

DOI:10.19651/j.cnki.emt.2210450

基于改进 YOLOv4 算法的玻璃杯缺陷识别方法研究^{*}张涛¹ 谢探阳¹ 李玉梅¹ 白俊华²

(1.北京信息科技大学高动态导航技术北京市重点实验室 北京 100101; 2.沧州格锐特钻头有限公司 沧州 062450)

摘要: 当前,玻璃杯表面缺陷检测主要依赖人力劳动来完成,存在耗时长且准确率不高等问题。提出了一种将 YOLOv4 与 MobileNetV3 结合的改进算法模型 YOLO-M 来解决该问题。首先,利用 MobileNetv3 网络替换 YOLOv4 原本的主干网络 CSPDarknet53,并修改激活函数,在减少模型大小和参数数量的基础上提升运行速度。然后,对玻璃杯缺陷样本进行拍照采样,将缺陷分为磨损、气泡、划痕三种,建立玻璃杯缺陷数据集。最后利用 YOLO-M、YOLOv4 以及 YOLOv4-tiny 三种算法对玻璃杯缺陷数据集进行训练,将不同算法下的平均精度均值、帧率等评价指标进行对比。实验结果表明,YOLO-M 算法在玻璃杯缺陷检测上的帧率达到 57.72 f/s,平均精度均值达到 91.95%,均为最高。YOLO-M 算法在玻璃杯缺陷识别的速度和精度上有明显效果,可作为后续分拣研究,以及其他玻璃制品缺陷识别的重要参考。

关键词: 深度学习;目标检测;深度可分离卷积;瓶颈残差结构

中图分类号: TP399 **文献标识码:** A **国家标准学科分类代码:** 520.6

Research on glass defect recognition method based on improved YOLOv4

Zhang Tao¹ Xie Tanyang¹ Li Yumei¹ Bai Junhua²

(1. Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science & Technology University, Beijing 100101, China; 2. Cangzhou Geruite bit Co., Ltd., Cangzhou 062450, China)

Abstract: At present, glass surface defect detection is mainly manual, which takes a long time and has low accuracy. An improved algorithm model YOLO-M combining YOLOv4 and mobilenetv3 is proposed to solve this problem. Firstly, the mobilenetv3 network is used to replace the original backbone network cspdarknet53 of YOLOv4, and the activation function is modified to improve the running speed on the basis of reducing the model size and parameters. Then, the glass defect samples were photographed and sampled. The defects were divided into wear, bubble and scratch, and the glass defect data set was established. Finally, the glass defect data set is trained by using YOLO-M, YOLOv4 and YOLOv4 tiny algorithms, and the evaluation indexes such as average precision and frame rate under different algorithms are compared. The experimental results show that the frame rate of YOLO-M algorithm in glass defect detection is 57.72 f/s, and the average accuracy is 91.95%. YOLO-M algorithm has obvious effect on the speed and accuracy of glass defect recognition, and can be used as an important reference for subsequent sorting research and other glass product defect recognition.

Keywords: target detection; deep learning; depth separable convolution; bottleneck residual

0 引言

在玻璃制品生产过程中,需要对原料进行高温加热,然后冷却成型。这个过程会使玻璃成品产生气泡、裂缝等缺陷。而目前,国内对于这种缺陷的识别与拣选主要依赖人工,但人工识别存在一定主观性,并且效率低,其准确率和

检测速度都不高。近年来,国内外对于玻璃表面缺陷识别以及 MobileNet 网络的研究越来越多。

张丹丹等^[1]提出了利用深度卷积网络与稀疏自编码器结合来进行玻璃缺陷检测。吉祥等^[2]将梯度因子和动量因子加入到 BP 神经网络算法的运算当中。但由于玻璃制品在特征采集过程中极易受到光源影响,导致识别准确

收稿日期:2022-06-24

^{*} 基金项目:国家自然科学基金青年项目(52104001)、北京信息科技大学重点研究培育项目(2121YJPY220)、北京市教育委员会科学研究计划项目(KM202111232004)、中国石油天然气集团有限公司-中国石油大学(北京)战略合作科技专项(ZLZX2020-03)资助

率欠佳。

元宁宁等^[3]对生产环境进行了模拟,采取一种背光式照明方案以及 CCD 摄像技术来初步采集图像样本。齐榕等^[4]将采集到的玻璃制品数据集进行了二次处理,以便于深度学习算法进行特征提取。

国外一些研究人员针对视觉算法准确性和计算速度进行了优化。Kulkarni 等^[5]提出了一种使用 MobileNet 改进的网络模型。Srinivasu 等^[6]采用基于 MobileNetV2 和长短期记忆神经网络对皮肤病进行分类。Murthy 等^[7]将 SSD 算法的主干特征提取网络替换为 MobileNet,在 Jetson Nano 开发板上运行并进行行人检测。

为了加快算法的运算速度,提升准确率,本文提出了一种将 YOLOv4 与 MobileNetV3 结合的算法模型:YOLO-M。玻璃杯缺陷数据集为自行拍照、标注制作,缺陷类型分为气泡、磨损、划痕 3 种。YOLO-M 算法结合了 MobileNetV3 参数和体量都不大,且运算速度快的优点,最大程度的保留了 YOLOv4 的逻辑结构,保证了预测的精确度。实验部分利用 YOLO-M、YOLOv4 以及 YOLOv4-tiny 三种算法在玻璃杯缺陷数据集上进行训练和测试。在帧率和平均精度均值等不同的评价指标以及不同类型的缺陷下进行检测对比。结果表明,YOLO-M 算法的平均精度均值达到 91.95%,帧率为 57.72 f/s,均为 3 种算法中的最优。

1 YOLOv4 算法原理

1.1 主干特征提取

YOLOv4 的主干特征网络是 CSPDarknet53^[8],其主要工作就是特征层提取计算。其中的残差卷积结构,会将输入进来的特征层的高、宽进行压缩,并把结果和原来的特征层叠加。残差结构数学上的表达为:

$$x_i = H_L(x_{i-1}) + x_{i-1} \quad (1)$$

式中: x_i 表示 i 层的输出; H_L 表示批量归一化、卷积运算等结合在一起的非线性变换体。

1.2 加强特征网络

在 CSPDarknet53 获得有效特征层后,加强特征提取网络会对输入的有效特征层进行融合运算,获得加强特征的特征层。

加强特征提取网的第一部分是空间金字塔结构 SPP^[9]。输入图像通过其中尺寸不同的池化层与全连接层,将会被区别提取出不同特征,以增强检测准确率。

第二部分是 PANet^[10]结构,其作用原理是增加特征融合率,建立底层特征和顶层特征之间的信息路径,来减少底层特征向顶层传递的过程中产生的特征缺失,保留足够的特征。

2 改进 YOLOv4 算法

MobileNet 系列网络是轻量级的深层神经网络,其主

干部分的作用是同样是进行特征提取。本文使用 MobileNetV3 特征提取网络代替 YOLOv4 中的 CSPDarknet53 进行特征提取工作。

2.1 深度可分离卷积

MobileNetV1 的主体部分使用了一种深度可分离卷积运算^[11]。将原本的卷积运算分解为两个更小的运算,如图 1 所示。深度可分离卷积运算量与标准卷积运算量的比较为:

$$\frac{D_A \times D_A \times I \times D_B \times D_B + I \times J \times D_B \times D_B}{D_A \times D_A \times I \times J \times D_B \times D_B} = \frac{1}{J} + \frac{1}{D_A^2} \quad (2)$$

式中:分子第一项为深度卷积运算量,第二项为逐点卷积运算量,分母为标准卷积运算量; D_A 为卷积核尺寸; D_B 为输入图片尺寸; I 为输入通道个数; J 为输出通道个数。

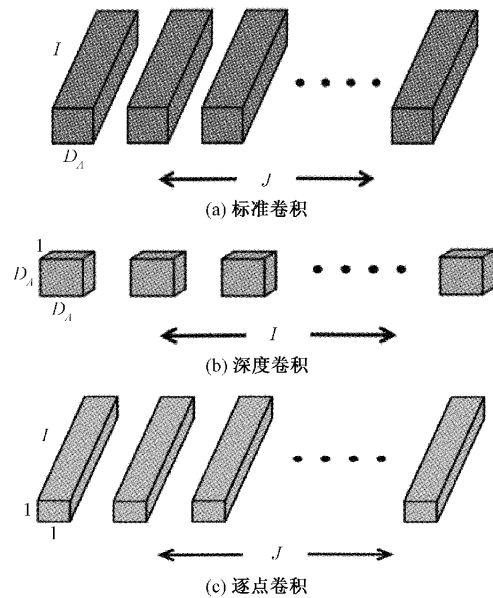


图 1 卷积分解过程

深度可分离卷积首先是采用深度卷积,对不同输入通道分别进行卷积,然后采用逐点卷积将前面的输出进行再结合,减少计算压力。

2.2 瓶颈逆残差结构

MobileNetV2 除了继续使用深度可分离结构之外,还使用了一种瓶颈逆残差结构^[12]。即是在可分离结构两端分别增加了扩展层和投影层。投影层的目的是把高维空间的特征映射到低维空间。扩展层的功能是将低维空间特征映射到高维空间。

瓶颈逆残差结构使用 ReLU6 作为激活函数,数学表达式为:

$$ReLU6 = \min(6, \max(0, x)) \quad (3)$$

在不使用激活函数的情况下,神经网络结构内每一层的输出数据,都是上层输入数据进行线性运算的结果。

瓶颈逆残差结构就是先将低维数据映射到高维,再用深度可分离卷积运算来获得特征,之后压缩数据,让数据维度收缩变小。

2.3 MobileNetV3

MobileNetV3 在沿用了深度可分离卷积和瓶颈逆残差结构的同时,增加了一种轻量级注意力模型:挤压与激励(SE-Net)神经网络^[13]。并对部分卷积层引入了 h-swish^[14] 激活函数。表达如下:

$$h-swish(x) = x \frac{ReLU6(x+3)}{6} \quad (4)$$

输入数据通过 SE-Net 神经网络,首先进行挤压运算,进行特征提取并压缩。然后激励运算会获取各个通道之间

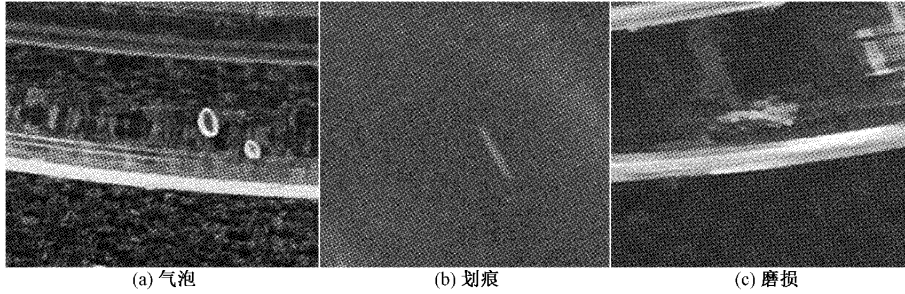


图 2 玻璃杯数据集样本示例

3.2 评价指标

本文使用帧率 (frame per second, FPS)、精确度 (Precision)、召回率 (Recall)、F1 值 (F1-score)、平均精确度 (average precision, AP) 和平均精度均值 (mean average precision, mAP) 作为检测效果的评价指标。平均精确度一定程度上整合了检测中的精确率和召回率,是检测算法中常用的评价指标^[15]。AP 和 mAP 值越高,说明算法对于目标缺陷的检测效果越好。

帧率即每秒检测多少帧的图像,用来衡量检测的速度,不同硬件 FPS 值并不一样。本实验在同一台机器上进行。FPS 计算式为:

$$FPS = \frac{1}{t} \quad (5)$$

式中: t 为检测一帧所需的时间。时间越短, FPS 值就越大,效果越好。

关于其他指标,假定: TP 为正类样本被判定为正类; FP 为负类样本被判定为正类; TN 为负类样本被预测为负类; FN 为正类样本被判定为负类。精确率表达式为:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

召回率表达式为:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

F1 值综合考量了召回率和精确度的情况,表达式为:

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

的权重,增加对信息量大的通道的特征提取程度,对于信息量少且不太重要的通道减少特征提取程度。

3 实验分析

3.1 数据集

使用型号为 LT-USB1080P 的高清工业摄像头拍摄玻璃杯样本图片,自行采集图片并进行标注,制作玻璃杯数据集。标注类别有气泡、划痕、磨损 3 种,原图片样本总计 1 200 张,气泡 346 张,划痕 402 张,磨损 452 张。图 2 为部分样本示例。因图像数据集数量过小会导致训练结果过拟合,所以图像原样本采用了一些数据增强手段,例如翻转、缩放、拼接等,在不破坏原样本的前提下增加数据量。

AP 事实上指的是,利用不同的精确度和召回率的点的组合,绘制出来的曲线的面积。平均精确度表达式为:

$$AP_i = \int_0^1 P(r) dr \quad (9)$$

式中: P 表示精确率; r 表示召回率。

当选取的置信度^[16]足够密集的时候,就可以获得足够的精确度和召回率数值。类别数为 m 的平均精度均值表达式为:

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i \quad (10)$$

式中: mAP 就是所有种类的 AP 值求平均。

3.3 损失函数

本文的损失函数由分类损失 L_{class} 、置信度损失 L_{conf} 和坐标回归损失 L_{coord} 3 个部分共同构成^[17]。计算式如下:

$$L_{coord} = \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{obj} [(x_i - x'_i)^2 + (y_i - y'_i)^2] + \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{obj} [(\sqrt{w_i} - \sqrt{w'_i})^2 + (\sqrt{h_i} - \sqrt{h'_i})^2] \quad (11)$$

$$L_{conf} = \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{obj} (C_i - C'_i)^2 + \lambda_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^B l_{ij}^{noobj} (C_i - C'_i)^2 \quad (12)$$

$$L_{class} = \lambda_{class} \sum_{i=0}^{s^2} l_i^{obj} \sum_{c \in classes} [p_i(c) - p'_i(c)]^2 \quad (13)$$

$$Loss = L_{coord} + L_{conf} + L_{class} \quad (14)$$

式中: x_i, y_i 为预测框中心坐标; w_i, h_i 为预测框宽、高; S

为输入图片网格划分系数; B 为预测框个数; l_{ij}^{obj} 表示第 i 个网格中第 j 个边框对应预测目标时, 其值为 1, 否则为 0; l_{ij}^{noobj} 表示第 i 个网格中第 j 个边框不对应预测目标时, 其值为 1, 否则为 0; classes 为预测类别; p 为预测框类别概率, p' 为真实框类别概率。为了使各项损失比重平衡, 通常设置坐标损失权重 $\lambda_{coord} = 5$, 设置不含目标的置信度损失 $\lambda_{noobj} = 0.5$ 。

3.4 模型训练

本文使用 YOLO-M、YOLOv4、YOLOv4-tiny 三种算法在自制的玻璃杯缺陷数据集上进行训练。每种算法均使用如下设置: 初始学习率为 0.001, 初始动量为 0.9, 训练批次大小为 8, 交并比阈值选择 0.5。首先进行 50 个轮次的预训练, 之后改变学习率为 0.0001, 再训练 50 个轮次。

3.5 结果分析

实验在 Windows10、CUDA10.2 环境下进行。调用 GPU 进行训练, 其主要配置为 NVIDIA1660, 6 G 显存。Python 版本为 3.7, 基于 TensorFlow2.2.0 框架。实验素材即为 3.1 节介绍的自制数据集。

3 种算法的训练损失曲线如图 3 所示。纵坐标为损失值, 无单位, 横坐标为训练轮次。3 种算法在前 40 轮的训练过程中, 损失值不断下降, 虽存在波动, 但下降趋势在不断减缓。YOLO-M 算法最先在 40 轮左右收敛。

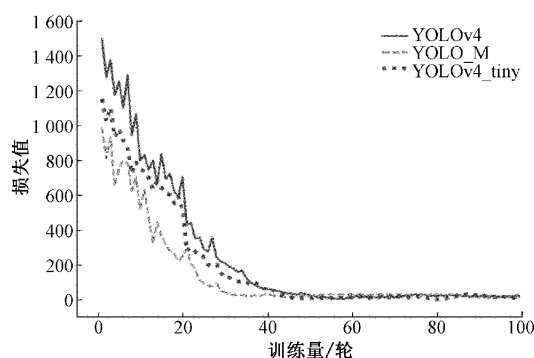


图3 三种算法损失值曲线

这个过程表明模型的真实结果和预测判断结果在不断接近, 预测的误差逐渐下降, 但训练还不够充分, 算法模型还在进行有效学习。60 轮之后, 三条损失曲线都不再发生明显波动, 学习过程可以结束, 无需再进行更多轮次的训练。

如图 4、5 为 3 种算法的精确度曲线与召回率曲线。横坐标为训练轮次, 纵坐标为百分率。随着训练轮次的增加, 三种算法的精确度与召回率在整体趋势上虽然存在波动, 但都在不断增加, 最终在 60 轮之后基本稳定。

橙色虚线所代表的 YOLO-M 算法的精确度、召回率以及 F1 值最终均高于其他两种算法。精确度、召回率以及 F1 值最终结果如表 1 所示。

如图 6 为 3 种算法的 mAP 曲线。横坐标为训练轮次, 纵坐标为百分率。随着训练过程不断加深, 3 种算法的

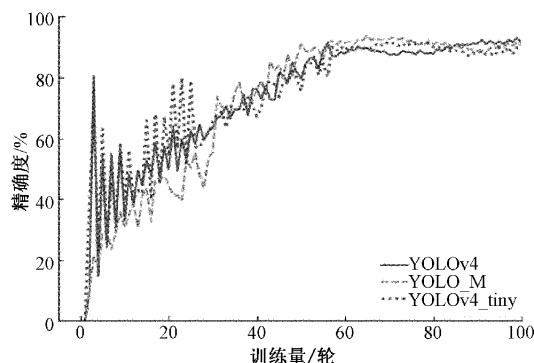


图4 三种算法精确度曲线

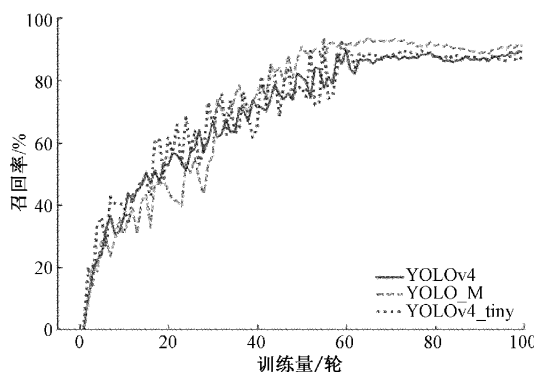


图5 三种算法召回率曲线

表1 部分训练结果指标

算法	精确度 P/ %	召回率 R/ %	F1-score/ %
YOLO-M	92.02	90.19	90.71
YOLOv4	91.48	88.44	89.97
YOLOv4-tiny	89.55	86.32	87.98

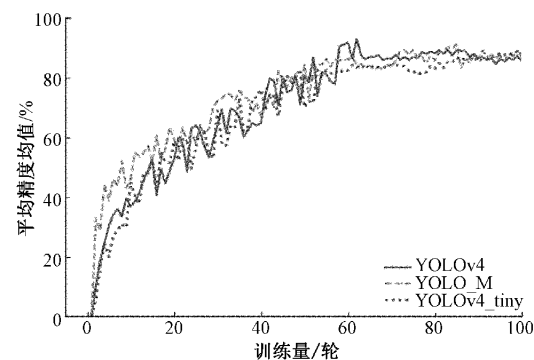


图6 三种算法 mAP 值曲线

mAP 值不断增加, 在 60 轮之后趋于稳定。

如图 7 为将自然光或者反光误识别为划痕的情况。4 张图片中均出现了光线投射在玻璃杯上, 在玻璃表面产生类似于划痕的线条。算法误将这些因光学原因出现的线条判定为划痕。

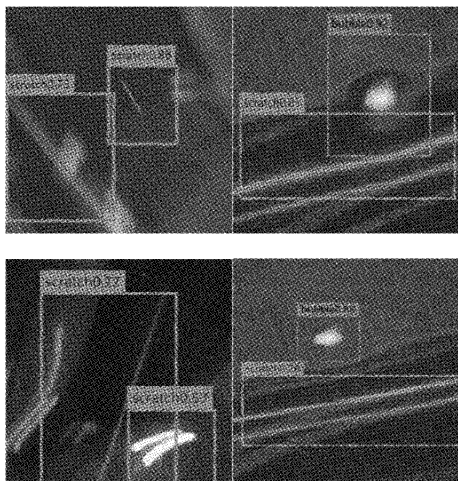


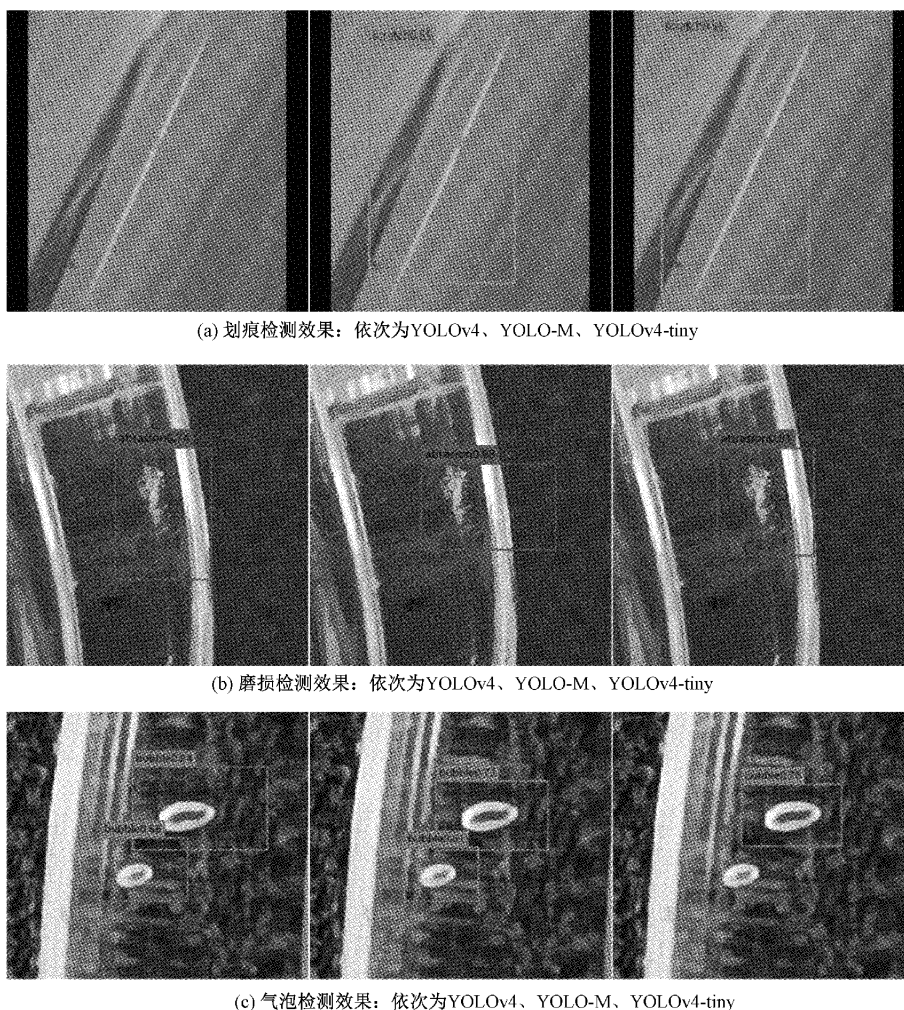
图 7 部分错误识别

在划痕的检测上,3种算法的整体效果都不如磨损和气泡的检测,主要原因是划痕效果受光线影响比较严重,划痕和自然光或者灯光在玻璃上的反射极易混淆,而造成一些错误识别。

YOLO-M算法的划痕检测和气泡检测在3种算法中的效果均为最佳,分别可以达到85.11%和96.38%。图8为3种算法的实际检测效果对比。YOLO-M算法在3种缺陷的检测过程中基本不存在漏检状况。

表2为3种算法对于不同缺陷的平均精确度对比。对于磨损和气泡的检测,3种算法没有较大的精度差别。

表3为3种算法在FPS值、每轮次训练时间与平均精度均值的对比,YOLO-M算法在平均精度均值上达到91.95%,为3种算法中最佳;FPS值为57.72,分别比YOLOv4算法和YOLOv4-tiny算法高出42.84和7.17;



(a) 划痕检测效果: 依次为YOLOv4、YOLO-M、YOLOv4-tiny

(b) 磨损检测效果: 依次为YOLOv4、YOLO-M、YOLOv4-tiny

(c) 气泡检测效果: 依次为YOLOv4、YOLO-M、YOLOv4-tiny

图 8 算法真实检测效果

每轮次训练时间相较YOLOv4算法缩短了50s。

YOLO-M在以上性能对比中结果均为最优,充分证明了此算法在保护检测精度的情况下,可以大幅度提升检

测速率。YOLO-M算法在玻璃杯缺陷识别的速度和精度上均有明显效果,可作为后续分拣研究,以及其他玻璃制品缺陷识别的重要参考。

表2 不同缺陷检测结果对比

算法	AP/% (气泡)	AP/% (划痕)	AP/% (磨损)
YOLO-M	96.38	85.11	94.36
YOLOv4	95.70	81.92	93.91
YOLOv4-tiny	95.91	78.42	95.61

表3 算法主要性能对比

算法	FPS (f/s)	训练总耗时/ s	mAP/ %
YOLO-M	57.72	8 083	91.95
YOLOv4	14.88	13 045	90.51
YOLOv4-tiny	50.55	8 766	89.98

4 结 论

本文将YOLOv4原本的主干特征提取网络替换为MobilenetV3,优化网络结构,提出YOLO-M算法。

建立玻璃杯缺陷数据集,对YOLO-M、YOLOv4、YOLOv4-tiny三种算法进行训练和测试,结果表明YOLO-M算法的平均精度均值达到91.95%,FPS值达到57.72 f/s,均为3种算法中最优。

实验结果表明,YOLO-M算法无论在检测速度还是检测精度上,均具有明显效果。将此算法应用于实际生产环境中,可极大提高整个生产、检测、分拣流程的作业效率。YOLO-M算法对于其他玻璃制品缺陷检测同样有一定的参考意义。

参考文献

- [1] 张丹丹,金永,胡缤予,等.基于集成深度学习的玻璃缺陷识别方法[J].计算机测量与控制,2019,27(2):216-220.
- [2] 吉祥,戴曙光.BP神经网络在玻璃缺陷识别中的应用[J].软件导刊,2019,18(4):137-140,144.
- [3] 亓宁宁,常敏,刘雨翰.基于机器视觉的玻璃缺陷检测[J].光学仪器,2020,42(1):25-31.
- [4] 齐榕,贾瑞生,徐志峰,等.基于YOLOv3的轻量级目标检测网络[J].计算机应用与软件,2020,37(10):208-213.
- [5] KULKARNI U, MEENA S M, GURLAHOSUR S V, et al. Quantization friendly mobileNet(QF-MobileNet) architecture for vision based applications on embedded platforms[J]. Neural Networks, 2021, 136:28-39.
- [6] SRINIVASU P N, SIVASAI J G, IJAZ M F, et al. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM[J]. Sensors, 2021, 21(8): 2852.

- [7] MURTHY C B, HASHMI M F, KESKAR A G. Optimized MobileNet + SSD: A real-time pedestrian detection on a low-end edge device[J]. International Journal of Multimedia Information Retrieval, 2021, 10(3): 171-184.
- [8] 孟浩,田洋,孙宇婷,等.全局注意力关系网络的小样本船舶识别[J].仪器仪表学报,2021,42(12):220-227.
- [9] 廖国庆,吴文海,曾鑫鹏.YOLOv4与ORB深度融合的绝缘子识别定位研究[J].电子测量与仪器学报,2022,36(2):131-138.
- [10] 李庆,龚远强,张玮,等.用于智能垃圾分拣的注意力YOLOv4算法[J].计算机工程与应用,2022,58(11):260-268.
- [11] KAISER L, GOMEZ A N, CHOLLET F. Depthwise separable convolutions for neural machine translation[J]. ArXiv Preprint,2017, ArXiv:1706.03059.
- [12] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [13] 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.
- [14] 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.
- [15] 化嫣然,张卓,龙赛,等.基于改进YOLO算法的遥感图像目标检测[J].电子测量技术,2020,43(24):87-92.
- [16] 陈晓艳,任玉蒙,张东洋,等.基于YOLOv3的轻量化高精度多目标检测模型[J].天津科技大学学报,2021,36(3):33-38.
- [17] 解尧婷,张丕状.基于改进的YOLOv4输电线路小目标检测[J].国外电子测量技术,2021,40(2):47-51.

作者简介

张涛,博士,副教授,主要研究方向为油气井复杂工况数值模拟;油气井智能测量与控制。

E-mail:chonpfl@163.com

谢探阳,硕士研究生,主要研究方向为智能测量与控制。

E-mail:1257587393@qq.com

李玉梅,助理研究员,主要研究方向为复杂环境下运动体多参数测量与控制。

E-mail:liyumei3680238@163.com

白俊华,高级工程师,主要研究方向为石油钻采设备制造。

E-mail:great166@126.com