

DOI:10.19651/j.cnki.emt.2209215

基于改进 YOLOv3 的酒瓶盖瑕疵检测算法*

段禄成¹ 谭保华¹ 余星雨²

(1. 湖北工业大学理学院 武汉 430068; 2. 湖北工业大学电气与电子工程学院 武汉 430068)

摘要: 在智能酿酒工艺中,对酒瓶外包装进行瑕疵检测是质检环节重要的一环。本文基于改进 YOLOv3 目标检测算法,将其应用到酒瓶盖瑕疵检测的环节中,最终结果符合工厂生产线对瑕疵检测精度和速度的要求。该方法在 YOLOv3 主干 Backbone 网络的残差模块中引入 SENet Module,应用注意力机制加强对特征的提取,在 Neck 特征金字塔网络中引入自适应特征融合网络(ASFF),融合不同尺度的特征信息,提高模型的预测能力,同时引入 Focus Loss 损失函数解决正负样本不均衡问题,加速损失函数的收敛速度。改进后的 YOLOv3-ASFL 在自制酒瓶盖瑕疵数据集上 mAP 达到 92.33%,单张图像检测时间仅为 0.085 s,比原始 YOLOv3 在相同数据集上的 mAP 提升了 6.59%。改进后的 YOLOv3 模型性能更好,符合酒瓶包装生产线对瑕疵检测的需求。

关键词: 瑕疵检测;自适应特征融合;YOLOv3;SENet;Focus Loss

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Defect detection for wine bottle caps based on improved YOLOv3

Duan Lucheng¹ Tan Baohua¹ Yu Xingyu²

(1. School of Science, Hubei University of Technology, Wuhan 430068, China;

2. School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China)

Abstract: Defect detection on outer packaging of wine bottles is of importance. A improved YOLOv3 algorithm is proposed to deal with that problem. The final result shows the improved YOLOv3 meets the production line's requirements for accuracy and speed well. First, the SENet module is introduced into YOLOv3 backbone network's residual block, applying the attention mechanism to enhance the feature extraction. Second, the adaptive feature fusion network (ASFF) is introduced into the feature pyramid network to fuse the feature information in different scales, which enhance the predictive ability of the model. Third, the focus loss function is used to solve the problem of unbalanced positive and negative samples, which will help accelerate the convergence speed of the loss function. The improved YOLOv3-ASFL achieves a mAP up to 92.33% in the self-made wine bottle cap dataset, which is 6.59% higher than the original YOLOv3, and the single image detection time is only 0.085 s. The improved YOLOv3 model has a better performance and meets the needs of the wine bottle packaging production line for defect detection.

Keywords: defect detection; ASFF; YOLOv3; SENet; Focus Loss

0 引言

白酒产业是实体经济的重要组成部分,截止 2021 年全国白酒产业总营收达 6 000 亿元人民币的规模,其营业收入和利润占全国工业行业的比例分别为 0.56%、2.26%^[1],而且这个规模还在继续地扩大。但是传统的白酒企业和白酒生产工厂自动化、智能化水平不高,生产工艺还停留在初级的自动化水平,因此白酒企业对其酿造和包装生产线进行

自动化、智能化、数字化改造有着迫切需要^[2]。

随着自动化技术的应用,生产线上的产品由于受到抖动、校准误差等相关因素的影响,生产出的产品可能出现质量问题,如瓶盖破损、瓶盖断点、瓶盖变形等问题,而这些存在质量问题的瑕疵品如果进入市场不仅会影响消费者的购买欲望、影响产品销售,而且还会影响企业在消费者心目中的形象,不利于企业的持续发展。因此对酒瓶盖进行瑕疵检测是质检环节必不可少的一环,而传统的人工视觉检测模

收稿日期:2022-03-09

* 基金项目:国家教育部大学生创新创业计划(201810500024)、湖北省高等学校哲学社会科学研究重大项目(21ZD054)、湖北工业大学绿色科技引领计划(CPYF2018009)项目资助

式,劳动成本高、工作环境差,而且容易忽略一些细小瑕疵,最终导致成品良率下降。因此越来越多的工厂开始利用机器视觉来代替人工进行质检^[3]。

随着机器视觉的发展,机器视觉已经广泛地运用到了工业生产中,其中在目标检测、产品质检领域有着广泛地应用。近年来,基于深度学习的方法在视觉系统、语音检测和文本分析等领域中取得了长足的发展。在视觉领域,相比于传统的人工提取特征,深度学习技术能够利用多层卷积神经网络对图像进行自动特征提取,进而获得图像的局部信息和深度信息。深度学习自从2011年 AlexNet^[4]提出后,深度学习相关的研究就进入爆发期,原因是多方面的,但简单概括主要是两点:一是神经网络算法的发展,二是GPU算力成本的下降^[5]。目标检测的主要目的是定位并识别图像中单个或多个目标,返回目标位置的同时对其所属类别进行正确分类^[6],目前目标检测算法已经发展成为两个大的分支^[7],一个是两阶段的网络模型,典型的代表是RCNN^[8]、Fast-RCNN^[9]、Faster-RCNN^[10]网络,两阶段网络通过分为特征和分类两个阶段来进行网络的设计,优点是精度高,预测效果好,但是检测速度慢。因此针对这个不足,研究人员又提出了一阶段的网络模型,典型代表是YOLO^[11]、SSD^[12]等系列网络模型,YOLO模型通过把特征和分类集成到一个网络进行,这极大地提高了检测速度,符合实时检测场景的需求,但是由于提高了速度,对于检测精度有所有损失。在实际应用中,薄景文等^[13]通过改进YOLOv3的主干网络,引入注意力机制,将其应用在口罩检测上,在保证检测精度的前提下,加速了模型的检测速度。李晖等^[14]通过改进SSD网络,引入FPN和CReLU,提高了对舰船小目标的识别能力。曹红燕等^[15]结合特征尺度改进YOLOv3,将其应用在红外图像车辆行人检测,取得了不错的效果。

本文通过改进YOLOv3算法来检测酒瓶盖瑕疵,通过在YOLOv3主干backbone网络DarkNet53的残差模块引入SE模块,应用注意力机制加强对特征的提取,在Neck特征金字塔网络中引入自适应特征融合网络(ASFF),融合不同尺度的特征信息,提高模型的预测能力,同时引入Focus Loss损失函数解决正负样本不均衡问题,加速损失函数的收敛速度。

1 酒瓶盖瑕疵检测流程

在酒瓶盖瑕疵检测过程中,第1步需要采集生产线上的酒瓶盖瑕疵图像。采集完成后,需要通过标注软件对图像中的瑕疵进行标注,标注瑕疵类别总共分为7类,分别是:{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},同时根据标注好的标签生成标准VOC数据格式,其中Annotations用来存放标注信息的xml文件,ImageSet用来存放数据集划分的txt文件,JPEGImages用来存放原始的图像。第2步是将图像数据

输入到改进后的YOLOv3网络中进行训练,通过改进backbone主干网络引入SENet Module,然后再输入到自适应特征融合网络(ASFF)中进行训练,训练过程中通过FocusLoss函数平衡正负样本。第3步是在完成模型训练后保存训练好的模型,再通过测试数据对训练完成的模型进行评估,然后将模型部署到实际包装生产线上,实现对生产线上的酒瓶盖进行实时检测,具体流程如图1所示。

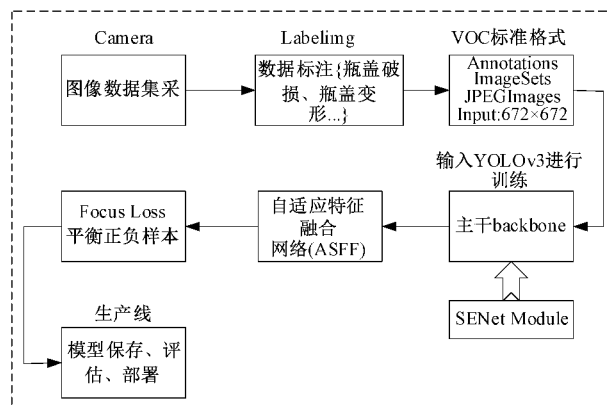


图1 检测流程

2 数据集标注及改进YOLOv3网络模型

2.1 数据集标注

本次使用的数据集是天池公开的企业生产数据集,本数据集总共有酒瓶盖瑕疵图像4105张,图片像素为492×658,如图2所示。

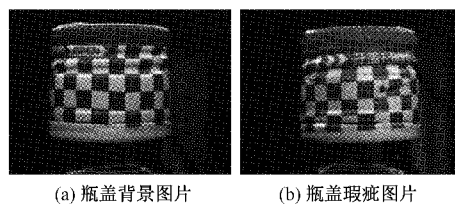


图2 酒瓶盖背景图及瑕疵图

1) 数据集标注

通过开源软件LabelImg对图像进行标注,来制作训练深度学习模型需要的数据集。标注过程图像按照瑕疵类别分为7种类别:{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},通过LabelImg将图像标注成标准的VOC数据格式,其过程如图3所示。同时每张图像标注完成后都会生成一个与之对应的标注文件,一般为xml格式,文件中会包括图片的路径、名称、像素和瑕疵类别以及对应的瑕疵位置信息**bbox**(x, y, w, h),其中 x, y 是起始点坐标, w, h 是宽和高。通过对4105张瓶盖图像数据进行标注后,共标注瑕疵5393处,如表1所示为各类瑕疵数量。

2) 训练测试集的划分

在深度学习模型的训练过程中为了防止模型出现过拟

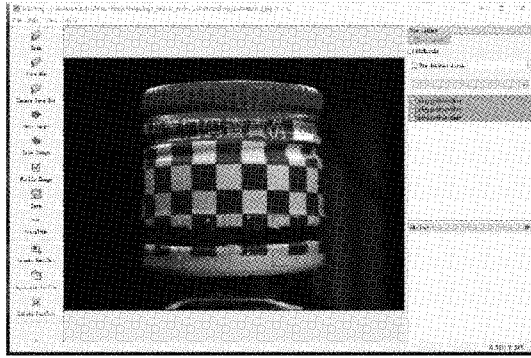


图 3 labeling 数据标注图

表 1 各类瑕疵数量表

类别	背景	瓶盖破损	瓶盖变形	瓶盖坏边
类别号	0	1	2	3
训练集	1 170	1 619	705	1 179
测试集	341	463	264	341
类别	瓶盖打旋	瓶盖断点	异常喷码	正常喷码
类别号	4	5	6	7
训练集	480	641	199	489
测试集	134	159	273	191

合的问题,使模型具有更好的泛化能力,往往需要将数据集拆分为训练集和测试集,利用训练集进行梯度下降训练,再通过测试集对训练完成的模型进行评估。本文采用 9 : 1 的比例划分数据集,选择 3 324 张图片作为训练集,370 张图片作为测试集。

2.2 改进 YOLOv3 网络模型

1) YOLOv3 主干网络 DarkNet53

YOLOv3 网络是由 Redmon 等^[16]提出的一种目标检测算法。如表 2 所示, YOLOv3 的 Backbone 网络是 Darknet53,借鉴了 Resnet^[17]的思想,引入了残差模块,残差模块可以很好地解决由于卷积层加深而导致的网络退化和梯度消失的问题,因此 YOLOv3 可以采用更深的骨干网络 DarkNet53 来提取图像。由表 2 中可以看出,整个网络分为 5 个残差模块,每个残差块中使用一个 3×3 卷积和一个 1×1 卷积,结构十分简单,通过多次残差模块和卷积操作,最终输出大中小 3 种尺度的特征图,分别是 32×32 、 16×16 、 8×8 ,分别用来检测小目标、中目标和大目标,其中取消池化操作,通过卷积核的步长来下采样。

2) 改进残差模块引入 SE Module

SENet 是由 Hu 等^[18]提出的网络结构,主要通过引入注意力机制 SE Module,增加少量参数,获得不同通道上的特征,使网络更关注通道间的特征融合,从而提高检测精度。SE Module 通过压缩(Squeeze)、激励(Exitation)来完成,如图 4 所示,通过 Squeeze 操作 F_{sq} 获得通道响应的全局分布,形成通道描述符,然后通过 Exitation 操作 F_{ex} ,学

表 2 DarkNet53 网络结构表

重复数	过滤器	过滤器数量	尺寸	特征图大小
1 ×	Convolutional	32	3×3	256×256
	Convolutional	64	$3 \times 3/2$	128×128
	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			256×256
2 ×	Convolutional	128	$3 \times 3/2$	64×64
	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	$3 \times 3/2$	32×32
8 ×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	$3 \times 3/2$	16×16
	Convolutional	256	1×1	
8 ×	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1 024	$3 \times 3/2$	8×8
	Convolutional	512	1×1	
	4 ×	Convolutional	1 024	3×3
Residual				8×8
Avgpool			Global	
Connected			1 000	
Softmax				

习对各通道的依赖程度,并根据依赖程度对不同的特征图进行调整,更新输入特征图的权重,输出具有更强语义的特征图。本文考虑将 SE Module 嵌入主干网络中的残差模块(Resblock)中,如图 5 所示,在输入图片进行卷积之后,将输出的结果输入 SE 模块,然后把 SE 的输出结果与原始输入图片进行残差相加作为 DarkNet53 的基本操作模块,如表 2 中的 Resblock 所示。

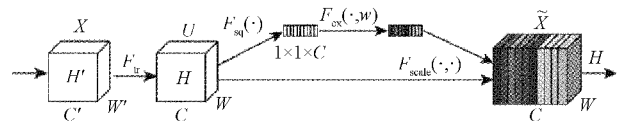


图 4 SE 模块操作图

其中,最重要的是 Squeeze 操作和 Exitation, Squeeze: 通过对空间维度的通道数进行特征压缩,将每个二维的特征通道变成一个实数,这个实数代表着全局的感受野,并且输出的维度和输入的特征通道数相同。它表征着在特征通道上响应的全局分布,而且使得靠近输入的层也可以获得全局的感受野。 Z_c 代表 Squeeze 操作的结果。式(1)代表的是普通卷积操作, V_c 代表第 c 个卷积核, X^s 代表第 s 个输入,最终得到的 U_c 是卷积后的三维矩阵中第 c 个二维矩

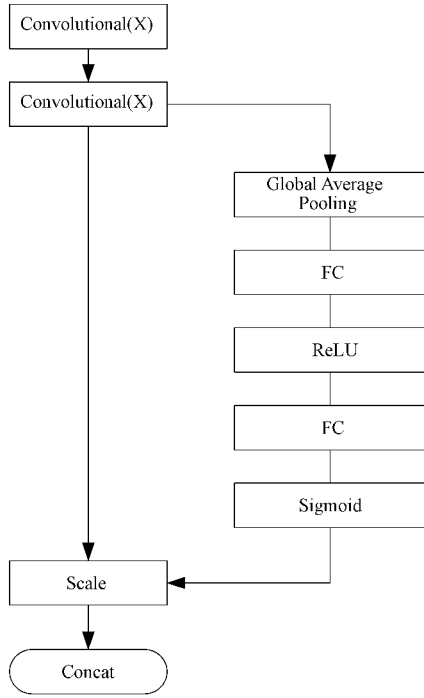


图5 插入SE模块后的残差模块流程

阵,下标 c 表示通道数。

$$U_c = \sum_{s=1}^c V_c^s \cdots X^s \quad (1)$$

式(1)是原网络的卷积操作,式(2)是 Squeeze 操作,

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (2)$$

式(2)将 $H \times W \times C$ 的输入转换为 $1 \times 1 \times C$ 的输出, H 和 W 分别代表图片的高和宽,最后得到的数值 Z_c 表示对应特征图的全局信息。

Excitation:通过 Squeeze 输出的 $1 \times 1 \times C$ 的特征图,再经过两个全连接神经网络,最后用一个类似于循环神经网络中的门控机制。通过参数来为每个特征通道生成权重,其中门控制单元 s 的计算方法如式(3)所示:

$$s = \sigma(g(W_2 \delta(W_1 z))) \quad (3)$$

式(3)中 z 是 Squeeze 操作得到数值结果的结合, W_1 和 W_2 分别是两个全连接层的权值矩阵,维度分别是 $C/r \times C$ 和 $C \times C/r$, r 是缩放参数。 δ 是 ReLU 激活函数, σ 是 Sigmoid 函数, s 表示通过全连接层和非线性层得到的 c 个特征图的权重集合。

3)改进特征金字塔结构引入自适应特征融

YOLOv3 采用类似 FPN^[19] 的特征金字塔结构进行上采样和融合做法,如图 6 所示,特征金字塔网络可以通过自顶向下的路径和横向的方法将语义强、低分辨率的特征与语义弱、高分辨率的特征有效结合,有效地提高模型对小目标的检测精度。YOLOv3 在 backbone 网络输出 3 种尺度的特征图,分别是小尺度、中尺度和大尺度,分别检测大目标、中目标和小目标,通过上采样和特征融合在 3 个不同尺

度分别做独立检测。

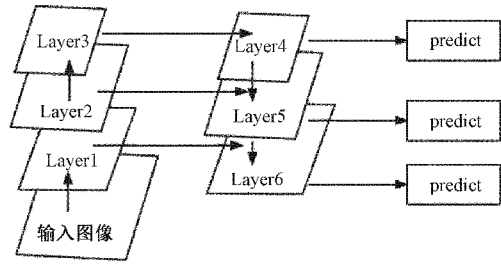


图6 特征金字塔图

由于 YOLOv3 在特征融合时借鉴 DenseNet^[20] 采用 concat 的做法,无法很好地利用浅层网络的语义信息,所以引入自适应特征融合(ASFF)的方法,通过对每一尺度的特征图学习一个融合权重,以便在预测时可以充分地利用不同尺度的特征图的语义信息。自适应特征融合(ASFF)是由 Liu 等^[21] 提出的,其核心的思想是在不同尺度进行特征融合时能自适应地调整权重,如图 7 所示。

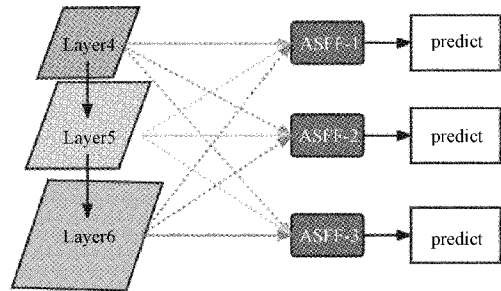


图7 自适应特征融合图

在输出的 3 个特征图后再引入 ASFF,以 ASFF-2 为例, x_1, x_2, x_3 分别来自 layer4、layer5、layer6 的特征矩阵,来自不同尺度的特征再乘以权重参数 α, β 和 γ 求和最终得到该尺度下的输出。如式(4)、(5)所示。

$$y_{ij}^l = \alpha_{ij}^l x_{ij}^{1 \rightarrow l} + \beta_{ij}^l x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l x_{ij}^{3 \rightarrow l} \quad (4)$$

其中, α, β 和 γ 满足:

$$1 = \alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l \quad (5)$$

如图 8 所示为改进后 YOLOv3 的网络结构。

2.3 优化损失函数

YOLOv3 的损失函数包含定位损失、置信度损失以及分类损失,如式(6)所示。

$$L = L_{xy} + L_{wh} + L_{iou} + L_{class} \quad (6)$$

中心点损失函数定义如式(7)所示。

$$L_{xy} = \lambda_{coord} \times \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \quad (7)$$

尺寸损失函数如式(8)所示。

$$L_{wh} = \lambda_{coord} \times \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} (2 - w_i \times h_i) \times [(\omega_i - \hat{\omega}_i)^2 + (h_i - \hat{h}_i)^2] \quad (8)$$

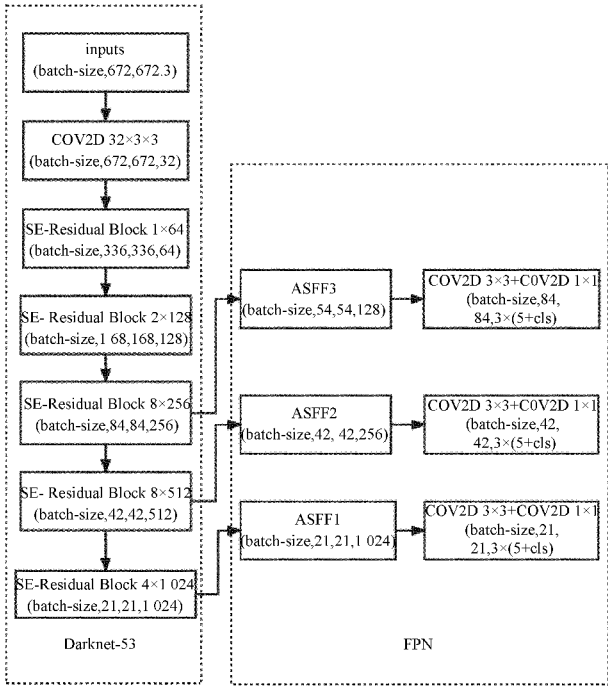


图 8 改进后 YOLOv3 网络结构

置信度损失函数公式如式(9)所示。

$$L = - \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} \times [\hat{C}_i \lg(C_i) + (1 - \hat{C}_i) \lg(1 - C_i)] - \lambda_{noobj} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{noobj} \times [\hat{C}_i \lg(C_i) + (1 - \hat{C}_i) \lg(1 - C_i)] \quad (9)$$

类别损失函数如式(10)所示。

$$L = - \sum_{i=0}^{K \times K} I_{ij}^{obj} \times \sum_{c \in class} [\hat{P}_i(c) \lg(P_i(c)) + (1 - \hat{P}_i(c)) \lg(1 - P_i(c))] \quad (10)$$

在式(7)、(8)中, x, y, w, h 表示预测框的坐标, $\hat{x}, \hat{y}, \hat{w}, \hat{h}$ 表示真实框的坐标, 在式(9)中 $I_{ij}^{obj} = 1$ 表示目标对象出现在网格 j 的边界框 i 中, $I_{ij}^{noobj} = 1$ 表示目标对象未出现在网格 j 的边界框 i 中, λ_{coord} 表示在 $I_{ij}^{obj} = 1$ 时预测框和真实框的坐标差, λ_{obj} 表示出现目标对象边界框的置信度, λ_{noobj} 表示未出现目标对象边界的置信度。

YOLOv3 是典型的基于 Anchor 先验框模式的目标检测算法, 在进行预测的时候会生成大量的先验框, 这会导致产生大量的负样本, 其本身的置信度损失函数不能解决正负样本比例失衡的问题。因此, 本文引入 Focus Loss^[22] 损失函数来控制置信度损失。

Focus Loss 方法通过减少易分类样本权重, 使模型在训练时更关注于难分类的样本, 原理就在标准交叉熵基础上加上权重, 标准交叉熵定义如式(11)所示。

$$F(ce) = -\lg(p_i) \quad (11)$$

Focus Loss 如式(12)所示。

$$FL(p_i) = -\alpha_i (1 - p_i)^\gamma \lg(p_i) \quad (12)$$

其中, α 和 γ 是超参数, p_i 是样本的真实的类别。 p_i 二分法定义如式(13)所示。

$$p_i = \begin{cases} p(x), & y_{true} < 1 \\ 1 - p(x), & \text{其他} \end{cases} \quad (13)$$

改进后, 最终采用的损失函数如式(14)所示。

$$L = L_{xy} + L_{sch} + FL(p_i) + L_{class} \quad (14)$$

3 实验结果与分析

3.1 实验环境

模型训练在 Windows10 操作系统, Pytorch 框架下完成, 处理器是 intel i7-9700, GPU 是 NVIDIA GETFORCE RTX2080TI, 软件环境 CUDA10.2, Python3.7。

3.2 训练过程

利用 YOLOv3 在 COCO 数据集上的预训练模型初始化参数, 可以加快模型的训练速度。共设置 100 个 epoch, 使用冻结训练, 在前 50 个 epoch 冻结主干网络, 初始学习率设置为 0.001 可以收敛得更快, 在后 50 个 epoch 解冻训练, 同时学习率减小为 0.0001, 更新整个参数权重, 总共迭代 4000 次以上。如图 9 所示为改进后的 YOLOv3-ASFL 损失函数曲线, 在训练到 10 个 epoch 时, loss 开始缓慢下降, 这个模型训练耗时在 5 h 左右。

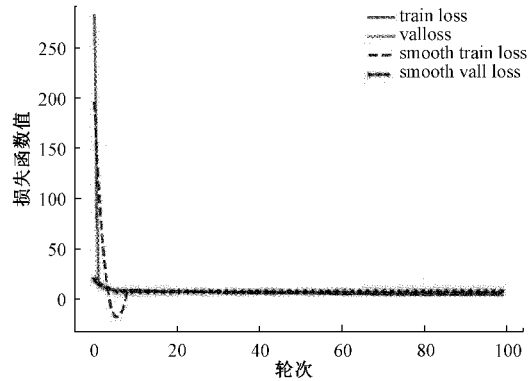


图 9 损失函数曲线图

3.3 评价指标

本文采用 IoU 函数对模型进行评估, 对测试样本计算查准率 (precision)、召回率 (recall), 平均精度 (AP)、平均精度均值 (mAP) 和检测单张图片耗时 t 作为评估指标。其计算公式分别如式(15)~(18)所示。

$$precision = \frac{TP}{TP + FP} \quad (15)$$

$$recall = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \frac{\sum_{i=0}^n P_i}{n} \quad (17)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (18)$$

式中:TP(true positive)表示正确预测的正样本;FN(false negative)表示错误预测的负样本;FP(false positive)表示错误预测的正样本; n 为采样P-R对, k 样本类别数(本文为7)。

3.4 实验结果

1)引入自适应特征融合、SENet和 Focus Loss 损失函数

本文采用IoU函数对模型进行评估,通过100个epoch训练后,原始的YOLOv3和改进后的YOLOv3结果对比如表3所示,在YOLOv3-ASFF中引入自适应特征融合,mAP提升了1.47%,在YOLOv3-SE中引入SE Module,mAP提升了3.60%,在YOLOv3-AS中同时添加SE Module和ASFF,mAP提升了5.82%,对比增加幅度可知,分别引入SE Module和ASFF对网络的性能都有提高,并且同时引入两者对模型性能提高更多,这是因为在主干网络引入SE注意力机制后再在此基础上进一步融合不同尺度的特征信息,两者是递进关系,所以能够进一步提高模型的性能。

表3 原始结构和改进结构对比表

模型	ASFF	SE	mAP/%	t/s
YOLOv3	无	无	85.75	0.056
YOLOv3-ASFF	有	无	87.22	0.067
YOLOv3-SE	无	有	89.35	0.081
YOLOv3-AS	有	有	91.57	0.085

YOLOv3-ASFL在原有的ASFF和SE Module的基础上引入了Focus Loss函数,如表4所示,相比于YOLOv3-AS,mAP提升了0.77%,如图10所示,通过引入3种优化策略,在实际的训练过程中,SE Module、自适应特征融合和Focus Loss都对YOLOv3的检测精度都有不同程度的提升,在检测速度与原来基本相同的情况下,改进后的模型精度更高,其结果符合实际生产线上对瑕疵检测的要求。

表4 优化损失函数对比表

模型	FI	mAP/%	t/s
YOLOv3	无	85.75	0.056
YOLOv3-AS	无	91.57	0.085
YOLOv3-ASFL	有	92.34	0.086

2)检测结果对比及各类瑕疵检测精度

如图11所示,为原本的YOLOv3和改进后的YOLOv3-ASFL对同一张瓶盖缺陷图片的检测的结果对比,通过对比图11(a)和(b)可以看出,在瓶盖变形和瓶盖破损处在原始的YOLOv3存在漏检情况,而改进后的YOLOv-ASFL,通过引入SEnet Module等注意力机制模

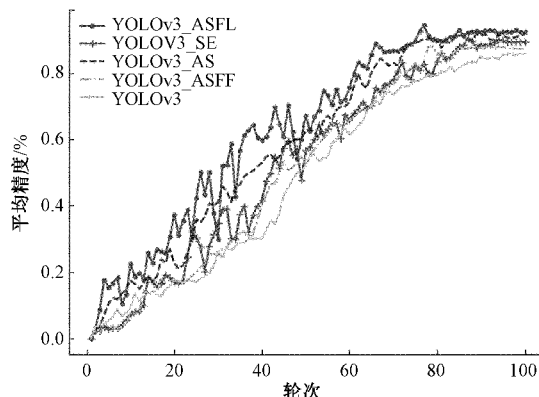
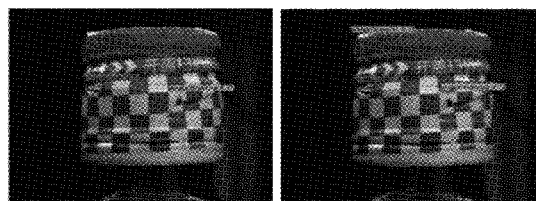


图10 各优化模型 mAP 对比曲线

块,使模型更加关注瑕疵目标,从而在瓶盖变形和瓶盖破损同时出现的情况下,依旧可以识别出来。



(a) 原始YOLOv3预测结果 (b) 改进后YOLO-ASFL预测结果

图11 检测效果对比

如图12所示为各类别瑕疵检测精度,其中,喷码是在固定位置通过喷码机喷射上去,而且是数字信息,所以相对而言容易检测出来,所以检测精度较高。瓶盖断点是瓶盖与瓶盖环之间出现断裂,一旦出现断点,断点处前后背景相差比较大,对比明显,所以瓶盖断点的检测精度也相对比较高。瓶盖打旋存在于瓶盖颈部与头部的交接处,一般是连接不正常的表现,由于其语义信息与背景图片相似,同时在本次使用的数据集中,瓶盖破损、瓶盖变形等其他的瑕疵类别数量比较多,存在类别不均衡问题,所以检测精度稍低。瓶盖坏边多是瓶盖底部的细小损坏,属于小目标检测,小目标一般都比较难检测出来,检测精度相对也低一些。相比

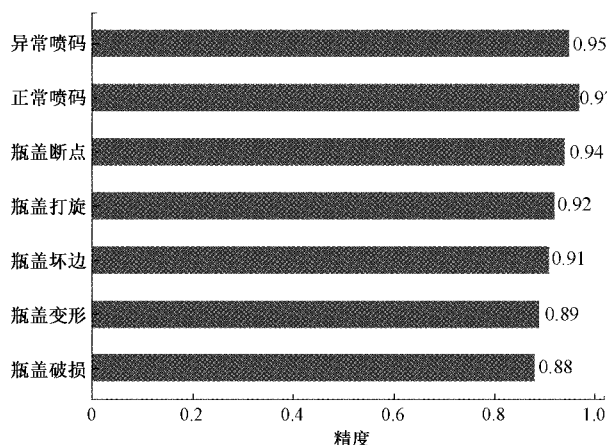


图12 各种瑕疵预测结果直方图

于其他类别的精度,瓶盖破损和瓶盖变形的检测精度要更低一些,这是因为瓶盖破损类间本身存在差异性,既存在大目标也存在小目标,语义信息复杂,且一般瓶盖破损伴有瓶盖变形,语义相互覆盖,不容易分辨出来,所以两者的检测精度都较低。但总体来看,各个类别瑕疵均能检测出来,并且都获得较高的检测精度。

4 结 论

从上文分析可知,YOLOv3 原始的 backbone 网络 Darknet53 引入残差结构,解决了深层卷积网络的网格退化和梯度消失的问题,在此基础上,在残差模块引入注意力机制,使网络更加关注输入图片的特征,再通过输出大中小 3 种不同尺度的特征图进行特征金字塔融合,在特征金字塔中引入自适应特征融合,更好地融合不同尺度的特征信息,实际结果也表明这两者对模型性能提升有很大的作用。在计算损失函数时,由于 YOLOv3 作为典型的生成先验框的检测算法,存在正负样本不均衡的问题,引入 Focus Loss 来平衡样本,能够加速收敛速度同时也能提高模型的检测精度。改进后的 YOLOv3-ASFL 在自制酒瓶盖瑕疵数据集上 mAP 达到 92.33%,单张图像检测时间仅为 0.085 s,比原始 YOLOv3 在相同数据集上的 mAP 提升了 6.59%。改进后的 YOLOv3-ASFL 满足酒厂生产线对酒瓶盖瑕疵的检测需求,对提高酿酒工艺的智能化水平具有指导意义。

虽然改进的 YOLOv3-ASFL 具有较好的性能,但是依旧存在继续优化的空间,正如实验结果小节所说,由于在自制酒瓶盖数据集中,首先数据集类别不均衡,存在过拟合及欠拟合的风险,其次像瓶盖破损和瓶盖变形存在类间差异性问题与语义覆盖问题,加大了检测难度。下一步优化需要针对数据集本身做处理,需要进一步处理类别不均衡问题。同时针对语义覆盖问题,可以尝试共享权重等方法来提高对语义的识别精度。

参考文献

- [1] 曾祥凤,陈一君. 白酒产业高质量发展的动能转换研究[J]. 四川轻化工大学学报, 2021, 36(4): 1-16.
- [2] 杨柳,王霜伶. 三新消费助力后疫情时代白酒产业高质量发展研究[J]. 酿酒科技, 2021(7): 135-142.
- [3] NAGATA F, TOKUNO K, MITARAI K, et al. Defect detection method using deep convolutional neural network, support vector machine and template matching techniques[J]. Artificial Life and Robotics, 2019, 24(4): 512-519.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advance in Neural Information Processing Systems, 2012, 25:1097-1105.
- [5] ZOU Z, SHI Z, GUO Y, et al. Object detection in 20 years: A survey [J]. ArXiv Preprint, 2019, ArXiv:1905.05055.
- [6] JIAO L, ZHANG F, LIU F, et al. A survey of deep learning-based object detection[J]. IEEE Access, 2019, DOI:10.1109/ACCESS.2019.2939201.
- [7] 张培培,王昭,王菲. 基于深度学习的图像目标检测算法研究[J]. 国外电子测量技术, 2020, 39(8): 34-39.
- [8] 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.
- [9] GIRSHICK R. Fast R-CNN[C]. In Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.
- [10] REN S, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2017, 39(6):1137-1149.
- [11] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:7263-7271.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016:21-37.
- [13] 薄景文,张春堂. 基于 YOLOv3 的轻量化口罩佩戴检测算法[J]. 电子测量技术, 2021, 44(23):105-110.
- [14] 李晖晖,周康鹏,韩太初. 基于 CRReLU 和 FPN 改进的 SSD 舰船目标检测[J]. 仪器仪表学报, 2020, 41(4): 183-190.
- [15] 曹红燕,沈小林,刘长明,等. 改进的 YOLOv3 的红外目标检测算法[J]. 电子测量与仪器学报, 2020, 34(8): 188-194.
- [16] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.
- [18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [19] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings

- of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [20] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [21] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection [J]. ArXiv Preprint, 2019, ArXiv:1911.09516.
- [22] 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

段禄成,硕士研究生,主要研究方向为深度学习、电子信息技

术。
E-mail:1277729218@qq.com

谭保华(通信作者),博士,博士生导师,教授,主要研究方

向为电子技术、测控技术、计算机应用。
E-mail:tan_bh@126.com

余星雨,本科,主要研究方向为缺陷检测、电子信息技术。
E-mail:1024049652@qq.com