

基于深度学习的红外目标成像液位检测方法

梁霄, 李家炜, 赵小龙, 臧俊斌, 张志东*, 薛晨阳

中北大学仪器科学与动态测试教育部重点实验室, 山西 太原 030051

摘要 容器液位检测是工业生产及化工原料储存、运输过程中的重要环节, 针对现有液位检测技术中传感器布置容易受空间限制, 在高温高压、灰尘、潮湿等特殊环境下传感器寿命短等问题, 提出了一种基于深度学习的红外目标成像液位检测方法。通过对容器红外图像标注数据集进行优化训练, 得到可以准确识别容器内液体百分比含量的模型。首先, 构建储罐液位标准数据集, 并搭建基于 Pytorch 的深度学习目标检测框架。然后, 在输入端对图像进行数据增强, 调整模型的宽度和深度, 优化训练检测模型。最后, 采用特征金字塔网络和路径聚合网络结构融合不同尺寸特征图的特征信息, 用联合交并比计算边界框的回归损失, 并在后处理过程中引入加权非极大值抑制。实验结果表明, 该模型具有较好的鲁棒性和识别效果, 在交并比为 0.5 时的平均精度均值可达到 0.804。

关键词 图像处理; 液位检测; 红外成像; 深度学习; 目标检测

中图分类号 TN911.73

文献标志码 A

doi: 10.3788/AOS202141.2110001

Infrared Target Imaging Liquid Level Detection Method Based on Deep Learning

Liang Xiao, Li Jiawei, Zhao Xiaolong, Zang Junbin, Zhang Zhidong*, Xue Chenyang

*Key Laboratory of Instrumentation Science & Dynamic Measurement, Ministry of Education,
North University of China, Taiyuan, Shanxi 030051, China*

Abstract The detection of container liquid level is an important link in the process of industrial production, storage and transportation of chemical raw materials. Aiming at the problems that the sensor layout in the existing liquid level detection technology is easily limited by space and the short service life of the sensor in special environments such as high temperature, high pressure, dust and humidity, a method of infrared target imaging liquid level detection based on deep learning is proposed in this paper. Through the optimization training of the infrared image annotation data set of the tank liquid level, the model that can accurately identify the percentage content of liquid in the container is obtained. First, construct a standard data set of tank liquid level and build an image detection framework based on Pytorch's deep learning. Then, enhance the data on the image at the input end, adjust the width and depth of the model, and optimize and train the detection model. Finally, the feature pyramid network and path aggregation network structure are used to fuse the feature information of different size feature maps, the complete intersection over union is used to calculate the regression loss of the bounding box, and the weighted non maximum suppression method is introduced in the post-processing process. The experimental results show that the model has good robustness and recognition effect, the mean average precision is up to 0.804 when intersection over union is 0.5.

Key words image processing; liquid level detection; infrared imaging; deep learning; target detection

OCIS codes 100.3008; 100.4996; 110.3080

收稿日期: 2021-03-23; 修回日期: 2021-05-09; 录用日期: 2021-05-18

基金项目: 国家自然科学基金(61727806, 61605177, 62001430)、山西省“1331 工程”重点学科建设项目(1331KSC)、山西省自然科学基金(201801D221200)、2020 年度山西省研究生教育创新项目(2020BY101)

通信作者: *zdzhang@nuc.edu.cn

1 引言

工业生产的迅速发展,使化工过程相较于以往更加复杂精细化。对液位的准确检测可以使人们更清晰地了解整个生产过程及进度,制定生产计划,一定程度上有利于提高生产效率,确保安全生产。因此,容器液位的检测对工业生产安全至关重要^[1]。在工业生产精细化的必然趋势下,检测设备不仅需要具有高精度和非接触检测的特点,还要能适应工业现场的高温高压、灰尘、潮湿等特殊环境,对液位检测技术提出了极大的挑战。

目前,液位检测技术主要有差压式液位检测、超声波液位检测、光纤液位检测、激光液位检测、雷达波液位检测方法^[2-3]。这些方法使用的传感元件通常被布置在容器内部,进行液面检测时需要同时对容器进行改造或重新设计容器,从而造成新的隐患或增大制造成本。此外,传感器长期处于容器内部,极易老化,从而增加了维护成本。因此,发展一种非接触且不需要改造容器的液位检测技术对于工程应用至关重要。基于红外图像的液位检测技术可通过红外相机拍摄的图像对液位进行直观估计,这种方式虽然易于实现,但液位判断的准确度较低^[4-6]。深度学习技术的迅速发展^[7-8]使利用红外热成像原理和目标检测技术进行液位测量成为了可能,采集制作储罐红外图像标准数据集,并搭建深度学习模型进行储罐液位识别,能有效提高液位识别的准确度。尤其是对于高温高压容器液面的检测,该技术具有容易布置、稳定性强、不存在视觉疲劳影响等优点,相比在储罐内部安装传感器的方法有极大优势。因此,利用计算机视觉算法辅助进行液位检测,既能有效提升检测效率和检测精度、减少人力资源的浪费,又能为车间智能化升级改造奠定基础。

针对工业应用对容器液位非接触检测技术的需求,本文提出了一种基于深度学习的红外图像液位检测方法。首先,采集不同百分比容量液位状态的红外图像进行标注,制作标准数据集,并搭建基于 Pytorch 的深度学习模型框架。然后,采用改进的 YOLOv5 (You only look once, v5) 算法进行训练学习,在输入端进行数据增强,并调整卷积核数量和跨阶段局部网络(CSPNet)模块的结构,使模型具有不同的宽度和深度。最后,采用特征金字塔网络和路径聚合网络(FPN+PAN)融合不同尺寸特征图的特征信息,在输出端考虑回归框的长宽比,用联合交并比(CIOU)计算边界框回归损失,在后处理过

程中引入加权非极大值抑制(NMS)方式,提升了网络对遮挡重叠目标的识别精度。此外,对模型进行量化处理,减小了部署模型对能源和计算速度的依赖,有利于模型在移动终端的应用推广。该技术在未来工业生产车间的智能化升级中具有重要的应用前景。

2 检测原理与实验装置

2.1 液位检测原理与算法

由黑体辐射原理可知,任何温度高于绝对零度的物体都在不停地发射红外辐射。只要容器内液体与容器壁比热容不同,红外图像中就会存在温度分层现象,而介质中热扩散的存在会使界面分层变模糊。因此,采集已知实际液体百分比容量储罐的热扩散红外图像,通过人工标注建立标准数据集,并建立深度学习算法模型,可实现对模糊界面液位的检测与识别。YOLO^[9]算法采用深层卷积神经网络实现,YOLOv5 可在 Pytorch 中原生实现,模型文件格式转换的简便性极大提升了该网络的实用性。实验使用 YOLOv5 优化配置训练出满足精度和速度要求的最小权重文件只有 14.4 MB,有效降低了对终端计算能力的需求。

2.2 实验准备

2.2.1 液位检测实验平台

实验数据使用 Hti 鑫思特 HT-A2 热成像仪采集,其波长范围为 8~14 μm ,热图像分辨率为 320 pixel \times 240 pixel,热图像帧率为 9 Hz。模型训练使用的硬件平台为 Intel i9-9900K 4.7 GHz 处理器,内存为 64 GB,显卡为 NVIDIA RTX2080Ti 11 GB。系统软件平台包括 Ubuntu18.04、Pytorch 1.7.0、CUDA10.1 以及 OpenCV4.1.15。

2.2.2 液位图像数据集的采集与预处理

由于红外热成像开放数据集较少,没有针对容器液位检测的红外图像数据集。因此,实验使用 HT-A2 热成像仪采集不同百分比容量储罐模型的红外图像,通过质量控制储罐模型中的液体,得到标准 10 等分容量下的液位状态。采集过程中罐体处于自然散热状态,与环境的温差为 25~50 $^{\circ}\text{C}$,可以得到比较清晰的液位边界。测试结果表明,HT-A2 热成像仪在液体与环境温差为 5 $^{\circ}\text{C}$ 时的成像效果依然较好。改变相机拍摄的距离和角度获得不同情况下的容器液面图像,以提升数据集的普适性。对每种液位状态采集 100 多张红外图像,共计 1000 多张数据图像,初步筛选后制作数据集。其中,训练集与

验证集图像数量的比例为 9:1。对红外图像的液位状态进行标注,生成 xml 文件以及 YOLO 格式的 txt 文件。数据集按照 VOC 数据集格式进行处理^[10],包括 Annotation、Images、ImageSets 文件夹,通过 split.py 程序对标注后的红外图像进行划分,生成 train.txt、val.txt、test.txt 文件,再通过 label.

py 程序根据 xml 文件生成训练所需的记录框位置、类别等信息的 label 文件以及图像路径的 train.txt、test.txt、val.txt 文件。不同角度、拍摄距离、液位状态的储罐红外图像数据集样本如图 1 所示。其中,左上角的温度为红外采集设备检测的焦点温度, e 为设定的被测物体发射率,实验中固定为 0.85。

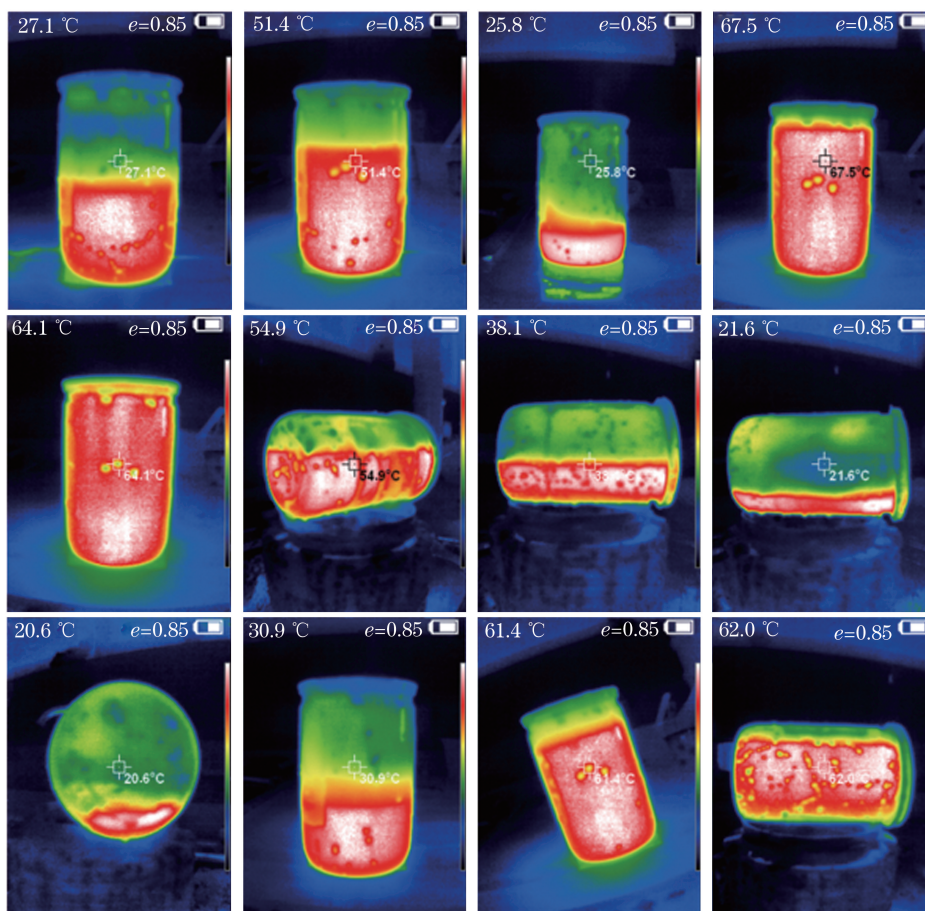


图 1 储罐红外图像数据集样本

Fig. 1 Sample of infrared image data set of storage tank

采集到相应图像后,对被测目标背景存在干扰和图像中存在的伪信号和噪声情况进行预处理,图像形态学操作腐蚀处理可以消除小区域及背景干扰,得到较为纯净的红外目标图像,原始图像及处理后的图像如图 2 所示。此外,实验也尝试了形态学梯度、顶帽、黑帽等处理方法,以应对复杂情况下获得的红外图像,消除干扰,凸显轮廓等特征^[11]。由罐体厚度及温差较小导致的液面分界不明显问题,可通过阈值分割预处理得到改善。对储罐液位红外图像进行阈值分割处理时,先由直方图得到大致的阈值,再根据实际液位实验得到准确分割的阈值,进而得到准确的液位边界,两种状态的原始图像及处理结果如图 3 所示。常见的阈值分割方法有实验法、根据直方图谷底确定阈值法、迭代选择阈值法和

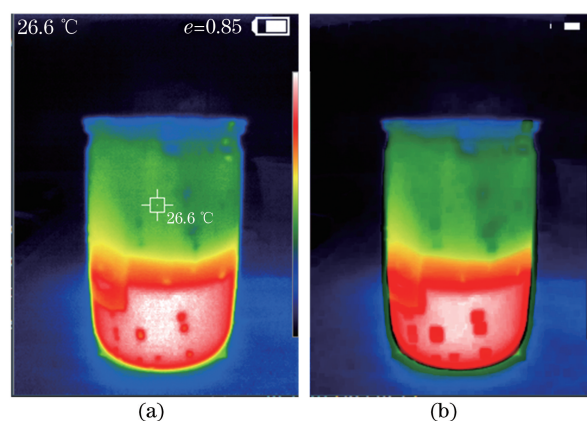


图 2 腐蚀处理的结果。(a)原始图像;(b)预处理后的图像
Fig. 2 Results of the corrosion processing. (a) Original image; (b) preprocessed image

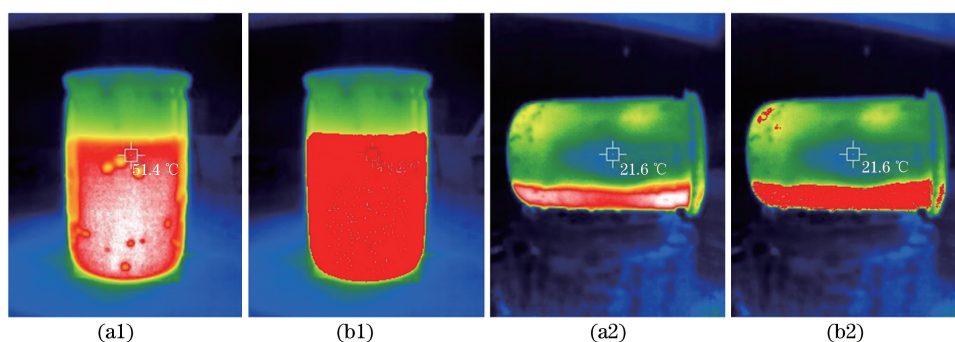


图 3 阈值分割的结果。(a)原始图像;(b)分割后的图像

Fig. 3 Result of the threshold segmentation. (a) Original image; (b) segmented image

最大类间方差法。此外,针对不同环境下的图像可以使用不同的图像预处理方法,如图像滤波去噪(均值滤波、中值滤波、高斯滤波等)及直方图均衡化处理。

3 模型优化训练与结果分析

3.1 输入数据的处理

输入端采用 Mosaic 增强方法^[12-13]对数据进行增强,主要目的是丰富红外储罐图像背景,扩增小目标数据量,优化对较小储罐目标的检测,同时防止过拟合,提高模型的鲁棒性,使模型在数据量较小的情况下依然能得到较好的检测效果。不同于常规数据增强方法,Mosaic 增强方法首先对随机读取的 4 张图像分别进行随机翻转、缩放、色域变化等操作;然后按照读取顺序分别将 4 张图像置于左上、左下、右下、右上方向;最后利用矩阵方式截取图像区域,即截取左上位置图像的左上部分,截取左下位置图像的左下部分,以此类推,得到 4 张截取图像,并拼接成新的图像。拼接后的图像包含原始的标注框,因此可用拼接后的图像进行特征提取与训练。图 4 为 Mosaic 方法处理得到的 4 张截取图像及原始标注框,标注框上的 9 表示标注为 100%容量状态的图像,2 表示标注为 30%容量状态的图像。

采用小批量梯度下降方法^[14]按批更新网络参数,即由同一批次的的数据共同决定此次梯度的更新方向,既能降低梯度下降过程的随机性,又能减少梯度更新的时间代价。针对不同数据集,目标检测算法会预设不同长宽的锚框作为预测框。训练中网络在初始锚框的基础上输出预测框,通过与真实框进行对比计算误差,并通过反向更新迭代网络参数,自适应计算不同数据集的最佳锚框^[15-16]。为了使该模型在液位数据集训练中获取更大的交并比(IOU),提高检测效果,对该数据集进行了自适应锚框计算,得到了相应的最佳锚框。

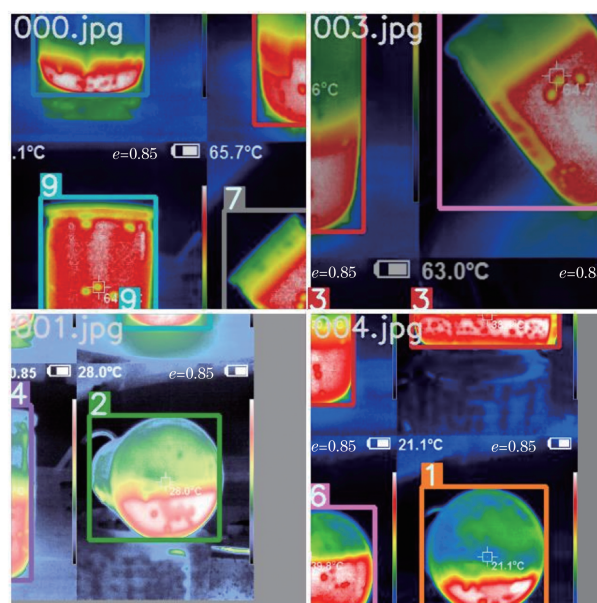


图 4 数据增强后的拼接图像

Fig. 4 Mosaic image after data enhancement

3.2 目标检测网络的结构

为了提高模型的训练效果和检测效果,采用了优化的 YOLOv5 网络结构。首先,在主干网络部分引入 Focus 结构对输入图像进行切片再拼接操作,最大程度减小下采样过程中的信息损失,如图 5 所

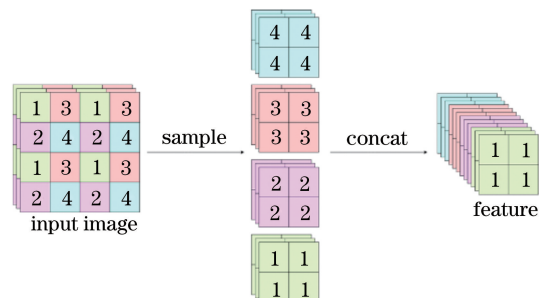


图 5 Focus 结构

Fig. 5 Focus structure

示。然后,采用不同数量的卷积核进行卷积操作,改变网络宽度。最后,在主干网络和融合特征图信息过程中分别使用 CSP1_N 和 CSP2_N 结构(N 为对应模块的数量),如图 6 所示。其中,Conv 为卷积操作,Conv2D 为二维(2D)卷积操作,Res unit 为残差单元,BN 为批归一化操作,激活函数 Leaky ReLU 为带泄漏的修正线性单元。CSP1_N 和 CSP2_N 结构均源于 CSPNet^[17],通过改变 CSP1_N 中的残差组件数量及 CSP2_N 中的卷积组数,实现对网络特征提取深度的改变。训练结果表明,在

满足应用需求的情况下,用 32 个卷积核、1 个残差组件和 2 组卷积模型能够兼顾模型的检测精度和推理速度,得到最小体积的权重文件。为了更好地利用主干网络提取的特征,采用 FPN+PAN^[18-20] 结构加强网络特征融合的能力。

参考文献[12, 21]对 IOU、GIOU(Generalized IOU)、DIOU(Distance-IOU)、CIOU 对应的损失函数 L_{IOU} 、 L_{GIOU} 、 L_{DIOU} 、 L_{CIOU} 计算方式进行了分析和对比。其中, L_{CIOU} 考虑了回归框的重叠面积、中心距离及长宽比。因此,在模型中采用 L_{CIOU} 计算边界

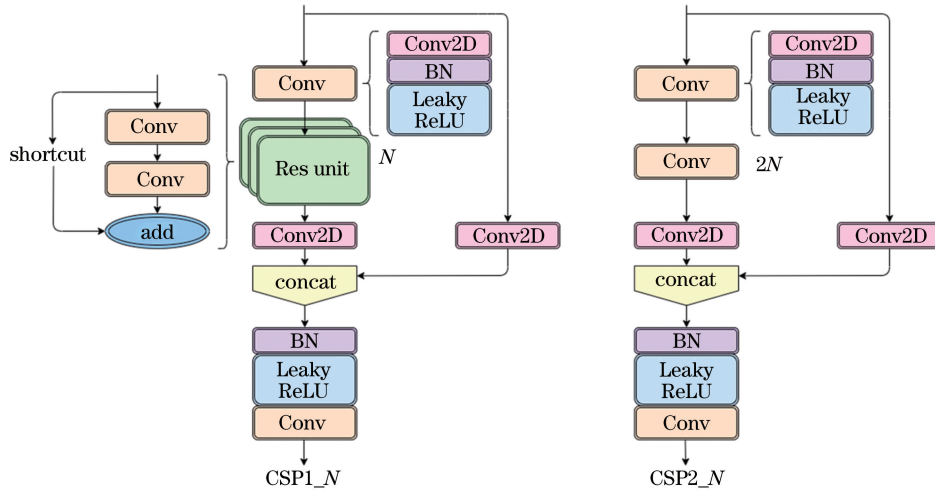


图 6 CSPNet 的结构

Fig. 6 Structure of the CSPNet

框,可以得到较好的收敛速度和精度,可表示为

$$L_{CIOU} = 1 - X_{IOU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu, \quad (1)$$

$$X_{IOU} = \frac{|A \cap B|}{|A \cup B|}, \quad (2)$$

$$\alpha = \frac{\nu}{(1 - X_{IOU}) + \nu}, \quad (3)$$

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \quad (4)$$

式中, b, b^{gt} 分别为预测框(A)和真实框(B)的中心点, ρ 为两个中心点间的欧氏距离, c 为能同时包含预测框和真实框最小闭合区域的对角线距离, X_{IOU} 为 A 框和 B 框的 IOU, α 为正权衡参数, ν 为度量长宽比相似性的参数, w, h 分别为目标检测框的宽和高。

3.3 边界框的处理方式

在目标检测后处理过程中,用训练得到的模型参数计算生成锚框的类别和偏移量,得到预测边界框。由于 IOU 设定的阈值和锚框数量不同,预测中一个检测目标可能会输出多个相似的预测边界

框^[22],不仅增加了计算量,还会导致输出结果不整齐。因此,还需对预测结果进行 NMS 操作^[23]。本模型在训练中采取加权非极大值抑制方法,相比传统的 NMS,该方法不直接剔除 IOU 大于阈值且类别相同的框,而是根据网络预测的置信度进行加权,得到最终预测的矩形框,再剔除其他框。加权非极大值抑制能获得更高的准确率和召回率^[24],得到更有效的预测框。

3.4 模型性能分析

3.4.1 模型评价指标

常用来评价模型目标检测性能的指标有准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F-Score,具体计算方法如表 1 所示。其中, β 为调整精确率和召回率权重的参数, X_{TP} 、 X_{FP} 、 X_{FN} 、 X_{PN} 的含义如表 2 所示。

通常情况下,精确率越高,召回率越低;反之,召回率越高,精确率就越低。F-Score 是对这两个指标的综合权衡,用 β 调整精确率和召回率的权重,当 $\beta=1$ 时,两者权重相同,简称为 F1-Score。在实际工程应用中,若要获得更准确的检测框,则需减小

表 1 评价模型的性能指标及计算方式

Table 1 Performance indicators and calculation methods of the evaluation model

Indicator	Describe	Calculation method
Accuracy	ratio of correctly classified samples to total samples	$\frac{X_{TP} + X_{TN}}{X_{TP} + X_{FN} + X_{FP} + X_{TN}}$
Precision (P)	ratio of correctly retrieved samples to total retrieved samples	$\frac{X_{TP}}{X_{TP} + X_{FP}}$
Recall (R)	ratio of correctly retrieved samples to samples that should have been retrieved	$\frac{X_{TP}}{X_{TP} + X_{FN}}$
F-Score	consider the reconciliation between accuracy and recall	$\frac{(1 + \beta^2) \cdot P \cdot R}{\beta^2 \cdot (P + R)}$

表 2 正负样本关系

Table 2 Positive and negative sample relationship

Case	Test outcome	Truth
X_{TP}	anomaly	anomaly
X_{FP}	anomaly	normol
X_{FN}	normol	anomaly
X_{TN}	normal	normal

β ;若优先考虑检测到目标,则需增大 β 。当召回率从 0 增大到 1,平均精度(AP)可以由 11 个不同召回率下的最大精确率计算,可表示为

$$S_{AP} = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1.0\}} S_{AP}(r) = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1.0\}} P_{interp}(r), \quad (5)$$

式中, S_{AP} 为准确率-召回率曲线下的面积,平均精度均值(mAP)是各类别检测精度的平均值, $P_{interp}(r)$ 为召回率超过 r 时对应的最大精确率,可表示为

$$P_{interp}(r) = \max_{\tilde{r} \geq r} p(\tilde{r}). \quad (6)$$

3.4.2 模型的检测结果及分析

在最终的模型训练中进行了 2000 轮迭代实验,训练过程的可视化结果如图 7 所示。图 7(a)、图 7(d)为回归框的损失均值,该值越小,表明预测结果越准确;图 7(b)、图 7(e)为目标检测中预测框的准确度损失均值,该值越小,表明检测结果越准确;图 7(c)、图 7(f)为分类损失均值,该值越小,表明分类结果越准确;图 7(g)、图 7(h)为随训练的进程模型精确率和召回率的变化情况。对于液位检测任务,最主要的目的是检测出储罐液位,而该模型的召回率可达到 0.87,意味着有较高的检出效果。mAP@0.5 表示正负样本阈值为 0.5 时所有类的 mAP。在模型训练过程中取准确率和召回率较高的模型设置,训练得到 mAP@0.5 为 0.804 的检测模型。在液位监测的大部分应用中只需要判断液位

是过高和过低。通过目标检测的方式识别检测液位状态时,状态分类越少,分类难度越低,相应的检测精度较高;若状态分类过多,相邻液位状态的特征区别不明显,导致分类难度增加,相应的检测精度会有一定程度的降低。实验中的液位检测研究对象为直径和高度比为 1:2 的储罐模型,储罐内液体量以容积的 10 等分递增,分为 10 个不同的液位状态,因此测量精度为罐体高度的 10%。该液位检测方法在不同应用条件下,测量精度可以通过目标储罐的粗细以及红外图像状态标注精度来提升,而模型检测精确度则需要网络模型来优化提升。

得到检测模型后,对实际检测效果进行测试,在储罐中加入固定百分比容量的液体,在液体温度高于室温 10 °C 时进行红外图像采集,将得到的图像输入训练好的模型中,得到目标检测结果。为显示对不同百分比容量储罐的检测效果,随机选取了不同液位状态的检测结果并进行拼接展示,结果如图 8 所示。可以发现,该模型能准确检测出储罐所在的位置,并在目标框左上角给出推理结果和置信度值。实际液体体积占罐体容量的 40%,模型推理得到储罐内液体百分比容量是 40,置信度为 0.95;相机倾斜一定角度采集的储罐红外图像实际百分比容量为 60%,模型推理结果也是 60,置信度为 0.89,液位状态检测结果准确。

传统液位测量方法通过连通器结构直接读取液位高度;在罐体外侧壁上下移动超声波传感器并根据收发信号判断出液面位置;在罐体正上方使用超声波传感器测量液面距离,得到液面高度;将电容式液位测量传感器浸入液面,根据输出的电信号间接得到液面高度。而实验中结合储罐红外图像和深度学习进行液位状态识别检测的方法,有灵活的量程范围,可根据使用场景调整测量精度,解决了传统测量方法存在的安装复杂、位置角度要求高、使用维修费用高等问题,在液位辅助检测以及苛刻的测量环

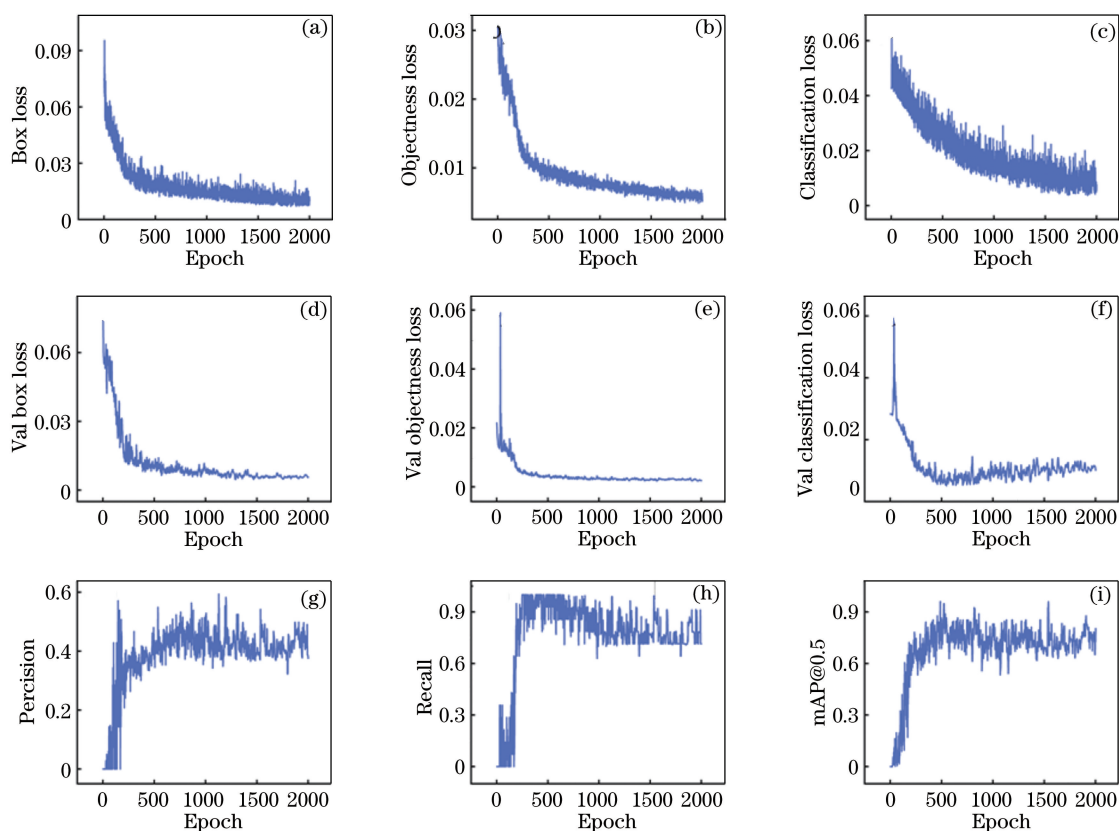


图 7 训练过程的可视化。(a) Box 损失;(b) Objectness 损失;(c) Classification 损失;(d)验证集 Box 损失;(e)验证集 Objectness 损失;(f)验证集 Classification 损失;(g)精确率;(h)召回率;(i) mAP

Fig. 7 Visualization of the training process. (a) Box loss; (b) Objectness loss; (c) Classification loss; (d) validation set Box loss; (e) validation set Objectness loss; (f) validation set Classification loss; (g) precision; (h) recall; (i) mAP

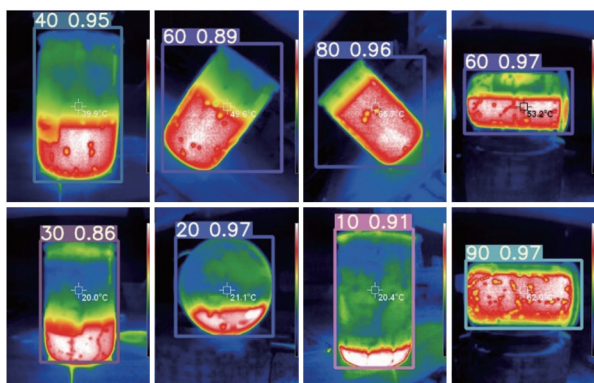


图 8 液位状态的检测结果

Fig. 8 Test results of liquid level state

境下有明显优势。

4 结 论

面向容器液位非接触检测问题,提出了一种基于深度学习的储罐液位红外图像检测方法。通过 Mosaic 数据增强方法丰富了红外储罐图像的背景,防止网络过拟合;通过缩放产生小目标,以优化对较小储罐目标的检测。在 Focus 结构之后使用 32 个

卷积核,在主干网络 CSP1_N 结构中使用 1 个残差组件,融合特征图信息时在 CSP2_N 结构中使用 2 组卷积,得到的模型能有效提取目标特征。实验结果表明,该方法的 mAP@0.5 达到 0.804,可给出不同角度采集的储罐红外图像中储罐液体占容积的百分比,且推理结果具有较高的置信度,从而实现对储罐容器液位状态的非接触检测。因此,该方法能应用于工厂储液罐、燃油罐等设备的液位状态检测。相比传统检测方法,本方法避免了布置特殊传感器的步骤,只需架设红外图像采集装置,用目标检测的方式代替人工判断,实现对储罐液体占容积百分比的检测输出。通过数据增强方式提高数据集的普适性,在数据量较小的情况下也能得到较好的检测效果。

参 考 文 献

[1] Bera S C, Mandal H, Saha S, et al. Study of a modified capacitance-type level transducer for any type of liquid [J]. IEEE Transactions on Instrumentation and Measurement, 2014, 63 (3):

- 641-649.
- [2] Terzic J, Nagarajah C R, Alamgir M. Fluid level measurement in dynamic environments using a single ultrasonic sensor and support vector machine (SVM) [J]. *Sensors and Actuators A: Physical*, 2010, 161 (1/2): 278-287.
- [3] Jiang Y, Yuan M T. The study of improving ultrasonic ranging accuracy based on the double closed-loop control technology [C]//2017 Chinese Automation Congress (CAC), October 20-22, 2017, Jinan, China. New York: IEEE Press, 2017: 2462-2466.
- [4] Miao Z, Zhang Y, Chen R M, et al. Method for fast detection of infrared targets based on key points [J]. *Acta Optica Sinica*, 2020, 40(23): 2312006.
苗壮, 张湧, 陈瑞敏, 等. 基于关键点的快速红外目标检测方法 [J]. *光学学报*, 2020, 40(23): 2312006.
- [5] Nguyen L S, Schaeli B, Sage D, et al. Vision-based system for the control and measurement of wastewater flow rate in sewer systems [J]. *Water Science and Technology*, 2009, 60(9): 2281-2289.
- [6] Melchionni D, Norgia M. Optical system for liquid level measurements [J]. *The Review of Scientific Instruments*, 2014, 85(7): 075113.
- [7] Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism [J]. *Acta Optica Sinica*, 2020, 40(5): 0504001.
赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测 [J]. *光学学报*, 2020, 40(5): 0504001.
- [8] Grauman K, Darrell T. The pyramid match kernel: discriminative classification with sets of image features [C]//Tenth IEEE International Conference on Computer Vision (ICCV'05), October 17-21, 2005, Beijing, China. New York: IEEE Press, 2005, 1: 1458-1465.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [10] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge [J]. *International Journal of Computer Vision*, 2010, 88 (2): 303-338.
- [11] Wang D, Shen T. Research on weak and small infrared target detection algorithm under complex sky background [J]. *Acta Optica Sinica*, 2020, 40(5): 0512001.
王笛, 沈涛. 复杂天空背景下的红外弱小目标检测算法研究 [J]. *光学学报*, 2020, 40(5): 0512001.
- [12] Bochkovskiy A, Wang C Y, Mark L H Y. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2021-03-11]. <https://arxiv.org/abs/2004.10934v1>.
- [13] Zoph B, Cubuk E D, Ghiasi G, et al. Learning data augmentation strategies for object detection [M]//Vedaldi A, Bischof H, Brox T, et al. *Computer vision-ECCV 2020. Lecture notes in computer science*. Cham: Springer, 2020, 12372: 566-583.
- [14] Li M, Zhang T, Chen Y Q, et al. Efficient mini-batch training for stochastic optimization [C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, August 14, 2014, New York, USA. New York: ACM, 2014: 661-670.
- [15] 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 and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [16] Lu Y X, Javidi T, Lazebnik S. Adaptive object detection using adjacency and zoom prediction [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2351-2359.
- [17] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.
- [18] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [19] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [20] Liu F, Wu Z W, Yang A Z, et al. Multi-scale feature fusion based adaptive object detection for UAV [J]. *Acta Optica Sinica*, 2020, 40(10): 1015002.
刘芳, 吴志威, 杨安喆, 等. 基于多尺度特征融合的自适应无人机目标检测 [J]. *光学学报*, 2020, 40(10): 1015002.

- [21] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression [EB/OL]. (2019-11-19) [2021-03-22]. <https://arxiv.org/abs/1911.08287>.
- [22] Li Z Q, Zhu R F, Ma J Y, et al. Airport detection method combined with continuous learning of residual-based network on remote sensing image[J]. Acta Optica Sinica, 2020, 40(16): 1628005.
李竺强, 朱瑞飞, 马经宇, 等. 联合连续学习的残差网络遥感影像机场目标检测方法 [J]. 光学学报, 2020, 40(16): 1628005.
- [23] Bodla N, Singh B, Chellappa R, et al. Soft-NMS: improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5562-5570.
- [24] Ning C C, Zhou H J, Song Y, et al. Inception single shot MultiBox detector for object detection[C]//2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), July 10-14, 2017, Hong Kong, China. New York: IEEE Press, 2017: 549-554.