

拟数据并添加至训练集是减少标注成本的最佳方法。YOLOv5 使用了 Mosaic 数据增强方法，即 4 张图片进行随机裁剪、缩放等操作后随机拼接成一张新的图像，增加小目标样本的数量，在进行归一化操作时，会一次性计算 4 张图片，提升网络训练速度。

由于采集得到的 O 型圈缺陷图像有限，但实际制造中产生的缺陷千变万化，并且其表面存在的缺陷通常不是单一类型的，可能存在多类型多数量的缺陷，因此提供给模型不同形态的缺陷数据越多，检测结果越准确。本文改进 Mosaic 方法，称之为 Mosaic-9，流程图如图 3(a)所示。从

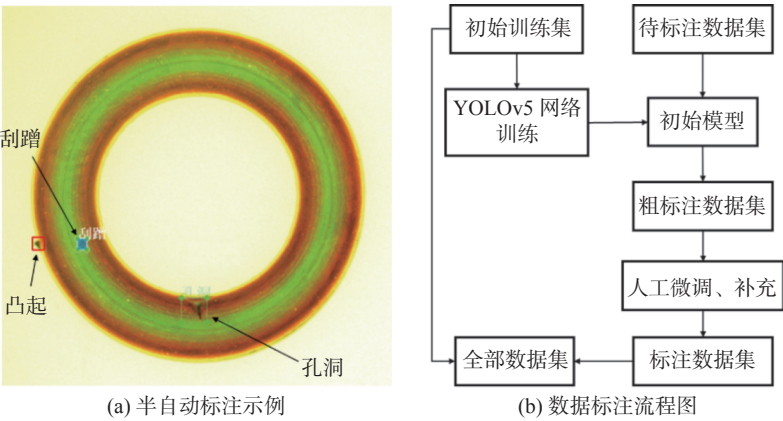


图 2 半自动标注示例与数据标注流程

Fig.2 Semi-automatic annotation example and data annotation process

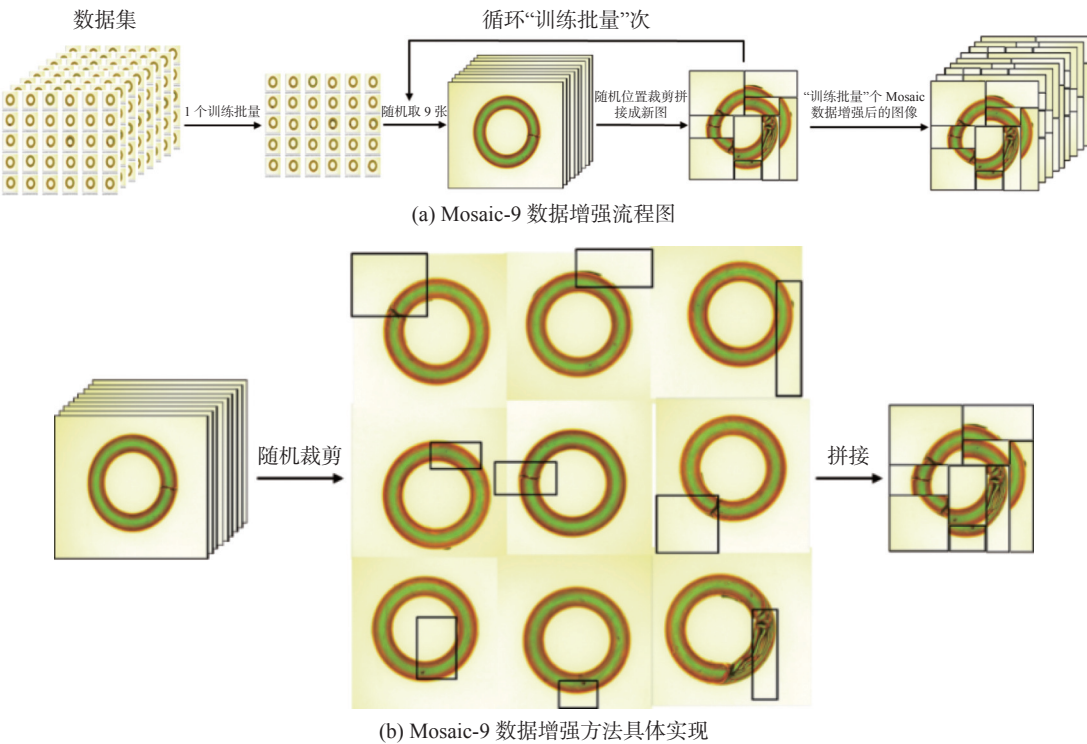


图 3 Mosaic-9 数据增强方法

Fig.3 Mosaic-9 data enhancement method

数据集中随机抽取 9 张图像作为 Mosaic 数据增强原图，9 张图像拼接的方法采用随机裁剪 5%~25%；由于缺陷在原图像的占比较小，对 9 张原图随机缩放的程度不宜过高，避免小型缺陷的特征丢失，采用随机缩放原图像 60%~110%。将 9 张图像随机裁剪缩放后拼接成一张新的图像，具体实现方法如图 3(b)所示。

1.3 超参数优化

超参数优化方法可分为两类：一类为可视化优化方法，比如 Zeiler 等^[22]利用反卷积可视化

AlexNet 网络；另一类为自动超参数优化方法^[23]，其中，网格搜索通过穷举的方式遍历参数的取值范围，具有简单、直观、全局优化的特点，适合低维度参数寻优，且不会陷入局部最优。本文采用网格搜索的方法来寻找最优的超参数数值。合适的超参数可以很快地达到训练效果，并且训练权重更加符合 O 型圈表面缺陷的检测。YOLOv5 微调了 29 个超参数，与默认超参数的模型相比，最终测试集上的检测正确率提高了 2.17%。超参数优化结果如表 2 所示。

表 2 超参数进化结果
Tab.2 Results of hyperparametric evolution

lr0	lrf	momentum	weight_decay	warmup_epochs	warmup_momentum	warmup_bias_lr	box	cls	cls_pw
0.01547	0.08892	0.95527	0.00032	1.71537	0.86814	0.12404	0.07157	0.70063	1.18413
obj	obj_pw	iou_t	anchor_t	fl_gamma	hsv_h	hsv_s	hsv_v	degrees	translate
0.79487	0.83363	0.2	3.44412	0.0	0.01473	0.47766	0.23671	6.67394	0.18679
scale	shear	perspective	flipud	fliplr	mosaic	mixup	copy_paste	anchors	
0.15021	0.0	0.0	0.23914	0.5	0.5	0.5	0.0	3.4203	

2 改进的 YOLOv5

$$y_j' = \begin{cases} \varepsilon, & y_j = 0 \\ 1 - \varepsilon, & y_j = 1 \end{cases} \quad (4)$$

2.1 标签平滑

标签平滑最早应用于分类算法中，之后才引入目标检测算法中，它们都在网络预测层中使用交叉熵损失。对于每一个输入样本 x ，损失函数 L_i 和 y_j 的计算方式如下所示^[19]：

$$L_i = - \sum_{j=1}^n y_j \ln y_j \quad (1)$$

$$y_j = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}} \quad (2)$$

式中： y_j 表示模型的输出； n 为样本数量； j 和 k 为初始值为 1 的两个变量。

通过式 (2) 可知，在交叉熵损失的计算过程中，模型会不断调整权重，使预测结果尽可能地接近 0 和 1。除此之外，YOLOv5 在生成真实标签的时候，存在偏差，因此通过自动标注获取的标签容易产生错标。针对此问题，在计算交叉熵损失的过程中引入了标签平滑技术。交叉熵损失的计算公式为^[19]

$$L_i = - \sum_{j=1}^n y_j' \ln y_j \quad (3)$$

基于实验结果，将 ε 值设置为 0.1。通过引入标签平滑机制，模型在训练过程中不会拟合 0 和 1 这样的极端值，避免了过拟合的现象，并在一定程度上提升了模型的泛化能力和学习速度，有益于 O 型圈数据半自动标注的效果。

2.2 骨干网络改进

在实际 O 型圈检测场景中，由于光照环境、缺陷相似及缺陷尺寸范围变化明显的影响，常规的目标检测算法受到置信度的影响，易出现漏检部分缺陷而造成损失。为了让骨干网络学习到更加细致的局部特征信息，并融合全局信息作出更加准确的判断，避免误检测和漏检测，本文对 YOLOv5 的骨干网络进行改进，在每一个跨阶段连接网络 (cross stage partial, CSP) 结构之后引入一个卷积注意力机制模板 (convolution block attention module, CBAM)，CBAM-Backbone 模块如图 4 所示。

卷积注意力机制模块可以在通道维和空间维分别推断出注意力的权重，使模型在训练过程中会重点关注缺陷区域，提取缺陷区域中更加全面的特征信息，加大特征提取网络的全局感受野，增强特征提取能力，有效地减少环境背景和相似缺陷对 O 型圈表面缺陷检测的影响。

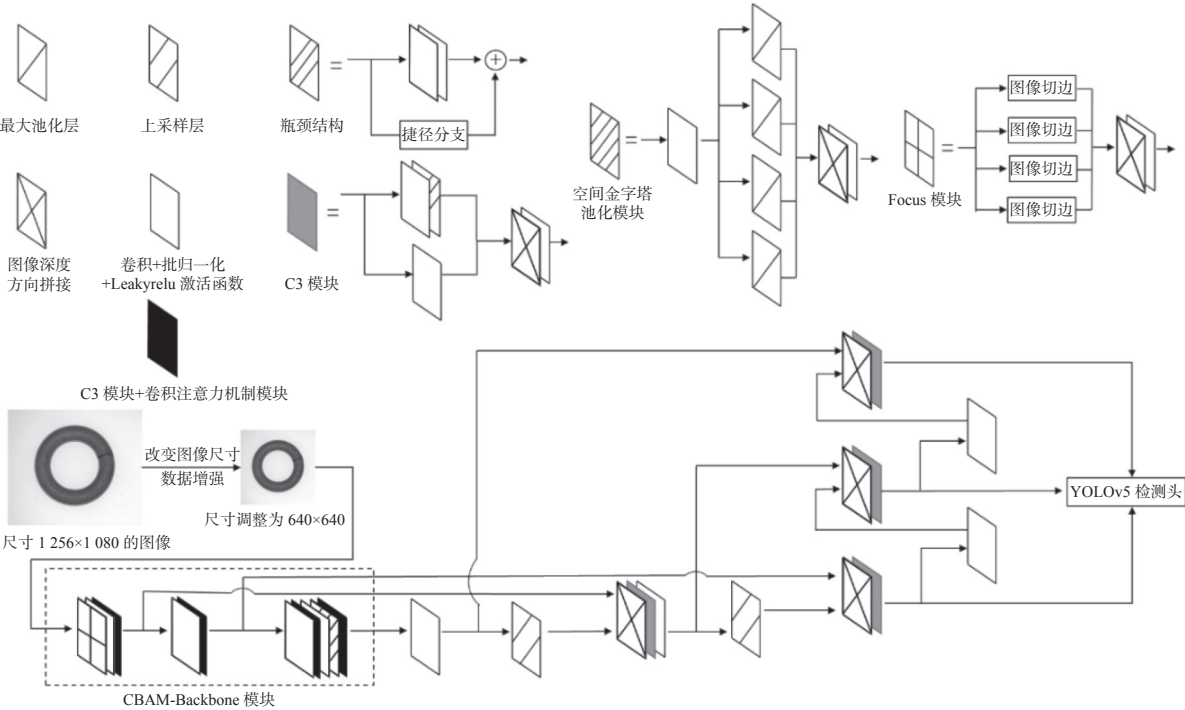


图 4 CBAM-Backbone 模块

Fig.4 Convolution block attention module-backbone module

通道注意力模块对输入特征图分别进行全局平均池化和全局最大池化来聚合图像的空间信息，经过 2 个多层感知机结构构建模型之间的相关性，最后经过 Sigmoid 激活函数获得每一个通道的权重。其表示形式为^[20]

$$M_c(F) = \sigma \left[W_1 \left(W_0 \left(F_{avg}^c \right) \right) + W_2 \left(W_0 \left(F_{max}^c \right) \right) \right] \quad (5)$$

式中： σ 表示对所有数据进行 Sigmoid 激活操作； F_{avg}^c 表示对输入特征矩阵进行全局平均池化； F_{max}^c 表示对输入矩阵进行全局最大池化； W_0 表示第 1 个全连接层的权重； W_1 表示第 2 个全连接层的权重； $M_c(F)$ 表示通道注意力模块的输出特征图。

空间注意力模块将通道注意力模块中的输出作为输入特征图。首先，对输入特征图做一个基于通道的全局平均池化和全局最大池化，将 2 个单通道特征图进行拼接；然后，通过一个 7×7 的卷积对特征图进行降维，获得一个单通道特征图，便于特征融合；最后，经过 Sigmoid 激活函数获得图像空间位置的每一个特征权重。对各个通道的特征进行加权操作，增强各个通道的信息交互，强化有效信息，抑制无效信息。其表示形式为^[20]

$$M_s(F) = \sigma \left(f^{7 \times 7} \left(\left[F_{avg}^s; F_{max}^s \right] \right) \right) \quad (6)$$

式中： $f^{7 \times 7}$ 表示对图像进行 7×7 的卷积； F_{avg}^s 表示对输入特征矩阵进行全局平均池化； F_{max}^s 表示对输入矩阵进行全局最大池化； $M_s(F)$ 表示空间注意力模块的输出特征图。

2.3 颈部结构改进

O 型圈表面不仅存在孔洞、亮斑和褶皱等尺度变化范围大的缺陷，还存在断裂和划痕等特征相似的缺陷。为了更好地进行不同尺寸的特征融合，捕捉小缺陷和相似缺陷的具体特征，提出剪枝的双向特征金字塔网络 (bidirectional feature pyramid network, BiFPN) 颈部结构，双向特征金字塔网络和剪枝的双向特征金字塔网络结构如图 5(a)和(b)所示。

具体带权特征融合方法的计算以图 5(b)的 P_2 输出为例，计算表达式为^[21]

$$P_2^{td} = Conv \left(\frac{w_1 P_2^{in} + w_2 Resize(P_3^{in})}{w_1 + w_2 + \epsilon} \right) \quad (7)$$

$$P_2^{out} = Conv \left(\frac{w_1' P_2^{in} + w_2' P_2^{td} + w_3' Resize(P_1^{out})}{w_1' + w_2' + w_3' + \epsilon} \right) \quad (8)$$

式中： $Conv$ 表示卷积操作； $Resize$ 表示下采样或上采样操作； $\epsilon = 0.0001$ 是一个较小的数，以避免数值不稳定； w 为学习到的参数，用于区分特征融合过程中不同特征的重要程度。

YOLOv5 作为通用物体检测器，在 O 型圈缺陷检测中无法达到实用效果，在特征金字塔网络^[24](feature pyramid network, FPN)结构中，将通过空间金字塔池化 (spatial pyramid pooling, SPP) 的 P_3 特征层作为最小特征层输出，但没有较好地

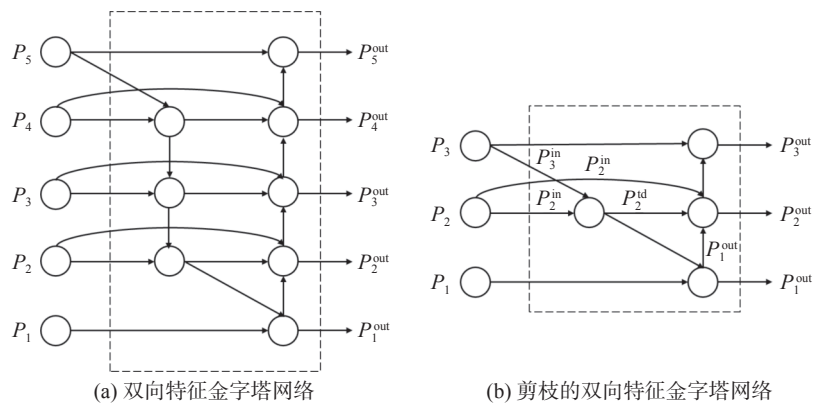


图 5 BiFPN 和剪枝的 BiFPN 结构对比

Fig.5 BiFPN and pruning BiFPN structure comparison

与其他尺寸的特征层融合。针对特征金字塔网络结构对 O 型圈表面小缺陷特征融合的不完全性, 对网络的颈部结构作出改进, 改进的颈部结构如图 6 所示。将骨干网络中提取的特征层 P_1 , P_2 , P_3 通过剪枝的双向特征金字塔网络结构进行特征融合, 使得骨干网络中提取到的特征层 P_3 通过上采样与特征层 P_2 融合。同理, 再次将上采样与特

征层 P_1 融合, 其中, 特征层 P_2 与上采样后的特征层 P_1 融合的特征层 P_2^{out} 再次融合, 最大限度获得 O 型圈表面缺陷中重要的特征信息。在通过颈部结构后端的金字塔注意力网络^[25](pyramid attention network, PAN)结构后得到 P_1^{out} , P_2^{out} , P_3^{out} 。YOLOv5 检测头分别输出 3 个尺寸为 19×19 , 38×38 和 76×76 的检测层, 用于检测大、中、小 3 种目标。

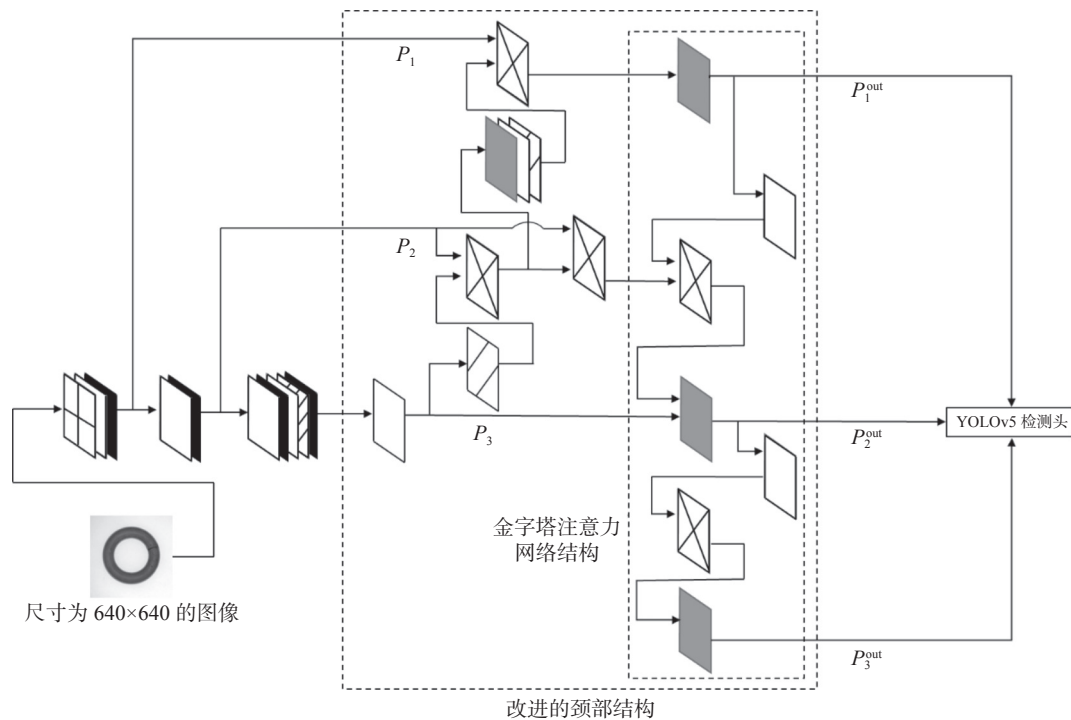


图 6 改进的颈部结构

Fig.6 Improved neck structure

3 实验分析与结果

3.1 实验分析

改进的 YOLOv5 模型搭建完毕, 通过计算机

完成相应实验, 计算机的具体配置如表 3 所示。利用训练集对改进的 YOLOv5 作训练优化, 实验采用随机梯度下降法, 具体参数如表 4 所示。准备待测试的图像数据, 将其全部放入改进的

表 3 实验平台相关配置表

Tab.3 Configuration table of experimental platform

环境	配置
CPU	AMD Ryzen Threadripper 3960X 24-Core Processor
GPU	NVIDIA GeForce RTX 3090
实验操作系统	Linux
深度学习框架	Pytorch
CUDA版本	11.3

YOLOv5 中，通过向前传播过程获取一个测试结果，将图像标签与测试结果作比对，计算出测试精度指标。

3.2 实验结果

在完成一轮训练并更新参数后，都要完成相应测试来检验此模型的泛化性和稳定性。如图 7 所示，box_loss 表示标签中边界框中心点横纵坐标 x 、 y 和宽高 w 、 h 带来的误差，class_loss 表示

表 4 模型训练参数

Tab.4 Model training parameters

训练模型	改进的YOLOv5
分辨率	640×640
训练样本数目	4 500
训练批量	64
训练迭代次数	500

类别带来的误差，mAP_50 表示交并比取 50% 时的平均精度均值。

不同算法实验结果对比如表 5 所示，针对各个缺陷类型，分别使用 SSD，YOLOv5 与改进的 YOLOv5 进行对比，改进的 YOLOv5 检测效果整体优于 SSD 算法和 YOLOv5 算法，其平均精度均值比 SSD 算法高出 5.93%，比 YOLOv5 高出 4.26%。在实际检测过程中，需要排查少数有缺陷的 O 型圈，因此，将候选框阈值适当降低来严格筛查有

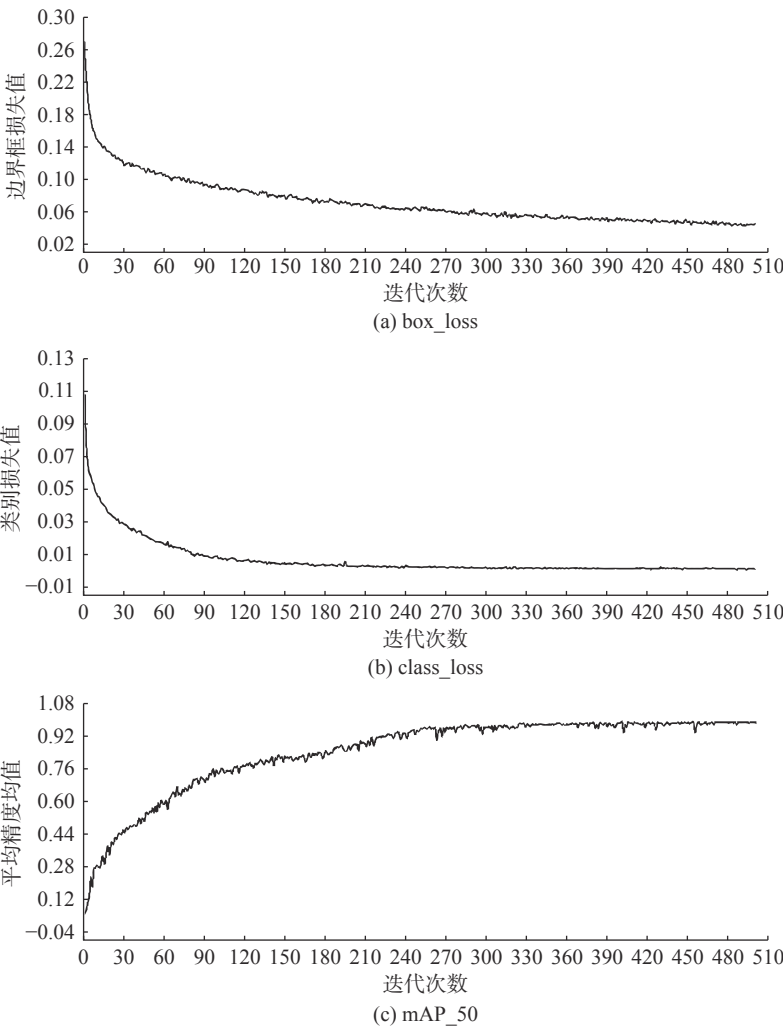


图 7 训练评价指标

Fig.7 Training evaluation index

表 5 不同算法实验结果对比

Tab.5 Comparison of experimental results of different algorithms

算法	平均精度/%								平均精度均值/%	召回率/%	检测时间/ms
	断裂	褶皱	亮斑	孔洞	划痕	刮蹭	凸起	缺边			
SSD	98.16	96.43	85.43	91.82	93.65	88.52	91.06	87.42	91.56	83	18
YOLOv5	98.32	96.52	89.75	93.45	94.17	90.46	92.23	90.94	93.23	86	15
改进的YOLOv5	99.57	99.52	94.64	97.96	97.65	96.23	97.87	96.50	97.49	94	24

缺陷的 O 型圈，可以接受将少数无缺陷的 O 型圈误检测为有缺陷以保证实际应用价值。经测试，本文所述改进的 YOLOv5 召回率为 94%，而 SSD，YOLOv5 检测召回率分别为 83% 和 86%。YOLOv5 与改进的 YOLOv5 检测结果示例如图 8(a) 和 (b) 所示，对于一些较小的缺陷，YOLOv5 出现漏检

情况，而改进的 YOLOv5 可以检测出较小缺陷的种类且精度符合实际应用要求，其推理时间由 15 ms 增加至 24 ms。但在实际的转盘机分拣系统中，当检测时间不超过 30 ms 时满足正常分拣要求。

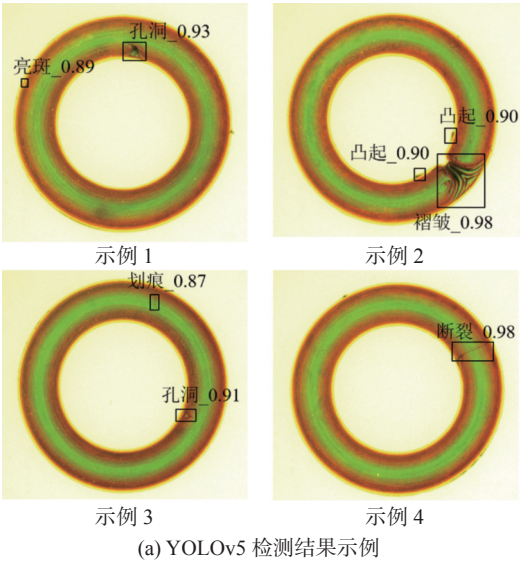
3.3 消融实验

为了更好地说明改进的 YOLOv5 中各个改进模块对检测效果的影响，在自制的 O 型圈数据集上进行消融实验，结果如表 6 所示。基于标签平滑方法的 YOLOv5 网络可以有效避免模型对半自动标注标签的过度依赖，与 YOLOv5 相比，其平均精度均值提高了 1.62%。通过添加 CBAM 注意力机制训练，获取更多重点关注目标的细节信息，抑制其他无用信息，该模块将平均精度均值从 94.85% 提高到 95.92%。将引入标签平滑方法和 CBAM 注意力机制的 YOLOv5 网络融入剪枝的双向特征金字塔网络结构，平均精度均值从 95.92% 提高到 97.49%。结果表明，改进后的特征融合网络结构可以有效检测 O 型圈表面的各种缺陷，提高网络的检测精度。

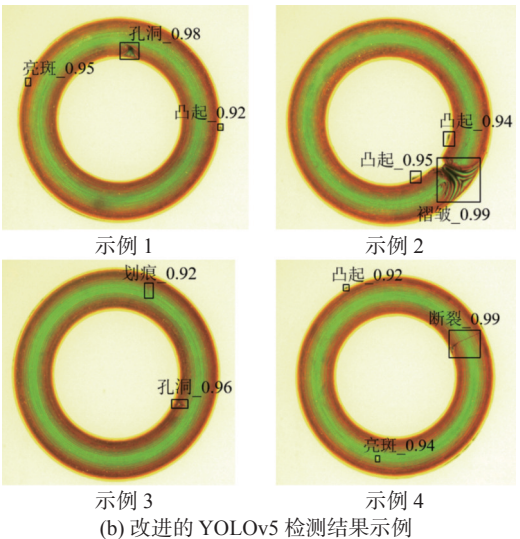
表 6 自制 O 型圈数据集上的消融实验				
Tab.6 Ablation experiments on a self-made O-ring dataset				
YOLOv5 基线	标签 平滑	卷积注意力 机制	剪枝的双向特征 金字塔网络	平均精度 均值/%
√				93.23
√	√			94.85
√	√	√		95.92
√	√	√	√	97.49

4 结束语

针对 O 型圈表面缺陷检测，提出一种改进的 YOLOv5 卷积神经网络模型，其性能优于原 YOLOv5 网络模型，有效地提升了检测 O 型圈表面缺陷的精度，为 O 型圈的智能分拣提供了一种可行的方法。本文通过改进 Mosaic 数据增强算法增大了小



(a) YOLOv5 检测结果示例



(b) 改进的 YOLOv5 检测结果示例

图 8 检测结果图

Fig. 8 Test result diagram

目标的数据量,减少批标准化计算的次数。在网络预测层使用标签平滑方法进行标签修正,有效提升了模型的泛化能力。针对原算法出现误检测和漏检测的情况,提出骨干网络中添加卷积注意力机制模块的策略,捕捉更多缺陷细节,融合全局信息对缺陷作出更加准确的判断。为了解决图像中小缺陷在特征提取过程中特征丢失的问题,通过剪枝的双向特征金字塔网络结构来加强对小缺陷的特征融合,将小缺陷的检测率由 89.46% 提升至 94.35%。在后续的研究中,将利用各种加速方法,在保证检测精度的前提下,提高算法的运行速度,以适应更为严苛的工业应用环境。

参考文献:

- [1] PENG G L, ZHANG Z J, LI W Q. Computer vision algorithm for measurement and inspection of O-rings[J]. *Measurement*, 2016, 94: 828–836.
- [2] 郑希鹏. 基于机器视觉的 O 型密封圈质量检测方法研究[D]. 合肥: 合肥工业大学, 2021.
- [3] 张涛, 刘玉婷, 杨亚宁, 等. 基于机器视觉的表面缺陷检测研究综述[J]. *科学技术与工程*, 2020, 20(35): 14366–14376.
- [4] DIESING G. How AI and machine vision intersect[J]. *Quality*, 2022, 61(2): 14.
- [5] 刘春, 江伟荣, 吕淑琴. 密封橡胶圈缺陷检测算法的研究[J]. *电子测量与仪器学报*, 2011, 25(5): 438–442.
- [6] CAO G M, XIE X M, YANG W Z, et al. Feature-fused SSD: fast detection for small objects[C]//Proceedings of the Ninth International Conference on Graphic and Image Processing (ICGIP 2017). Qingdao: SPIE, 2018: 106151E.
- [7] 王浩楠, 张晓青, 郭阳宽, 等. 基于机器视觉的轮胎胶料表面字符识别[J]. *电子测量与仪器学报*, 2021, 35(1): 191–199.
- [8] 曹勇. 基于机器视觉的工件检测系统若干问题的研究[D]. 合肥: 合肥工业大学, 2011.
- [9] 黄连, 刘晓军, 雷自力, 等. 基于奇异值分解的橡胶密封圈表面缺陷检测方法[J]. *润滑与密封*, 2021, 46(11): 84–88.
- [10] CHEN J Q. Image recognition technology based on neural network[J]. *IEEE Access*, 2020, 8: 157161–157167.
- [11] 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 & Machine Intelligence*, 2017, 39(6): 1137–1149.
- [12] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016: 379–387.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21–37.
- [14] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517–6525.
- [15] 刘颖, 刘红燕, 范九伦, 等. 基于深度学习的小目标检测研究与应用综述[J]. *电子学报*, 2020, 48(3): 590–601.
- [16] NAJIBI M, SINGH B, DAVIS L. AutoFocus: efficient multi-scale inference[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9744–9754.
- [17] LI Y C, ZHANG W D, CAI Y M, et al. Sniper based multi-target and multi-scale aerial image processing method[J]. *Journal of Physics: Conference Series*, 2020, 1659(1): 012003.
- [18] KUZNETSOVA A, MALEVA T, SOLOVIEV V. Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images[C]//Proceedings of the 17th International Symposium on Neural Networks. Cairo: Springer, 2020: 233–243.
- [19] ZHANG C B, JIANG P T, HOU Q B, et al. Delving deep into label smoothing[J]. *IEEE Transactions on Image Processing*, 2021, 30: 5984–5996.
- [20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3–19.
- [21] 董红召, 方浩杰, 张楠. 旋转框定位的多尺度再生物品目标检测算法[J]. *浙江大学学报(工学版)*, 2022, 56(1): 16–25.
- [22] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014: 818–833.
- [23] 陈森朋, 吴佳, 陈修云. 基于强化学习的超参数优化方法[J]. *小型微型计算机系统*, 2020, 41(4): 679–684.
- [24] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 936–944.
- [25] ZHAO T, WU X Q. Pyramid feature attention network for saliency detection[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 3080–3089.