

基于轻量化的指针仪表检测算法研究

骆东松,张杰锋,魏義民

(兰州理工大学 电气工程与信息工程学院, 甘肃 兰州 730050)

摘要:通过轻量化的神经网络算法部署在边缘计算设备是解决老式仪表读数问题的主流方向。该文提出了 YOLOv5s - Pointer 轻量化的指针仪表检测网络模型,以 YOLOv5s 网络模型作为基础,通过引入 Mixup 数据增强技术,使用 MobileNetV3 网络替换 CSPDarkNet53 特征提取网络,采用 SLoU Loss 定位损失函数,并引入动态样本权重思想,对网络进行改进。实验结果表明,相较于 YOLOv5s 网络在验证集上的参数量和计算量分别减少了 78% 和 57%,精确率提升了 1.3%。

关键词:指针仪表检测;YOLOv5s;数据增强;MobileNetV3;轻量化

中图分类号:TP391.41

文章编号:1000-0682(2024)01-0098-06

文献标识码:A

DOI:10.19950/j.cnki.CN61-1121/TH.2024.01.021

Research on pointer gauge detection algorithm based on lightweighting

LUO Dongsong, ZHANG Jiefeng, WEI Yimin

(Lanzhou University of Technology, Gansu Lanzhou 730050, China)

Abstract: Deployment of neural network algorithms via lightweight neural network algorithms to edge computing devices is the mainstream direction to solve the old meter reading problem. In this paper, we propose the YOLOv5s - Pointer lightweight network model for pointer meter detection, using the YOLOv5s network model as the basis, replacing the CSPDarkNet53 feature extraction network with the MobileNetV3 network by introducing the Mixup data enhancement technique, using the SLoU Loss to localize the loss function, and introducing the dynamic sample weighting idea to improve the network. The experimental results show that compared to YOLOv5s network the number of parameters and computation on the validation set are reduced by 78% and 57%, respectively, and the accuracy rate is improved by 1.3%.

Keywords: pointer meter inspection; YOLOv5s; data enhancement; MobileNetV3; lightweighting

0 引言

目前,厂区存在人工读数的现象。考虑到传统仪表的特性,通过机器视觉的方法来进行示数读取,可以更大程度的降低人工读数存在的问题^[1]。仪表示数读取问题,现场仪表表盘的检测是重要环节,绝大多数仪表表盘检测算法以满足不同环境下检测

精度为首要目标,在带来高精度检测的同时,使得网络模型更加复杂,运行时产生大量的计算和参数,在实际项目中,采用智能终端进行边缘计算的方式进行仪表示数读取上传,并且要保证上传的实时性,使得边缘计算芯片面临巨大压力。

考虑到深度学习对边缘计算芯片带来的巨大压力,该文致力于深度学习算法的轻量化研究,提出针对指针仪表检测的算法模型 YOLOv5s - Pointer,并通过拍摄现场指针仪表自建数据集进行验证。采用单阶段 YOLOv5s 算法为基础网络,首先,考虑到指针仪表数据集难易样本对训练结果的影响,在网络输入端加入 Mixup 数据增强技术。其次,在主干网络中引入轻量化特征提取网络 MobileNetv3 替换 CSPDarkNet53,大大减少了网络计算量。最后,将

收稿日期:2023-09-18

第一作者:骆东松(1970—),男,甘肃武山人,本科,教授,研究方向为嵌入式系统、计算机控制系统的研究与开发等。

E-mail:2677007909@qq.com

通信作者:张杰锋(2000—),男,甘肃会宁人,研究生,研究方向为计算机控制系统的研究与开发。

E-mail:2829940332@qq.com

YOLOv5s 原有的定位损失函数 CLoU 替换为 SLoU, 使目标框回归速度更快更稳定, 提升模型检测精度, 并通过引入动态样本加权模块, 动态改变难易样本的权重, 使网络对优秀样本的关注度更高, 防止网络过度依赖简单样本, 使训练网络过拟合。

1 YOLOv5s 模型

YOLOv5 算法不仅在处理速度方面有很大的优

势, 其检测精度也很高^[2]。通过不同的网络宽度, YOLOv5 分成了 s 、 m 、 l 、 x 四个模型, 其中 YOLOv5s 网络模型层数最浅, 处理速度最快。考虑到系统需要兼顾精度和速度, 采用 YOLOv5s 作为基础网络。YOLOv5s 模型结构包括输入端 (Input)、主干 (Backbone) 网络、颈部 (Neck) 网络和输出端 (Head)。YOLOv5s 模型结构如图 1 所示。

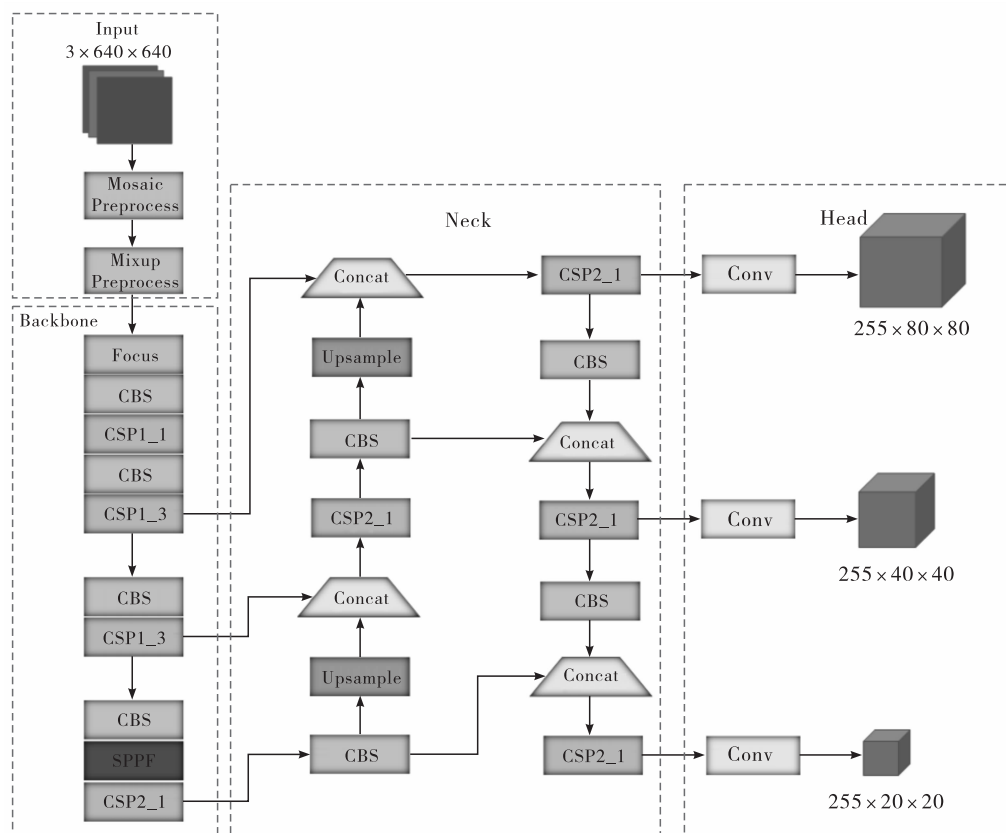


图 1 YOLOv5s 模型结构图

输入端主要对图片进行数据增强处理, 数据增强的方法包括自适应图片缩放、马赛克 (Mosaic) 数据增强、自适应锚框计算^[3]。

主干网络负责提取特征并生成特征图。结构包括 Focus、CBS、CSP、SPPF。Focus 通过 Slice 技术对图片进行裁剪。YOLOv5s 模型中采用 CBS 基本卷积模块。CSP 有 CSP1_X 和 CSP2_X 两种结构。快速-空间金字塔池化 (SPPF), 模块将输入通过 3 个大小为 5×5 的最大池化, 将产生的特征图通过 Concat 操作, 提升特征提取速度。

颈部采用 FPN + PAN 的结构, 能够更好的提升提取特征的能力: FPN (Feature Pyramid Network) 结构, 通过上采样的方式, 从上向下, 将高层特征信息传递到底层信息并进行融合, 对小物体的检测精度

提升很高^[4]; PAN (Path Aggregation Network) 结构则与 FPN 结构相反。YOLOv5s 输出端可以输出对应小、中、大三种不同检测尺度, 在经过 8、16、32 倍的下采样后得出。

输出端主要实现分类任务和回归任务。YOLOv5s 中用于回归优化损失函数采用 CLoU Loss, CLoU 综合考虑了预测框与真实框的距离、重叠率、尺度以及惩罚项等因素, 使目标框回归变得更加稳定。并采用非极大值抑制寻找最佳的目标边界框, 剔除冗余的边界框^[5]。

2 YOLOv5s - Pointer 轻量化模型

研究提出一种 YOLOv5s - pointer 轻量化网络模型, 通过引入 MobileNetV3^[6] 轻量化网络单元, 重

新构建 YOLOv5s 主干特征提取网络;在头部引入 Mixup 数据增强模块,提升自制数据集优秀样本的数量,训练模型更加精确且普适性更强;采用更加先进的 SLoU Loss 作为回归损失函数,并引入动态样本权重方法对其进行改进,使训练模型对优秀样本的关注度更强,减少简单样本造成的训练损失。

2.1 Mixup 数据增强引入

在网络头部引入 Mixup 数据增强,进一步扩展指针仪表数据集,防止网络模型过拟合。在训练过程中,设置每个 epoch 中,出现 Mosaic 数据增强的概率为 60%,且在出现 Mosaic 数据增强时,Mixup 数据增强才启动。

2.2 MobileNetv3 卷积模块引入

YOLOv5s 主干采用 CSPDarkNet53 卷积网络,有 104 层,在特征提取过程中进行大量的卷积操作,占用运算资源,不适合边缘计算设备上运行,需要对特征提取网络进行轻量化改进。

使用 MobileNetv3 代替 CSPDarkNet53 进行特征提取时的卷积网络。MobileNet^[7] 网络专注点在于移动设备中的卷积神经网络。MobileNetV3 使用轻量级倒残差结构模块,它在网络的不同层级上引入了倒残差结构,以增加网络的非线性表示能力;MobileNetV3 还引入了一种称为“可分离卷积”的操作,它将标准卷积操作分解为深度卷积和逐点卷积两个步骤,以减少计算量和参数数量。此外,MobileNetV3 还采用 SE 注意力机制^[8],通过学习通道间的相关性来增强网络的表达能力;通过改进 swish 激活函数,采用了更加节省计算量的 H-swish 激活函数,增强了量化过程;重新设计了耗时层结构,将第一个卷积层的卷积核个数由 32 个减少为 16 个,并精简了 Last-Stage,检测效果基本没有改变,但处理过程更加简洁。替换 YOLOv5s 特征提取网络后,MobileNet v3 相较于 CSPDarkNet53,帧率提升了 33.3%,计算量减少了 150.1%,识别准确度只降低了 11.2%。

2.3 损失函数优化

YOLOv5s 采用完全交并比损失函数,CIoU Loss 相较于 DLoU Loss^[9],将 Bounding box 的纵横比引入到损失函数中,提升了回归精度。CIoU 在 DLoU 惩罚项基础上增加了影响因子 $\alpha\gamma$,通过这个影响因子将预测框的纵横比拟合真实框的纵横比考虑进去。SIoU 在 CIoU 的基础上把预测框和真实框之间的向量角度作为惩罚项加入损失函数,防止预测框四处漂移,并使它更加贴合真实框,进一步提升网络的收敛速度和训练效果。因此本文选择用 SIoU 替代 CIoU。

SLoU Loss 通过加入角度惩罚项,使得收敛速度更快。角度成本示意图如图 2 所示。

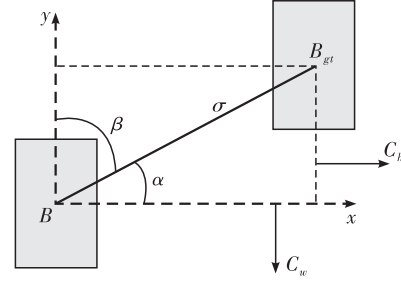


图 2 角度成本示意图

在回归过程中,当角度 α 小于 $\pi/4$ 时,预测框向 x 轴移动,并沿 x 轴迅速向真实框靠近。否则将 β 快速最小化,预测框靠近 y 轴,并沿 y 轴迅速靠近真实框。角度成本函数如式(1)所示:

$$\varphi = 1 - 2 \sin^2(\arcsin x - \frac{\pi}{4}) \quad (1)$$

其中:

$$x = C_h / \sigma = \sin \alpha \quad (2)$$

引入角度成本后,可以最大限度的减少与距离相关的变量数量。

重新定义了距离成本,如式(3)所示:

$$\Delta = \sum_{t=x,y} (1 - \exp(-\gamma \rho_t)) \quad (3)$$

其中: $\gamma = 2 - \varphi$; $\rho_x = (b_{cx}^{gt} - b_{cx})^2 / c_w^2$; $\rho_y = (b_{cy}^{gt} - b_{cy})^2 / c_h^2$, c_w 为真实框与预选框质心横坐标差。当 α 越接近 $\pi/4$ 时,距离损失越大。

形状成本表达式如式(4)所示:

$$\Omega = \sum_{t=w,h} (1 - \exp(-\omega_t))^\theta \quad (4)$$

其中:

$$\omega_w = \frac{|w - w^{gt}|}{\max(w - w^{gt})} \quad (5)$$

$$\omega_h = \frac{|h - h^{gt}|}{\max(h - h^{gt})} \quad (6)$$

其中: w 、 h 为预测框的宽和高; w^{gt} 、 h^{gt} 为真实框的宽和高。

LoU 成本表达式如式(7)所示:

$$LoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (7)$$

SLoU Loss 最终表达式为:

$$L = 1 - LoU + (\Delta + \Omega) / 2 \quad (8)$$

好的损失函数应该在锚框与目标框较好地重合时削弱几何度量的惩罚,不过多地干预训练将使模型有更好的泛化能力^[10]。为解决这个问题,引入动态样本加权模块对样本权重进行干预。

样本不平衡的问题一直存在与目标检测任务中,在指针仪表检测任务中,通过分析现场仪表分布情况,样本中大部分为简单易检测的样本,困难样本存在比较稀疏,因此该文将样本加权函数的思想引入损失函数中。

Huang^[11]等人根据每个样本的难易程度,动态地调整样本权重。自适应求出所有边界框的 LoU 值的平均值,作为划分样本难易的阈值 μ ,将 LoU 值在 μ 以下作为负样本,在 μ 以上的作为正样本。

在阈值附近的样本通常遭受的损失较大,在训练过程中,希望模型对阈值附近的样本能够更充分的学习优化。处在阈值附近的样本占有所有样本的比例只是少数,可以尝试为这些样本增加训练时的权重,通过 slide 对所有样本进行强调,slide 加权函数表达式(9)所示:

$$f(x) = \begin{cases} 1, & x \leq \mu - 0.1 \\ e^{1-\mu}, & \mu < x < \mu - 0.1 \\ e^{1-x}, & x \geq \mu \end{cases} \quad (9)$$

通过上式我们可以看出作者对简单样本权重函数通过引入非线性函数来解决,非线性函数提高了精度,但其部署在边缘设备时,会带来计算成本的提升,我们通过线性函数来模拟这一部分中的非线性函数,如式(10)所示。通过在验证集中验证,线性函数带来的损失几乎可以无视,降低了计算成本。样本划分阈值示意图如图 3 所示。

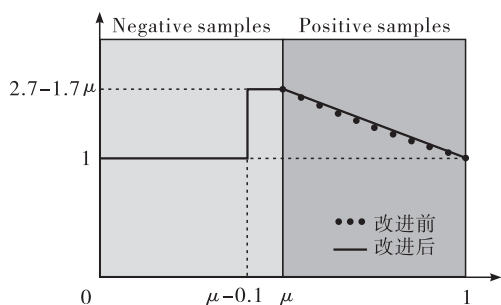


图 3 样本划分阈值

$$f(x) = \begin{cases} 1, & x \leq \mu - 0.1 \\ 2.7 - 1.7\mu, & \mu < x < \mu - 0.1 \\ -1.7x + 2.7, & x \geq \mu \end{cases} \quad (10)$$

3 试验与分析

3.1 环境搭建

该文使用 Pycharm 编程软件,Pytorch 深度学习框架,Pytorch 版本为 1.12.1,Python 版本为 3.8,CUDA 版本为 11.3.1,在 Windows 11 操作系统下训练搭建的环境,电脑 CPU 为 AMD Ryzen 7 5800H,Gpu 为

NVIDIA GTX 3050Ti,显存为 4 G,运行内存为 16 GB。

在模型训练开始前,输入图片的像素设置为 640×640 ,学习率设为 0.05,每次处理图片数量设置为 4,使用 SGD 作为梯度下降的优化器,训练次数设置为 250。

3.2 数据集建立

3.2.1 指针仪表图片采集

搭建指针仪表图像采集环境,图片采集设备为海康威视摄像头。根据不同的天气情况,不同角度进行拍摄,包括阴天拍摄、晴天拍摄、白天拍摄、夜间拍摄、近距离拍摄、远距离拍摄、仰角拍摄、俯角拍摄等多种情况。现场仪表环境中,可能存在其他圆形设备,将指针仪表置于圆形设备之中进行大量拍摄。通过以上方法提升数据集的多样性,提升模型训练精度。在经过三天连续采集指针仪表图片后,得到了 3600 张现场仪表照片,通过对比筛选,将拍摄不够清晰,距离太近以及太远的图片去除,实际保留与项目相符合的照片共 3100 张,分辨率为 960×540 。

3.2.2 数据集制作

使用 LabelImg 图像标注软件对 3100 张指针仪表图片进行标注,采用矩形框将不同场景中的指针仪表表盘框出。本次研究着重对指针仪表表盘的检测,所以在标注类名时,全部标注为“meter”。目前,在训练集和验证集的划分中,经常采用的划分比例为 9:1 或 8:2,考虑到训练集过多导致的过拟合现象,按照 8:2 的划分方式对数据集进行划分。划分后训练集包含 2480 张,验证集包含 620 张。

3.3 评价指标

对训练模型精度的评价指标主要有精确度 P 、召回率 R 、平均精度均值(mAP);用训练模型产生的权重大小、参数量、计算量来体现模型的复杂程度;用帧率(FPS)来体现模型实时检测的性能。

精确度 P (Precision) 表示预测正确的样本占有所有样本的比例,如式(11)所示:

$$P = \frac{TP}{TP + FP} \quad (11)$$

召回率 R (Recall) 表示在所有样本中,预测正确样本所占的比例,如式(12)所示:

$$R = \frac{TP}{TP + FN} \quad (12)$$

mAP 是平均精度 AP 的均值,计算平均精度 AP 的方法为 $P-R$ 曲线下方的面积,如式(13)所示:

$$V_{AP} = \int_0^1 P(r) dr \quad (13)$$

式中: TP 表示预测正确的正样本数量; FP 表

示被错误的预测为正样本的数量; FN 表示被错误的预测为负样本的数量; N 表示的数据集中的类数, 本研究训练模型过程中, 只有指针仪表一类, 所以 N 的值为 1。

3.4 试验结果分析

3.4.1 算法性能对比分析

通过对比 YOLOv5s - Pointer 与 YOLOv5s 在指针仪表数据集上性能表现, 来验证本研究的可行性。搭建训练环境, 在相同的实验环境和设置下, 分别对两种模型进行 250 轮的训练, 通过验证集进行验证, 得出 YOLOv5s - Pointer 在验证集上的损失函数曲线如图 4 所示。网络模型在前 50 epoch 收敛速度较

快, 到训练结束时, 损失达到最小, 没有出现过拟合现象。将训练结果进行对比如表 1 所示。

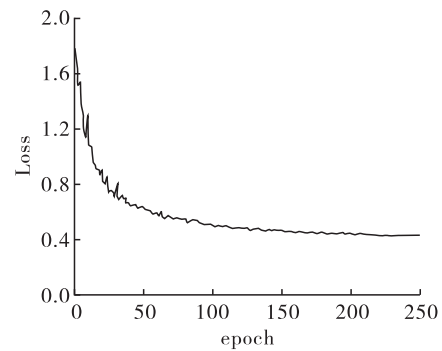


图4 损失函数

表1 YOLOv5s - Pointer 和 YOLOv5s 实验结果对比

模型	$P/\%$	$R/\%$	$AP_{0.5}/\%$	GFLOPS	参数量	模型大小/MB
YOLOv5s	97.5	95.0	98.1	14.3	6 990 532	12.9
YOLOv5s - Pointer	98.8	96.1	98.0	6.2	1 503 348	3.0

通过对表 1 进行分析, 可以看出 YOLOv5s - Pointer 网络模型在验证集上精确率达到 98.8%, 相较于原始网络模型 YOLOv5s 提高了 1.3%; YOLOv5s - Pointer 网络模型在验证集上的召回率达到 96.1%, 增长了 1.1%; $AP(0.5)$ 为 98.0%, 与 YOLOv5s 相比降低了 0.1%; 通过对比实验, 初步验证了改进算法的可行性。模型在验证集上的检测结果如图 5 所示。

3.4.2 轻量化网络消融试验分析

本节主要通过消融试验对改进的轻量化模型进行性能分析和验证。YOLOv5s 仪表盘检测时, 会产生大量的参数和计算, 同时伴随着较大的模型权重训练文件。通过消融试验分析引入 Mixup 数据增强、MobileNetv3 卷积模块、SLoU 定位损失函数、动态样本权重模块后对网络模型性能影响。具体试验结果如表 2 所示。

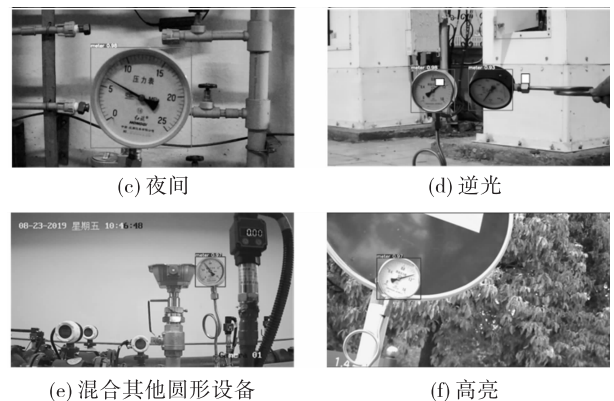
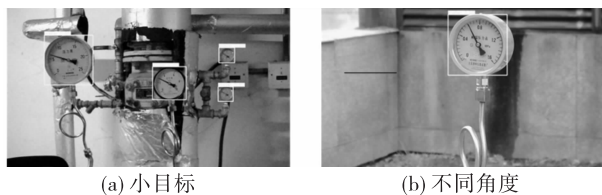


图5 仪表样例

表 2 中, 模块 1 是在模型头部加入 Mixup 数据增强, 模块 2 是使用 MobileNetv3 替换 CSPDarkNet53 网络, 模块 3 是用 SLoU 替换 CLoU, 模块 4 是引入动态样本加权模块。

表2 消融实验

模型	模块 1	模块 2	模块 3	模块 4	$P/\%$	$R/\%$	$AP_{0.5}/\%$	GFLOPS	参数量	模型大小/MB
YOLOv5s	✓				97.5	95.0	98.1	14.3	6 990 532	12.9
					97.7	94.9	98.1	14.4	6 990 532	12.9
		✓			95.7	93.9	96.6	7.87	3 495 322	4.4
YOLOv5s - Pointer	✓		✓	✓	98.6	96.5	98.5	13.9	5 843 981	10.3
		✓	✓	✓	98.8	96.1	98.0	6.2	1 503 348	3.0

由表 2 中数据可以看出,实验单独加入模块一后,对实验精度提升有一定帮助,提升了 0.02%;实验单独加入模块二后,模型的精确率降低 1.8%,召回率降低 1.1%,平均精度降低 1.5%,MobileNetv3 轻量化网络带来损失在可控范围内,但模型参数量减少了 50%,计算量减少了 55%;实验加入模块三和模块四后,精确率和召回率都有明显提升,分别提升了 1.1% 和 1.5%,平均精度也提升 0.4%。通过消融实验可以看出,YOLOv5s - pointer 算法取得了较好的检测结果,同时模型更加轻量化。

4 结语

目前仪表识别算法模型只在特定环境中表现良好,且模型较为复杂,不适合边缘设备部署。鉴于以上存在的问题,该文提出了 YOLOv5s - Pointer 网络模型,将 Mixup 数据增强加入 YOLOv5s 网络头部,训练过程中使网络可以更好的适应不同的场景;引入 MobileNetv3 替换 YOLOv5s 中原有的 CSPDarkNet53 网络,极大的降低原有网络的参数量,提升检测速率;替换 YOLOv5s 中定位损失函数 CLoU,使用更有利的 SLoU,并通过引入动态样本加权思想,使得模型对困难样本的关注度更高,避免了过多的关注简单样本导致训练出现过拟合现象。通过自建指针仪表数据集对改进模型进行训练和验证,实验结果表明,YOLOv5s - Pointer 网络对比 YOLOv5s 网络检测精度提升 1.3%,同时参数量减少了 78%,可以满足边缘设备的部署要求。后期将对模型进一步改进,并通过引入其他改进方法,满足指针仪表读数系统在设备上的运行。

参考文献:

- [1] 陈从平,王钦,姚威. 基于机器视觉的指针式仪表自动读数算法[J]. 传感器与微系统, 2022,41(10): 130 - 133.
- [2] 董文轩,梁宏涛,刘国柱. 深度卷积应用于目标检测算法综述[J]. 计算机科学与探索, 2022,16(05): 1025 - 1042.
- [3] 邵延华,张铎,楚红雨. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报, 2022,44(10): 3697 - 3708.
- [4] 赵英,王琦,沙捷. 基于 YOLOv5s 的轻量化乒乓球目标检测算法[J]. 计算机应用, 2023,43(S1): 229 - 234.
- [5] 金鑫,庄建军,徐子恒. 轻量化 YOLOv5s 网络车底危险物识别算法[J]. 浙江大学学报(工学版), 2023,57(08): 1516 - 1526 + 1561.
- [6] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [J]. 2019. DOI: 10. 48550/arXiv. 1905. 02244.
- [7] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- [8] 任欢,王旭光. 注意力机制综述[J]. 计算机应用, 2021(S1 vo 41): 1 - 6.
- [9] YIN Y, LEI L, LIANG M, et al. Research on Fall Detection Algorithm for the Elderly Living Alone Based on YOLO [C]//2021 IEEE International Conference on Emergency Science and Information Technology (IC-ESIT).
- [10] TONG Z, CHEN Y, XU Z, et al. Wise - IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[Z/OL]. arXiv, 2023(2023 - 04 - 08) [2023 - 09 - 15].
- [11] YU Z, HUANG H, CHEN W, et al. YOLO - FaceV2: A Scale and Occlusion Aware Face Detector[Z/OL]. arXiv, 2022(2022 - 08 - 04) [2023 - 09 - 15].
- [3] 金路,陶笑昱,王鹏. 浅谈空盒气压表和气压计的计量检定[J]. 仪器仪表用户, 2020,27(12): 110 - 112.
- [4] 刘海珍,郑树芳,姜雨蒙,等. 计量标准测量能力若干关键技术考核方法的研究[J]. 内蒙古气象, 2022(05): 43 - 48.
- [5] 郑树芳,刘海珍,温晓辉,等. 精密露点仪期间核查方法的研究[J]. 现代电子技术, 2023,46(08): 164 - 168.
- [6] 郑树芳,刘海珍,温晓辉,等. 电子探空仪基测箱校准结果的不确定度评定方法的研究[J]. 计算机测量与控制, 2022,30(12): 339 - 347.
- [7] 国家质量监督检验检疫总局. JJF1094 - 2002 测量仪器特性评定[S]. 北京:中国计量出版社, 2002.
- [8] 蒋涛,孟宪罗,甄树勇. 气象用 PTB 系列气压传感器通用性的研究[J]. 计量与测试技术, 2016,43(10): 20 - 22 + 24.
- [9] 刘海珍,郑树芳,温晓辉,等. 气象用数字式气压计期间核查方法的研究[J]. 现代电子技术, 2022,45(24): 70 - 73.
- [10] 刘海珍,郑树芳,温晓辉,等. 关于自动气象站气压传感器的测量误差受环境温度影响的研究[J]. 现代电子技术, 2023,46(10): 21 - 25.
- [11] 李扬,刘明光,钱学成,等. 压力传感器的温度补偿研究及其应用[J]. 电测与仪表, 2016,53(18): 107 - 112 + 123.
- [12] 李亮,杜利东,赵湛,等. 数字气压传感器的一种高效标定和补偿方法[J]. 传感器与微系统, 2014,33(11): 44 - 47.
- [13] 国家质量监督检验检疫总局. JJG1084 - 2013 数字式气压计[S]. 北京:中国质检出版社, 2013.