

基于改进 RetinaNet 的白酒瓶盖缺陷检测方法^{*}

车璇 朱文忠 李韬 胡友鑫

(四川轻化工大学计算机科学与工程学院 宜宾 644000)

摘要:针对瓶装白酒包装质检存在的检测准确度低,小目标重合度高导致误检漏检的情况,提出一种基于 RetinaNet 的目标检测优化算法,主要使用白酒瓶盖瑕疵数据集进行检测。方法将网络 Backbone 替换为 Swin Transformer,其包含的窗口注意力机制运算有效提升瓶盖瑕疵检测精度同时降低复杂度节省了计算量。在 Neck 阶段使用神经架构搜索特征金字塔网络(FPN)代替 FPN,利用自动架构搜索选出最佳特征融合层,为后续检测提供更高质量的模型,最后采用 Soft-NMS 降低检测框置信度保留一定真实框,有效的防止瓶盖瑕疵过近或重叠造成漏检。实验证明,改进算法能够精准的识别出各类瓶盖瑕疵,检测精度在白酒瓶盖瑕疵数据集达到了 93.53%,相较于原网络提升了 8.02%。

关键词:瑕疵检测;RetinaNet;Swin Transformer;NAS-FPN;Soft-NMS

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.2

Defect detection method of liquor bottle caps based on improved RetinaNet

Che Xuan Zhu Wenzhong Li Tao Hu Youxin

(College of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin 644000, China)

Abstract: Aiming at the situation that the quality inspection of bottled liquor packaging has low detection accuracy and high overlap of small targets, resulting in false detection and missed detection, this paper proposes a target detection optimization algorithm based on RetinaNet, which mainly uses the defect dataset of liquor bottle caps for detection. In this method, the network backbone is replaced by Swin Transformer, which contains window attention mechanism calculation, which effectively improves the accuracy of cap defect detection, reduces complexity, and saves calculation. In the neck stage, neural architecture is used to search for FPN instead of FPN, automatic architecture search is used to select the best feature fusion layer to provide a higher quality model for subsequent detection, and finally Soft-NMS is used to reduce the confidence of the detection frame and retain a certain real frame, which effectively prevents the leakage of cap defects caused by too close or overlapping. Experiments show that the improved algorithm in this paper can accurately identify various bottle cap defects, and the detection accuracy reaches 93.53% in the liquor bottle cap defect dataset, which is 8.02% higher than the original network.

Keywords: defect detection; RetinaNet; Swin Transformer; NAS-FPN; Soft-NMS

0 引言

白酒行业作为中国的传统产业,在国民经济中的地位举足轻重。根据白酒企业半年报报道,在 2021 年白酒销售中,5 家企业营收规模近 100 亿元,6 家企业营收超 20 亿元,营收规模微低于 20 亿元也包含 8 家企业。随着市场需求,各大小型酿造企业高度重视生产设备现代化,适

应国内白酒行业机械化、自动化发展的趋势^[1]。在白酒包装生产环节,首先需要将空瓶及时提供到灌装线,再经过灌装-装盖-粘贴贴标-装盒-装箱-粘贴箱标签-封箱-检验-码垛等过程^[2]。其中装盖环节,由于生产技术,瓶盖会出现各类的瑕疵,而人工检测瓶盖存在速度慢、精度低、劳动强度大的问题,因此机器视觉代替人工检测逐渐成为主流。

近年来国内外均有众多研究者及团队针对物体瑕疵

收稿日期:2023-01-14

^{*} 基金项目:四川省科技研发重点项目(2019YFG0200)、四川省科技创新(苗子工程)培育项目(2022049)资助

检测提出各类改进算法,郑云峰^[3]对瓶盖检测提出了一种基于直线拟合的算法,能够检测出瓶盖是否存在高盖、歪盖或无盖的瓶盖异常情况。此方法主要是根据瓶盖在图片上的投影位置与直线位置进行比较来判断异常。徐锋等^[4]提出一种改进型粒子群优化-变分模态分解的特征参数提取及缺陷检测方法,新特征参数有效的为后续融合多参数特征,构建人工智能识别系统,及缺陷检测提供可靠依据。单明陶等^[5]则提出 Faster-RCNN 网络引入 ResNeXt50 为主干网络,并加入改进后的特征金字塔网络(FPN)结构,然后使用 ROI Align 解决区域匹配,有效在不增加特征量的情况下提高了特征提取能力并定位了缺陷特征信息,实现了缺陷的快速检测。杭芹等^[6]以处理缺陷尺寸跨度大、相似特征、识别难等问题,提出增加特征检测器,加入特征金字塔结构还有边界框融合算法的 YOLOv5 模型。段禄成等^[7]采用了 YOLOv3 算法检测瓶盖缺陷,并将 SENet 模块引入加强特征提取,将自适应特征网络(ASFF)加入 Neck 层,融合特征提高预测能力,同时为了解决样本均衡问题引入 Focus Loss,有效适应了工厂检测需求。Biradar 等^[8]使用了有监督的 3 层深度卷积神经网络(DCNN),每层卷积都包含支持向量机(SVM)作分类器,网络检测出每个像素点之间局部连通性,有助于学习检测物体结构并辨别缺陷。Krummenacher 等^[9]则对缺陷监测分步处理,首先根据监测目标数据设计小波特征,并使用监督学习,然后通过深度学习设计训练对于不同缺陷类型的卷积神经网络,该方法对缺陷进行了有效且目标性强的检测。Rudolph 等^[10]提出了 DifferNet,利用卷积神经网络提取的特征的描述性,使用归一化估计它们

的密度,并采用了一个多尺度特征提取器,使归一化流为图像分配有意义的可能性。在训练和评估中利用了多重转换,实现了高鲁棒性和性能,使网络在样本过少时也能表现良好。

基于深度学习的目标检测算法目前主要分为两大类,一类是以 RCNN^[11]、Faster-RCNN^[12] 等为代表的 2 阶段算法,另一类是以 YOLO^[13]、SSD^[14] 等为代表的单阶段算法。两阶段算法通过特征提取生成候选框后利用卷积神经网络对样本进行回归和分类,具有精度高、速度慢的特点,而为了改进速度问题提出的单阶段检测又损失了一定的精度。本文综合考虑选用 RetinaNet^[15] 单阶段网络,其检测精度一度超过双阶段网络。

为了适应检测所用的数据集以及缺陷类型,主要对 RetinaNet 做了如下改进。首先,为了提升网络对多尺度特征的检测能力,将主干网络由 ResNet50 替换为 Swin Transformer^[16],特征加强层使用 NAS-FPN^[17] 代替 FPN,加强特征层融合。最后为了防止检测框重叠面积大被误删,使用 Soft-NMS^[18] 抑制回归框的分数,增大检测率。实验证明,本文模型检测结果优于各类神经网络,满足工厂对瓶盖瑕疵检测精度的需求。

1 RetinaNet 算法及改进

1.1 RetinaNet

原 RetinaNet 使用 ResNet 提取不同尺度特征图后通过特征金字塔网络进行特征融合,最后对融合的特征图采用权重共享基于卷积操作的预测器,对分类和回归进行预测,得出结果。网络结构如图 1 所示。

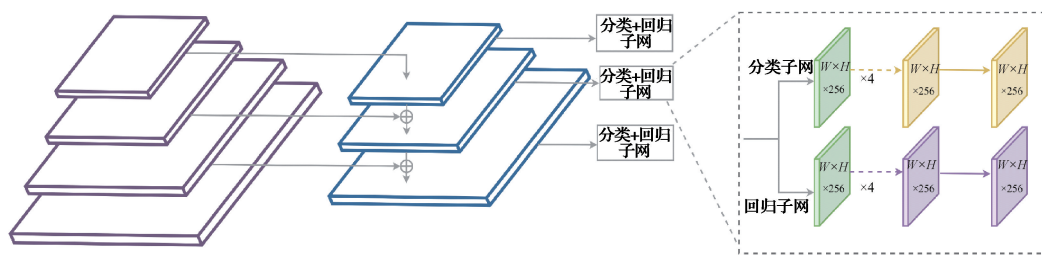


图 1 原 RetinaNet 网络结构

RetinaNet 的核心贡献点在于使用 Focal loss^[19] 作为损失函数,由于单阶段网络产生候选框过多,正负样本的比例失调,Focal loss 作为一个动态缩放的交叉熵损失函数,可以在训练过程中降低容易辨别的样本权重,让 loss 重心放在难辨别样本上,解决正负样本比例失衡问题。瓶盖缺陷检测中有些类别目标大,特征明显易检测,有些类别不仅重叠且目标极小,特征难捕获到,带有 Focal loss 的 RetinaNet 可以使模型聚焦到不易检测的样本上,提高效率。Focal loss 公式如下:

$$FL(p_i) = -(1 - p_i)^\gamma \log(p_i) \quad (1)$$

式中: γ 为超参数,取值范围 $0 \sim 5$; p_i 用来区分样本的难

易程度。 p_i 趋于 1 时则是易区分样本,此时 $1 - p_i$ 作为调制因子趋于 0,降低易区分样本的损失程度,遇到难区分样本时函数会反向调节。

网络中分类使用 Focal loss 函数对 FPN 提取的正负样本中的损失都进行计算,实现了复杂样本的聚焦,回归损失仅对正样本进行损失预算。整体损失计算公式如下:

$$Loss = \frac{1}{N_{pos}} \sum_i L_{cls}^i + \frac{1}{N_{pos}} \sum_j L_{reg}^j \quad (2)$$

式(2)由分类损失和回归损失相加, i 为所有的正负样本, j 为所有的正样本。 N_{pos} 是正样本个数。 L_{cls} 和 L_{reg} 为分别为 Sigmoid Focal loss 和 L1 loss 函数。

1.2 本文改进模型

本文在 RetinaNet 的基础上使用在视觉领域中可处理目标检测的 Transformer, Swin Transformer 为 Backbone, Neck 选取神经架构搜索和特征金字塔结合的 NAS-

FPN,得到最佳融合特征便于分类和回归处理,最后加入Soft-NMS,根据目标预测框交并比作分数抑制并对置信度修改,极大保留真实框提升了准确度,改进后网络结构如图2所示。

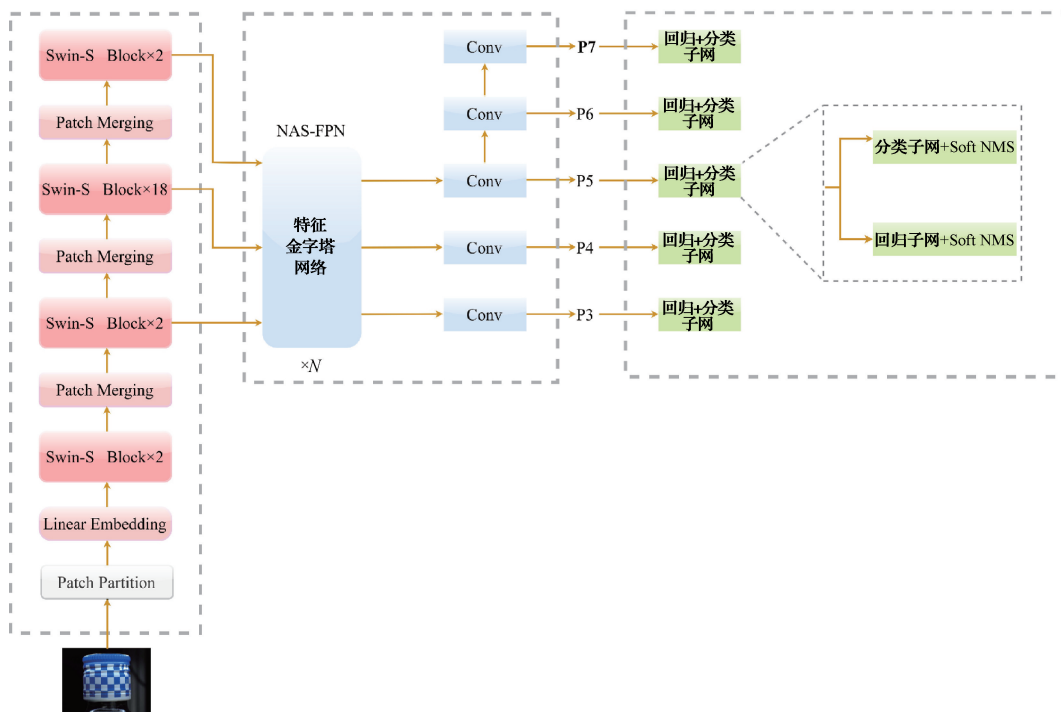


图 2 改进后 RetinaNet 网络结构

1) 更换主干网络为 Swin Transformer

Swin Transform 基于自注意力 (self-attention) 这种独特机制, 更好的利用了特征状态及不同像素点之间的信息, 比起 ResNet、VGG 等卷积神经网络只针对区域像素点单独检测, 获取了更多的信息。网络构建了分层模式并在较深的 Transformer 层逐渐合并相邻的 Patch, 对特征图的尺寸从 4 倍、8 倍、16 倍进行下采样处理, 有助于构建目标检测任务, 利用了分层特征图, 可以方便的结合类似 FPN 这种特征加强算法进行密集检测。RGB 图像首先被输入到 Patch Partition 模块进行分块处理, 将相邻像素 4×4 组合为 1 个 Patch 然后展平, 图像的长和宽缩为原本的 $1/4$, 通道数由 3 转换为 48, 经过 Stage1 的 Linear Embedding 对每个 Patch 数据做线性变换, 通道数由 48 转换为 C , 不同类型的 Swin Transformer 中 C 的大小不同, 本文实验使用模型为小型移动窗口 (Swin-tiny, Swin-T) 中 C 的取值设为 96。4 个 Stage 的目的就是构建大小不同的特征图, Stage 2—3 的 Patch Merging 是下采样模块, 核心模块 Swin Transformer 由两种结构组成, 数据先经窗口自注意力机制 (window based self-attention, W-MSA) 再由滑动窗口自注意力 (shifted window based self-attention, SW-MSA) 处理, 有效的加快了检测速度并更好的进行特征融合。

图 3 右半部分为 Swin-T 模块的两种结构, Swin-T 基于窗口进行自注意力计算以降低序列长度过长产生的庞大计算量, 自注意力公式如下:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = SoftMax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (3)$$

式中: Q, K, V 为输入数据线性变换生成的 3 个矩阵; d 与 QK 相关, 作用是为了在训练过程中保持梯度值稳定。MSA 将自注意力分窗口应用到特征图, 很大程度减少了计算量。

$$\Omega(MSA) = 4\hbar\omega c^2 + 2(\hbar\omega)^2 c \quad (4)$$

$$\Omega(W - MSA) = 4h\omega c^2 + 2M^2 h\omega c \quad (5)$$

式中:变量 h 、 w 、 c 分别表示特征图的高度、宽度和深度; M 则是每个窗口的大小。式(4)为多自注意力机制并引出式(5)窗口多头自注意力机制,均由式(3)推出。

但由于窗口之间检测信息未交互,又引入 SW-MSA 对窗口进行偏移,重新组合 W-MSA 分块的窗口,有效传递了信息。多层感知机 (multilayer perceptron, MLP) 和层级归一化 (layer normalization, LN) 与 Vision-Transformer 结构一致,目的是为了解决非线性问题和数据归一化。经过 Swin-T 模块不会改变数据的大小,只会增加特征序列的权值。

$$\hat{z}^l = W - MSA(LN(z^{l-1})) + z^{l-1} \quad (6)$$

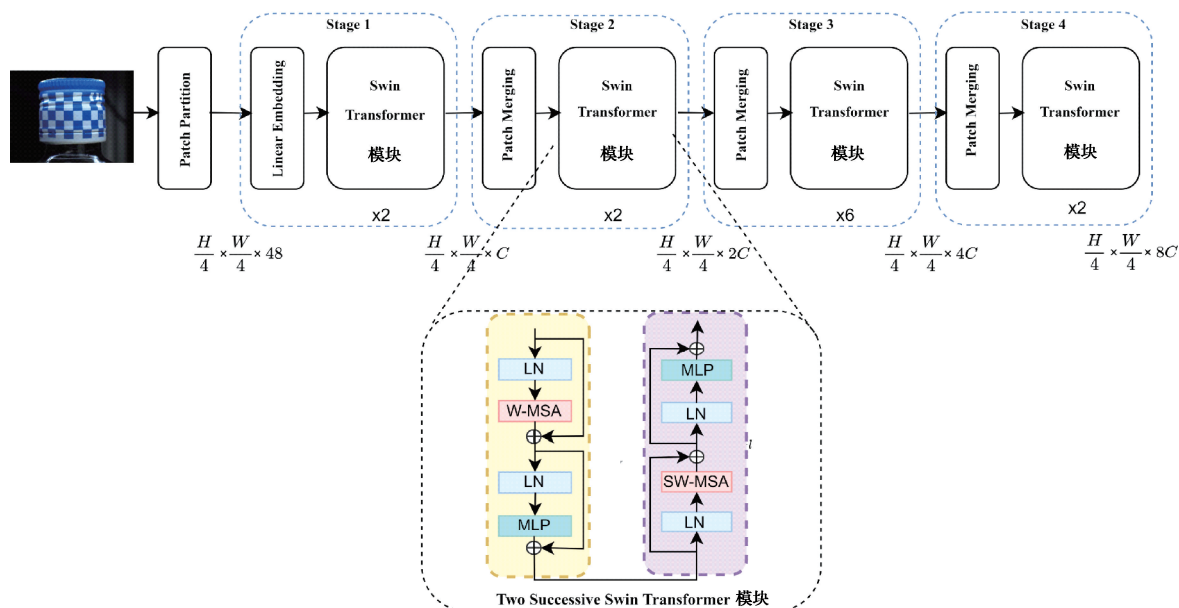


图3 Swin Transformer 结构

$$z' = MLPLN(\tilde{z}') + \tilde{z}' \quad (7)$$

$$\tilde{z}^{l+1} = SW-MSA(LN(z')) + z' \quad (8)$$

$$z^{l+1} = MLPLN(\tilde{z}^{l+1}) + \tilde{z}^{l+1} \quad (9)$$

式中： $\tilde{z}'$ 和 z' 分别表示第 l 块 SW-MSA 和 MLP 模块的输出特征；W-MSA 和 SW-MSA 分别表示采用有规律且可移位窗口划分配置的基于窗口的多头自注意力机制。

2) 引入神经架构搜索 FPN

NAS-FPN 将神经架构搜索^[19]技术应用到了 FPN 上,神经网络搜索架构在训练过程中使用强化学习选择了最佳的模型结构。NAS-FPN 算法流程与 NAS 技术相同,由控制器和评估器两个模块轮流交替,控制器是一个循环神经网络,根据评估器结果更新参数生成新的网络模式,评估器再对控制器生成网络进行效果评估,如此循环往复,两个结构最终生成一个高效能的网络。NAS 在给定的搜索空间使用上述算法对 FPN 重复训练,根据重复次数能够衡量出最佳的精度和速度。图 4 所示为 FPN 和 NAS-FPN 的网络架构,FPN 网络对 backbone 提取的不同尺度特征图进行采样融合,首先将最上层低分辨率的特征图经过上采样后与相邻高分辨率的特征图相融,以此类

推,然后在每个融合后的新特征图进行 bounding box 回归及类别预测。NAS-FPN 则对网络预测出的不同尺度特征图使用自动架构搜索网络,选出新的重组方案,灵活地获取了更优的 FPN 结构。能够在预测同一瓶盖的不同角度,不同大小,不同类型的缺陷时,跨尺度的组合出适合的特征层,提升精度。

NAS-FPN 中 RNN 控制器会对待融合的特征图和融合后的特征图进行采样,这个过程称为合并单元(merging cell, MC),也是网络的核心部分,根据合并单元得出新的不同比例的特征图,然后用于检测目标最后的类别和位置预测,MC 步骤如图 5 所示。(1)从候选列表选取两个特征图作为输入特征层;(2)选择特征层输出的分辨率,由于计算量的考虑,步长为 8 的特征层不加入采样空间;(3)运用二进制运算,将步骤(1)所选特征层融合后使用步骤(2)分辨率输出。其中,步骤(3)的二进制融合有方法,一种是直接 sum 运算,另一种是全局注意力池化(global pooling, GP),流程如图 6 所示。GP 主要作用是将高层包含的语义信息用作浅层的特征图权值来计算,高尺度特征全局池化后通过 1×1 卷积与低尺度特征相乘再相加然后输出,这样高层特征给低级别的特征作了重要指

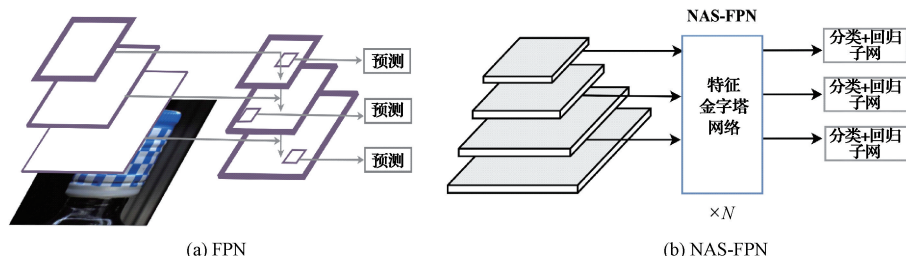


图4 FPN 结构和 NAS-FPN 结构

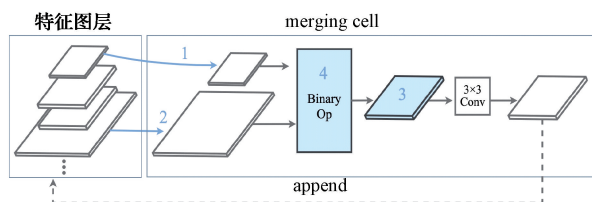


图5 Merging Cell

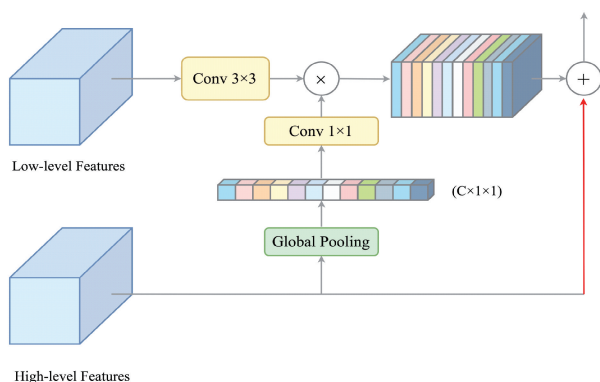


图6 全局注意力池化

导,更紧密的融合了特征。

3) Soft-NMS

非极大值抑制(NMS)一直是许多目标检测算法的组成部分,算法对模型预测后选出得分最大的建议框,将其他与最大建议框重叠明显的框丢弃,且过程不断递归实现。本文酒瓶瓶盖检测中,各种类别的真实框都存在重叠与包含部分,而 NMS 算法对重叠物体检测中,只保留分数较大的框体,很大程度上会误删目标框。为了防止真实检测物体被漏检,本文算法选用 Soft-NMS 对预测框进行抑制。

Soft-NMS 不是直接将交并比(intersection over union, IOU)大于阈值的框都去掉,而是降低置信度来抑制,算法在 NMS 的基础上设置 M 为得分最高框, b_i 为待处理框,计算出两者之间的 IOU 后,通过权重函数,与 M 重叠程度越高的检测框分数越衰减,根据衰减程度得出的分数为 s_i ,最后将 s_i 小于阈值的框丢弃。针对重叠度的打分而不是统一置 0 让 Soft-NMS 剔除多余框的同时保留了物体框,有效的提升了检测精度。由于线性加权的连续可能出现分数断层,Soft-NMS 也可以更改为高斯加权。具体公式如下:

$$s_i = \begin{cases} s_i, & IOU(M, b_i) < N_t \\ s_i(1 - IOU(M, b_i)), & IOU(M, b_i) \geq N_t \end{cases} \quad (10)$$

$$s_i = s_i e^{\frac{-IOU(M, b_i)^2}{\sigma}}, \quad \forall b_i \quad (11)$$

式中: s_i 为各建议框的得分; b_i 为检测到的建议框并等待处理; N_t 为手动设置的阈值; M 为当前得分最高框。

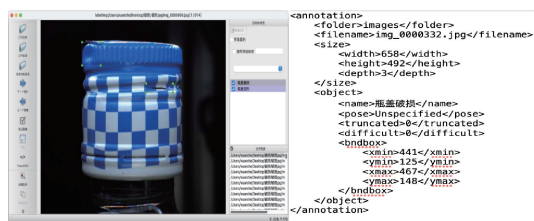
2 结果与分析

2.1 数据集预处理

数据集来自阿里云天池官网,是天池又上工业视觉检测算法大赛提供的瓶装白酒疵品质检数据集,其中共包含 4 105 张瑕疵瓶盖图像,图像大小为 658×492 。

1) 标签制作

首先使用目标检测开源标注工具 LabelImg 将图像制作为标准的 VOC 数据集格式,根据人工标注的锚框,生成带有文件名、图像尺寸、object 和 bndbox 的 xml 文件,其中 object 包含瑕疵类别名称,识别是否难易, bndbox 包含锚框的位置信息,左下角和右上角坐标和 x, y 的长度 ($x_{\min}, y_{\min}, x_{\max}, y_{\max}$) 可以确定一个框体位置。制作过程如图 7 所示。



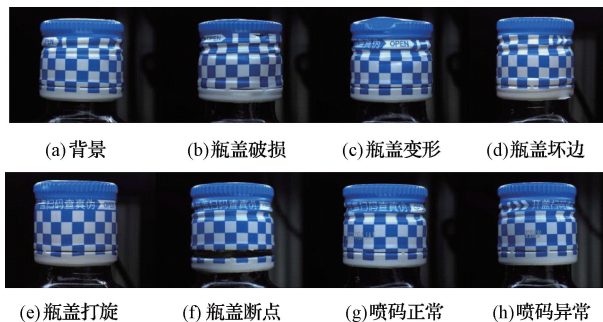
(a) LabelImg 标注

(b) VOC 数据集

图7 LabelImg 标注和 VOC 格式

2) 白酒瓶盖瑕疵分类

制作过程中根据瓶盖瑕疵出现的特点大小位置,将除背景外瑕疵分为瓶盖破损、瓶盖变形、瓶盖坏边、瓶盖打旋、瓶盖断点、正常喷码、异常喷码 7 个类别,如图 8 所示。



(a) 背景

(b) 瓶盖破损

(c) 瓶盖变形

(d) 瓶盖坏边

(e) 瓶盖打旋

(f) 瓶盖断点

(g) 喷码正常

(h) 喷码异常

图8 瑕疵类别

3) 数据集划分

合理划分数据集可以提升训练效率,在模型构建过程中需要检验模型的状态表现以及训练程度是处于过拟合还是欠拟合。所以在本实验中按照 8:1:1 的比例将数据集划分 1 训练集、验证集和测试集。同时为了扩充数据集使用了数据翻转、剪裁、移位、缩放、变色等方法。扩充后数据集共 6 568 张,其中训练集占 5 272 张,验证集 640 张,测试集 656 张,具体划分如表 1 所示。

表1 类别划分

标签类别	背景	瓶盖破损	瓶盖变形	瓶盖怀边
序号	0	1	2	3
训练集	913	1 270	670	1 278
验证集	452	554	370	426
测试集	453	526	366	410
标签类别	瓶盖打旋	瓶盖断点	喷码异常	喷码正常
序号	4	5	6	7
训练集	446	587	336	501
验证集	231	264	225	240
测试集	244	256	217	246

2.2 实验设备

实验使用 Windows 10, 64 位操作系统, mmdetection 框架进行训练, 显卡为 GETFORCE RTX 3060, 处理器为 i5-12400F, cuda 环境, 使用 Python 进行编程。具体运行状态如表 2 所示。

表2 实验环境

类别	参数
系统	Windows 10
GPU	GETFORCE RTX 3060
Cuda	11.1
开发软件	Pycharm
框架	mmdetection
Python	3.8

2.3 实验评价指标

实验主要以 IOU 测试出样的精确度, 平均准确率 (average precision, AP)、平均精度均值 (mean average precision, mAP)、召回率 R 作为评价指标。

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

$$AP = \int P(R) dR \quad (14)$$

$$mAP = \frac{\sum AP}{n} \quad (15)$$

式中: TP 表示正样本预测为正; TN 表示负样本预测为负; FP 表示负样本预测为正; FN 表示正样本预测为负; n 为被检测样本类别数, 本文类别为 7。

2.4 实验过程

实验过程使用 RetinaNet 在 PASCAL VOC 2007 的训练权重进行迁移学习, 对数据集在训练中设置了随机翻转, 增强亮度等数据增强策略, 优化器算法选择 AdamW, 是在 Adam 基础上加入 L2 正则化的改进算法。对输入图片使用多尺度训练 (multi scale training, MST) 处理, MST 在训练时设置多种不同的输入尺寸, 可以有效的提

升多尺度的物体检测, 提升了网络的鲁棒性, 本文模型输入尺度为 (1 333, 640) 和 (1 333, 800)。batch_size 受图片尺寸影响设置为 6, 初始学习率为 0.000 6, 针对学习率使用了 Warmup 优化算法, 也称为学习率预热方法, Warmup 在训练开始会使用一个较小的学习率训练一定的 epoches 后逐步增大为预先设置的数值, 之后又将学习率从优化中设置的预设值线性降低到 0, 便于模型的稳定训练。Epoch 次数设为 60 轮。训练结果如图 9 所示, 对比原网络和本文改进网络 loss 训练图可以看出, 改进后网络的损失曲线函数波动小且收敛快, 证明了本文改进模型提高训练速度和收敛的有效性。

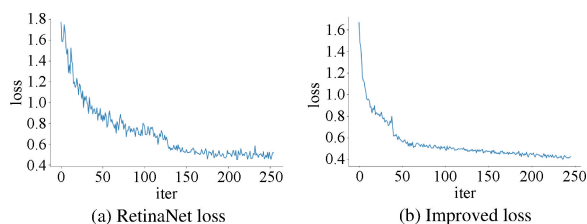


图9 RetinaNet loss 和改进网络 loss

2.5 各类别瑕疵检测精度及总体检测效果

图 10 所示为各类缺陷的具体精度值, 瓶盖打旋、喷码异常及喷码正常由于目标大, 特征清晰与瓶盖背景对比明显, 检测效果较高。而瓶盖断点后几个瑕疵, 检测目标小且分布不均, 瓶盖破损、瓶盖怀边、瓶盖破损 3 个类别, 在图片中的目标不仅数量多且大小不一, 特征信息复杂, 检测效果相对低一些。瓶盖断点虽然也是小目标检测, 但瑕疵位置集中在一处, 瑕疵大小基本统一, 目标之间的覆盖性也少, 检测精度表现较好。改进网络主要提升了小目标和重叠度较高类别的检测精度, 使各类瑕疵检测精度差距缩小, 网络整体提升效果较好, 满足酿酒工厂生产线对缺陷检测准确及速度的要求。

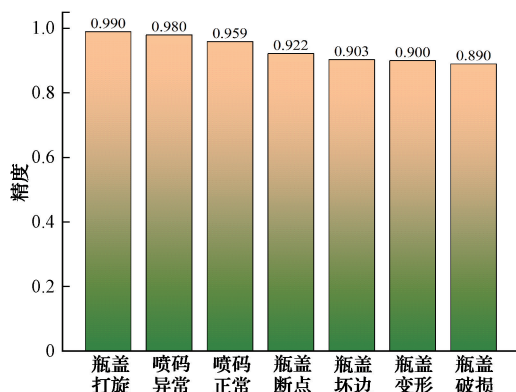


图10 各类别检测效果

改进后的算法检测效果如图 11、12 所示, 检测位置更具体, 难分辨的漏检目标也精准的给出预测框。检测优势明显, 模型整体检测精度提高。

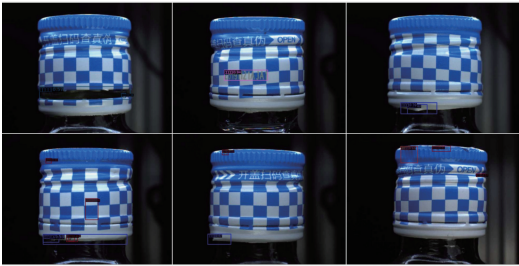


图 11 RetinaNet 检测效果

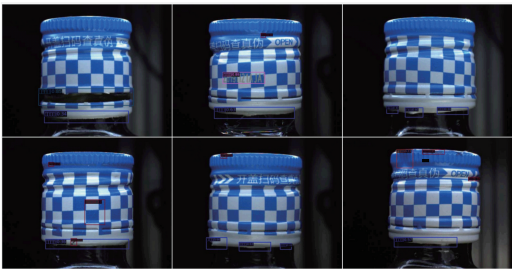


图 12 本文算法检测效果

2.6 实验结果对比

为了验证改进后的 RetinaNet 网络对酒瓶瓶盖缺陷检测的效果,实验分别训练了原始网络,网络使用 Swin Transform 为 backbone 后,更换神经架构搜索 FPN 为 neck 后及使用 Soft-NMS 抑制预测框后的网络在测试集上表现效果,并进行了对比,结果如表 3 所示。

表 3 消融实验

模型	mAP/%	Memory
RetinaNet	85.51	2 466
Swin-T	89.47	3 138
Swin-T+Soft-NMS	91.24	3 138
Swin-T+Soft-NMS+NAS-FPN	93.53	3 510

在 RetinaNet 中使用 Swin Transform 后 mAP 提升了 3.9%,引入 Soft-NMS 后 mAP 提升了 1.77%,最后添加 NAS-FPN 精度达到了 93.53%,提升了 2.29%,总精度提升为 8.02%。根据实验结果可知,每个模块都有不同程度的进步,精度对比如图 13 所示,充分说明本文改进网络对瓶盖缺陷检测数据集识别效果的升高及各增加模块的有效性。

同时为了验证本文改进算法的有效性,在同一数据集下,实验中训练了 Faster-RCNN、Cascade-RCNN、Mask-RCNN、YOLOv3、SSD 与本文算法作对比实验,实验结果如表 4 所示。

实验证明,本文改进网络的精度相比 Faster-RCNN,提升了 3.59%,检测时间缩小 90 ms;比 Cascade-RCNN 精度提升 2.9%,检测时间缩小 110 ms;比 Mask-RCNN 精度提升了 4.77%,检测时间缩小 60 ms;对比同类型的

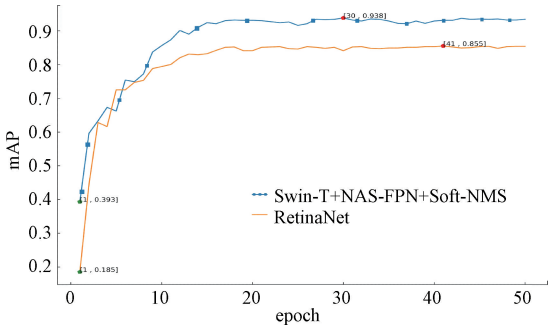


图 13 mAP 对比

表 4 对比试验

模型	mAP/%	检测时间/ms
Faster-RCNN	89.94	280
Cascade-RCNN	90.63	300
Mask-RCNN	88.76	250
YOLOv3	87.23	120
SSD	85.41	150
本文改进网络	93.53	190

单阶段网络 YOLOv3 本文精度提升了 6.3%,检测时间增加了 70 ms;比起 SSD 算法,本文精度提升了 8.12%,检测时间增加 40 ms。由数据可知,本文改进后的模型,算法精度对比双阶段和单阶段网络都效果更佳,且同时保持了单阶段检测网络在检测速度方面的优势。对比同类型单阶段网络精度显著提升,改善较大,检测速度增加幅度低,对比双阶段网络精度提升明显,检测速度差距大。综上所述,本文算法在白酒瓶盖缺陷检测方面优于其他网络,满足工厂包装对白酒瓶盖检测精度及速度方面的需求。

3 结 论

白酒瓶盖瑕疵数据集缺陷尺寸不一,差异大,类别中难检测图像分布不均,本文使用 RetinaNet 模型处理数据集,不仅使模型聚焦到不易检测特征中,解决数据不平衡问题,同时引入 Swin Transform 网络,使模型学习跨窗口信息并交互,关注到特征全局及局部的信息,网络所包含的自注意力机制使检测更加聚焦,NAS-FPN 在传统特征金字塔的基础上选取最佳特征融合层,Soft-NMS 在不增加模型计算量的基础上有效防止瑕疵目标漏检,采用 Pascal-Voc 格式的数据集及 coco 评价指标遵循目标检测算法通用的检测方法,精度及数据的鲁棒性都优于原网络,算法可行有效。通过对比实验,本文模型在速度和精确度整体效果都优于其他网络,对于工厂现代化高效准确检测需求具有重大的现实意义。

目前改进的网络只针对了瓶盖缺陷检测取得了较好的检测效果,且检测背景较为单一,后续考虑会增加检测目标融入不同的特征及对 neck 方面加入适合的注意力机制,使算法目标性更强,满足应对复杂背景也能准确识别

的网络,同时对检测速度进行优化,使网络轻量化,追求适用范围更广,更有效率的网络模型。

参考文献

- [1] 谢永文,肖曙光,谢永俊,等.论传统白酒酿造向机械智能化酿造设备成套生产线转型的必要性[J].酿酒科技,2014(1):63-64.
- [2] 任芳.白酒包材物流自动化升级与模式创新[J].物流技术与应用,2022,27(3):78-80.
- [3] 郑云峰.基于视觉的PET瓶封装质量自动检测系统的设计与实现[D].北京:清华大学,2011.
- [4] 徐锋,吴寅,刘云飞,等.阔叶材原木声信号特征提取及内部缺陷检测[J].仪器仪表学报,2022,43(10):205-214.
- [5] 单明陶,高玮玮,杨亦乐,等.基于改进Faster R-CNN的刹车零件缺陷检测算法[J].国外电子测量技术,2022,41(4):22-28.
- [6] 杭芹,程成,袁彪,等.基于YOLOv5-EA-FPNs的芯片缺陷检测方法研究[J/OL].电子测量与仪器学报:1-10[2023-03-30].<http://kns.cnki.net/kcms/detail/11.2488.TN.20230222.0912.008.html>.
- [7] 段禄成,谭保华,余星雨.基于改进YOLOv3的酒瓶盖瑕疵检测算法[J].电子测量技术,2022,45(15):130-137.
- [8] BIRADARA M, SHIPARAMATTIA B, PATIL B. Fabric defect detection using deep convolutional neural network[J]. Optical Memory and Neural Networks, 2021(30):250-256.
- [9] KRUMMENACHER G, CHENG S O, KOLLER S, et al. Wheel defect detection with machine learning[C]. IEEE Transactions on Intelligent Transportation Systems, 2018: 1176-1187.
- [10] RUDOLPH M, WANDT B, ROSENHAHN B. Same same but DifferNet; Semi-supervised defect detection with normalizing flows[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021:1907-1916.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.
- [12] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Neural Information Processing Systems, 2017, 39(6):1137-1149.
- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [14] LIU W, ANGUELOV D, ERHAN D. SSD: Single shot MultiBox detector[C]. European Conference on Computer Vision, 2016:21-37.
- [15] LIN T, GOYAL P, GIRSHICK R, HE K M, et al. Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision. IEEE, 2017:2999-3007.
- [16] LIU S Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:10012-10022.
- [17] GHIASI G, LIN T Y, PANG R, et al. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:7036-7045.
- [18] BODLA N, SINGH B, CHELLAPPA R. Soft-NMS improving object detection with one line of code[C]. 2017 IEEE International Conference on Computer Vision, 2017: 5562-5570.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017:2980-2988.

作者简介

车璇,硕士研究生,主要研究方向为计算机视觉与图像处理。

E-mail:2223186297@qq.com