DOI: 10.19650/j.cnki.cjsi.J2514198

基于 YOLO-MCSL 的轻量化智能电能表热缺陷目标检测方法*

陈方彬[1],赵仲勇[1],王 建[2],胡文杰[1],张开迪[3]
(1. 西南大学工程技术学院 重庆 400716; 2. 国网新疆电力有限公司电力科学研究院 新疆 830046;
3. 国网重庆市电力公司北碚供电分公司 重庆 400014)

摘 要:针对智能电能表及其接线盒热缺陷红外检测中存在的小目标漏检率高、复杂背景干扰严重及现有模型精度与效率难以兼顾等问题,提出了一种基于改进 YOLOv8s 架构的轻量化智能电能表目标检测算法 YOLO-MCSL,旨在满足电力现场巡检对实时检测的迫切需求。首先,将 MobileNetV4 轻量化网络作为骨干网,显著降低模型参数量与计算开销;其次,引入 RT-DETR 模型中的 CCFF 跨尺度特征融合模块,增强对多尺度微小热缺陷的感知能力;随后,设计轻量化 C2f_Star 模块替代原 C2f 结构,进一步压缩模型并提升特征提取效率;同时,构建 LSCD 轻量化共享卷积检测头,通过参数共享机制减少冗余计算;此外,结合 Focaler-SIoU 损失函数优化边界框回归过程,提升难易样本区分度;最后,应用层自适应幅值剪枝算法对模型进行结构化剪枝,实现性能与轻量化的平衡。基于自建的智能电能表热缺陷红外图像数据集开展实验,结果表明,在 3 类关键部件:接线盒、电池模块与显示屏的检测中,YOLO-MCSL 的检测精度分别达到 91.6%、99.2% 和 99.5%,整体 mAP@0.5 为 97.9%。相比 YOLOv8s 基准模型,参数量为 1.749 M,减少了 84.3%,计算量为 5.7 GFLOPs,降低了 80.2%,模型内存占用为 3.8 MB,减少了 82.3%。该方法为智能电能表缺陷检测提供了高精度、轻量化、可嵌入部署的解决方案,具备良好的工程应用前景。

关键词: YOLOv8s;智能电能表;热缺陷检测;轻量化
中图分类号: TH183.3 TM755 TP389 **文献标识码**: A **国家标准学科分类代码**: 470.40

A lightweight thermal defect detection method for smart electricity meters based on YOLO-MCSL

Chen Fangbin[1], Zhao Zhongyong[1], Wang Jian[2], Hu Wenjie[1], Zhang Kaidi[3]
(*1. College of Engineering and Technology, Southwest University, Chongqing 400716, China; 2. State Grid Xinjiang Electric Power Research Institute, Xinjiang 830046, China; 3. Beibei Power Supply Branch of State Grid Chongqing Electric Power Company, Chongqing 400014, China*)

Abstract: To address the issues of high miss rates for small targets, severe interference from complex backgrounds, and the inability of existing models to balance accuracy and efficiency in infrared detection of thermal defects in smart electricity meters and their junction boxes, we propose a lightweight smart electricity meter target detection algorithm, YOLO-MCSL, based on an improved YOLOv8s architecture. This algorithm aims to meet the urgent need for real-time detection in power field inspections. First, the MobileNetV4 lightweight network is adopted as the backbone to significantly reduce the number of model parameters and computational overhead. Second, the CCFF cross-scale feature fusion module from the RT-DETR model is introduced to enhance the detection capability for multi-scale small thermal defects. Subsequently, a lightweight C2f_Star module is designed to replace the original C2f structure, further compressing the model and improving feature extraction efficiency. Additionally, we construct the LSCD lightweight shared convolution detection head, which reduces redundant computation through parameter sharing. Furthermore, we combine the Focaler-SIoU loss function to optimize the bounding box regression process, enhancing the differentiation between easy and hard samples. Finally, we apply a layer-wise adaptive amplitude pruning algorithm to structurally prune the model, achieving a balance between performance and

收稿日期:2025-06-28 Received Date: 2025-06-28
*基金项目:中央高校基本科研业务费项目(SWU-KT22027)资助

lightweight design. Experiments were conducted on a self-constructed infrared image dataset of thermal defects in smart electricity meters. The results show that in the detection of three key components—junction boxes, battery modules, and displays—the detection accuracy of YOLO-MCSL reached 91.6%, 99.2%, and 99.5%, respectively, with an overall mAP@0.5 of 97.9%. Compared to the YOLOv8s baseline model, the number of parameters was reduced to 1.749 M (a reduction of 84.3 %), computational complexity was reduced to 5.7 GFLOPs (a reduction of 80.2%), and model memory usage was reduced to 3.8 MB (a reduction of 82.3%). This method provides a high-precision, lightweight, and embeddable solution for smart electricity meter defect detection, showing promising prospects for engineering applications.

Keywords: YOLOv8s; smart meters; thermal defect detection; lightweight

0　引　　言

随着国民经济的高速发展,社会用电需求持续增长,对电能计量装置的可靠性和准确性要求日益提高。目前居民智能电能表由于数量庞大、运行环境复杂且质量参差不齐,故障率相对较高。此外,电能计量的准确性对于维持电力系统稳定运行以及保证相关产业的经济效益具有重要意义[1-2]。因此需确保智能电能表长期稳定工作。然而,因安装质量不达标、接线错误、运行老化等因素影响导致智能电能表及接线盒的内外部易产生缺陷,随着设备的持续运行,可能会诱发发热故障,严重影响电能计量的精准性、可靠性和安全性。尤其在近年来夏季持续高温的外部环境影响叠加下,其热缺陷风险更加突显。当前针对智能电能计量回路设备热缺陷的检测,主要依赖人工红外测温并进行目视检查。该方法易忽略缺陷的异常特征,导致漏判引起的电能表故障和电能计量异常,甚至发展为设备起火等严重事故。

针对上述问题,近年来,有学者采用机器学习等人工智能技术对电气设备进行缺陷检测,取得了不错的效果[3-4]。然而,针对智能电能表、接线盒等计量回路关键设施的热缺陷检测的研究仍相对缺乏。另外,随着深度学习技术发展,以 YOLO(you only look once)系列为代表的目标检测算法逐步应用到电气设备缺陷检测方面,如绝缘子破损判断[5]、输电线路鸟害识别[6]、变电站安防等方面[7]。部分学者使用深度学习算法对仪表设备进行故障检测,但由于该类算法计算复杂度高、网络参数量大,往往需要借助高性能图形处理器(graphics processing unit, GPU)运算平台来进行实时检测[8],因此难以满足移动端或边端部署快速计算设备的需求。而轻量级的目标检测算法在保留较高准确率的同时具有较小的网络参数量,为智能电能表及接线盒热缺陷的检测提供了快速、精准的本地化部署能力。因此,研究智能电能表热缺陷的轻量化目标检测技术至关重要。

目前,针对仪器仪表图像的目标检测研究,已有轻量级目标检测算法的相关报道。崔昊杨等[9]提出一种基于轻量级 EF-YOLOv4 网络的电力仪表图像目标检测模型。该模型融入最近邻快速特征匹配方法并通过单位符号特征细粒度检测仪表目标。邬开俊等[10]提出了一种结合 FasterNet-tiny 和 YOLOv5s 网络模型改进的缺陷快速检测算法 FasterNet-YOLOv5,使网络模型更精确地定位目标缺陷区域。毛爱坤等[11]使用 ShuffleNet V2 替换 YOLOv5 的主干网络,在降低模型计算复杂度的同时提高了检测速度。何永春等[12]在 Faster R-CNN(faster region-based convolutional neural networks)的基础上提出一种基于空间注意力机制的多尺度仪表检测算法,以解决实际应用中仪表背景复杂的问题。上文大多是针对仪表其他缺陷的目标检测,正如前文所述,针对智能电能表及接线盒热缺陷的红外目标检测研究仍存在显著空白,且当前热缺陷检测面临两大难题:1)红外图像中热缺陷目标尺度变化大,要求模型兼具多尺度特征捕捉能力;2)电力现场需在有限算力下实现实时响应。现有方法难以平衡轻量化与高精度的双重需要。

为解决智能电能表及其附属接线盒热缺陷检测中存在的目标多样、运行环境复杂、边缘设备算力受限等实际问题,故提出了一种基于改进 YOLOv8s 架构的轻量化目标检测算法 YOLO-MCSL。整体思路是,通过“全链路”轻量化设计,在显著减少模型参数量和计算量的同时,保障检测精度,满足电力现场实时检测需求。具体而言,YOLO-MCSL 引入 MobileNetV4 作为骨干网络大幅压缩模型体积,结合 RT-DETR(real-time detection transformer)模型中的跨尺度特征融合模块(cnn-based cross-scale feature fusion, CCFF)提升复杂场景下的特征提取与多尺度感知能力,利用 C2f_Star 模块和轻量化共享卷积检测头(lightweight shared convolutional detection head, LSCD)进一步降低计算消耗并增强定位效果,采用 Focaler-SIoU 损失函数优化正负样本区分与边框回归精度,最后通过层自适应幅值剪枝(learning-aware magnitude-based pruning, LAMP)实现极致模型压缩。多项创新模块共同作用,使 YOLO-MCSL 在保证轻量化的同时,有效保持了热缺陷检测的准确率。实验结果表明,本方法在模型体积、参数和推理速度等方面优于对比算法,能够有效解决电力现场因算力有限而无法进行实时热缺陷检测的难题,具有重要的实际应用价值。

1 材料与方法

1.1 电能计量设备热缺陷数据获取

训练网络模型所用数据来源于所搭建的实验平台拍摄的数据集,实验平台如图 1 所示。

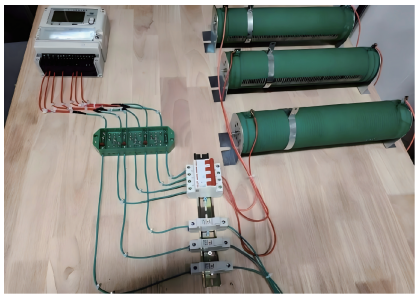
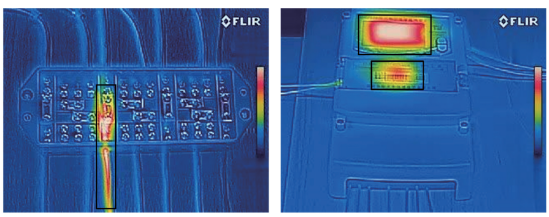


图 1 实验平台

Fig. 1 Experimental platform

经研究调查后,发现接线盒[13-14]、智能电能表显示屏处的热缺陷较为常见[15-16],但鉴于作为智能电能表电源的锂电池在长期充放电或高温环境下会提升其失效概率[17],因此以接线盒、智能电能表电池以及显示屏 3 种位置热缺陷作为检测目标。拍摄设备为 FLIR-T440 红外摄像仪,共包括 701 张计量设备热缺陷图像,所采集图像大小为 320 pixels×240 pixels。

自建的红外图像数据集主要包含 3 类位置的热缺陷:接线盒、智能电能表电池以及显示屏,分别如图 2 中矩形框标注所示。所用智能电能表的型号包括:DSZ1122、DSZ22、DSZ178、DTZY9599-Z、DTZ545、DTSD148。模拟热缺陷的方法分别为:通过人工干预松动接线盒连接处的螺栓来模拟接线盒热缺陷、在利用开关电源供电情况下将热贴片分别贴于电能表的显示屏和电池位置来模拟智能电能表发生相应位置故障时的情况。



(a) 接线盒处热缺陷 (b) 电池与显示屏处热缺陷

(a) Terminal_defect at junction box (b) Thermal defect at battery and screen

图 2 3 种不同类型的缺陷

Fig. 2 Three different types of defects

1.2 数据增广与预处理

采用旋转、平移、镜像、亮度调整、增加噪声的方式进行数据增强并得到 6 010 张图像作为数据集。同时在训练目标检测模型时采用 Mosaic 数据增强方法,通过对热缺陷图片通过随机缩放、裁剪及排布、色域变化中的一种或多种方法进行拼接[18],使网络模型能够学习到多样的场景和目标组合,有助于其更好地泛化到真实世界中的复杂场景,从而达到优化训练效率和数据利用率的效果。

1.3 建立图像数据库与标签数据库

使用目标检测标注工具 Labelimg 对拍摄图像中的接线盒接线柱处、智能电能表电池处以及显示屏处 3 种热缺陷进行了手工标注矩形框,标签名分别为 Terminal_fault、Battery_fault、Screen_fault。标注完成后按照 8∶1∶1 划分训练集、验证集和测试集。最终得到训练集、验证集和测试集样本数分别为:4 808、601 和 601 张。图 3 为热缺陷数据集中包括的种类和对应标签信息的可视化结果。

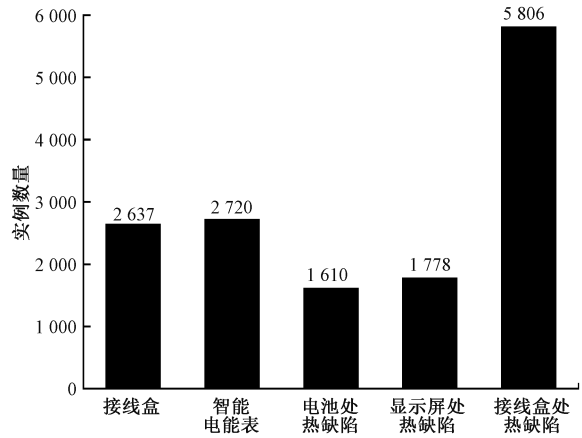


图 3 热缺陷数据集信息可视化

Fig. 3 Visualization of thermal defect dataset

2 算法原理

2.1 YOLO-MCSL 整体网络结构

于 2023 年被正式提出的 YOLOv8s 网络模型[19],具有优秀的检测性能,但存在不适用于边端设备部署、对复杂背景或遮挡场景的鲁棒性有限、检测速度较慢等难点。为解决上述问题,故提出一种轻量化的 YOLO-MCSL 网络模型,其结构如图 4 所示,主要由 Backbone、Neck 和 Head 这 3 部分组成。

YOLO-MCSL 网络模型主要对原 YOLOv8s 网络进行了 5 个不同部分的改进,即:

1)使用 MobileNetV4 作为主干网络提取特征。

2)将 Neck 部分替换为 RT-DETR 模型中的 CCFF 模块。

3)结合星形运算的思想,将颈部的 C2f 模块替换 C2f_Star 模块。

4)将原有的检测头替换为 LSCD 检测头。

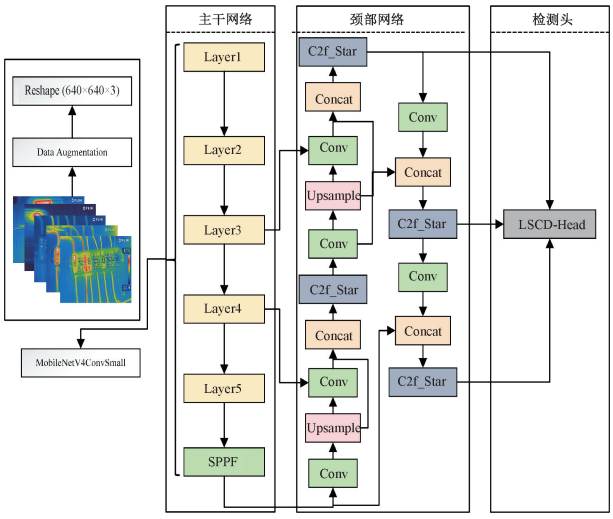


图 4　改进 YOLOv8s 结构

Fig. 4　Improved structure of YOLOv8s

5)将原本的 CIoU 损失函数替换为 Focaler-SIoU。

YOLO-MCSL 网络模型的工作流程为:

1)数据预处理阶段:对采集的红外热图像进行旋转、平移、镜像等几何变换,并调整亮度、添加噪声,通过 Mosaic 方法将图像随机拼接增强,提升模型对不同场景的适应能力。

2)特征提取阶段:采用轻量化 MobileNetV4 作为骨干网络,利用其深度可分离卷积(depthwise separable convolution, DSC)和通用反转瓶颈(universal inverted bottleneck, UIB)结构高效提取多尺度特征,在减少计算量的同时保持特征表达能力。

3)多尺度特征融合阶段:通过 RT-DETR 的 CCFF 模块处理骨干网络输出的特征,利用卷积块融合相邻尺度特征,增强模型对不同尺寸热缺陷的检测能力。同时,利用轻量化的 C2f_Star 模块,进一步降低参数量并提升特征表达能力。

4)目标定位与分类阶段:使用 LSCD 检测头处理融合后的图像特征,通过参数共享机制减少计算量,分离回归分支和分类分支,分别预测边界框坐标和类别概率。

5)后处理优化阶段:采用 Focaler-SIoU 损失函数优化边界框回归,关注不同难度的样本;基于 LAMP 剪枝去除冗余连接,在 50% 剪枝率下显著减小模型体积。

6)结果输出阶段:最终输出热缺陷的类别、位置及置信度,为后续的目标分析提供定量依据,便于进一步评估和处理。

2.2　轻量级 MobileNetV4 网络模型

由于原 YOLOv8s 网络的参数量与计算量较大,不利于对目标设备热缺陷开展实时检测。因此,将 MobileNetV4 轻量化网络[20]作为 YOLOv8s 的骨干网络。其核心理念在于利用深度可分离卷积替换传统卷积。图 5 展示了深度可分离卷积的处理流程。其在减少网络计算开销的同时,仍能有效提取空间特征和通道特征,保持较高的特征表达能力,为在移动设备等资源受限的环境中部署提供关键优势。

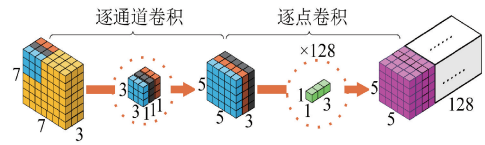


图 5　深度可分离卷积

Fig. 5　Depthwise separable convolution

MobileNetV4 是 MobileNet 系列的最新版本,是专为移动设备设计并引入了多种新颖且高效的架构组件。它的关键之处在于 UIB。而 UIB 是一个灵活的架构单元,其扩展和增强了 MobileNetV2 的反转瓶颈模块(inverted bottleneck, IB),将深度可分离卷积与逐点卷积结合起来,增强了灵活性和计算效率,同时融合了纯卷积神经网络、前馈网络。这是一种适用于高效网络设计的可调节模块,能够适应各种优化目标,且不会使搜索复杂度激增。在每个网络阶段,具有灵活性的 UIB,不但可以在空间和通道混合之间进行临时权衡、根据需要扩大感受野还能最大化计算利用率。通用反转瓶颈模块的结构图如图 6 所示。

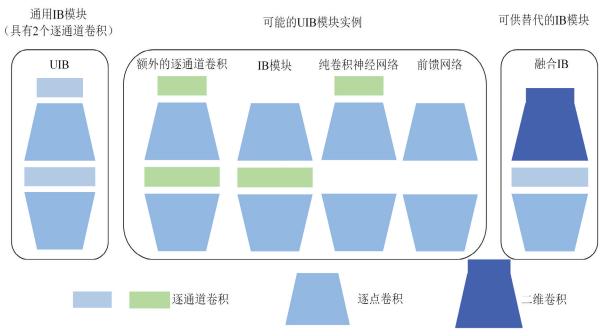


图 6　通用反转瓶颈模块

Fig. 6　Universal inverted bottleneck block

为了优化移动加速器的性能,MobileNetV4 还包括一种多查询注意力机制(multi-query attention, MQA)。该机制通过优化算术运算与内存访问的比率,显著提高了移动加速器上的推理速度。MQA 的核心计算式如式(1)和(2)所示。

$$Mobile\_MQA(\boldsymbol{X}) = Concat(A_1, \cdots, A_n)\boldsymbol{W}^o \tag{1}$$

$$A_j = \mathrm{Softmax}\left(\frac{(\boldsymbol{X}\boldsymbol{W}^{Q_j})(SR(\boldsymbol{X})\boldsymbol{W}^K)^{\mathrm{T}}}{\sqrt{d_k}}\right)(SR(\boldsymbol{X})\boldsymbol{W}^V) \tag{2}$$

式中:*SR* 表示空间缩减,即设计中步幅为 2 的深度可分离卷积,或者在不使用空间缩减的情况表示恒等函数,这种计算方法有效地利用了非对称空间子采样,提高了运算效率。

为让算法在保持准确率的情况下提升实时响应能力,选用 MobileNetV4 系列中参数量最小的 MNv4-Conv-S 版本替换 YOLOv8s 模型中的主干网络。

2.3 CCFF 跨尺度特征融合模块

如何处理多尺度特征是目标检测的难点之一,YOLOv8s 颈部采用 PANet(path aggregation network)[21]来提高特征表达能力,但这也使网络模型的计算复杂度和内存占用增加。

Zhao 等[22]提出了 Transformer 实时检测模型 RT-DETR。作为第 1 个实时端到端对象检测器,其设计了一种基于 CNN 的跨尺度特征融合模块,通过跨尺度融合快速处理多尺度特征,将特征转换为图像特征序列,从而显著提高推理速度。

为使网络模型在保持较高运算效率的同时,提升对不同大小目标的识别能力。采用 CCFF 模块来替代 YOLOv8s 中的特征融合模块。CCFF 模块在原网络模型的特征融合路径中插入若干个由卷积层组成的融合块,用于融合相邻尺度的特征。其结构如图 7 所示。

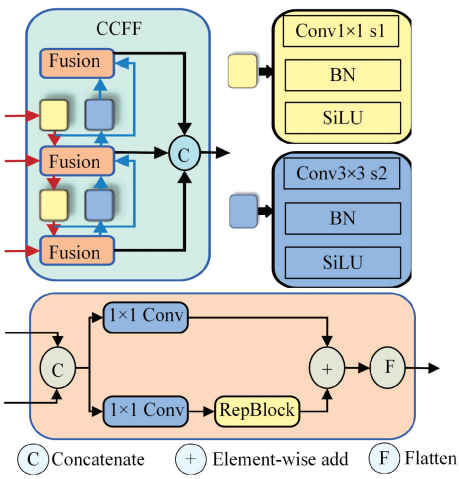


图 7 CCFF 模块结构

Fig. 7 CCFF module structure

2.4 C2f_Star 模块

由于现实世界图像数据的复杂性,高维和非线性特征在传统机器学习算法[23-24]和深度学习网络[25-28]中都至关重要。然而,在原 YOLOv8s 网络模型中,传统的标准卷积操作虽然能对上述特征进行有效提取,但计算成本较高。因此为了进一步加速模型训练,减少模型参数,提高对智能电能表以及接线盒热缺陷的检测效果,故结合星形运算的思想,将模型颈部中原本的 C2f 模块修改为 C2f_Star 模块。该修改环节可扩大网络的维度,降低网络扩大的增量效益,缩减模型复杂度,提高模型对目标的识别和收敛效果。C2f_Star 模块由 n 个子模块构成,其中子模块的处理流程如图 8 所示。

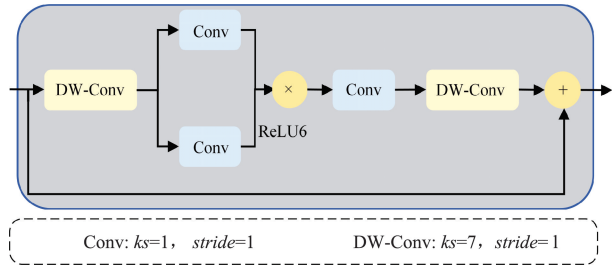


图 8 C2f_Star 子模块结构

Fig. 8 C2f_Star submodule structure

Ma 等[29]发现可以通过逐元素乘法融合不同的子空间特征。他们将这种范例称为“星形运算”(由于逐元素乘法符号类似星形)。将权重矩阵和偏差合并为一个统一的实体,记为:$\boldsymbol{W}=\begin{bmatrix}W\\B\end{bmatrix}$,$\boldsymbol{X}=\begin{bmatrix}X\\1\end{bmatrix}$,因此在单层神经网络中,星形运算如式(3)所示。

$$(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{X})\times(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{X})=\sum_{i=1}^{d+1}\sum_{j=1}^{d+1}w_1^i w_2^j x^i x^j \tag{3}$$

式中:在单元素特征输入、单输出通道变换的情况下,可以定义:w_1、w_2、$x\in R^{(d+1)\times 1}$,d 为输入通道数。扩展到多元素特征输入和多输出通道变换的情况下,则可以定义为:$\boldsymbol{W}_1$、$\boldsymbol{W}_2\in R^{(d+1)\times(d'+1)}$,$\boldsymbol{X}\in R^{(d+1)\times n}$。改写式(3)中描述的星型运算后,可以将其扩展为 $(d+1)(d+2)/2$ 个不同项之和,如式(4)所示。

$$\underbrace{\alpha(1,1)x^1x^1+\cdots+\alpha(3,4)x^3x^4+\cdots+\alpha(d+1,d+1)x^{d+1}x^{d+1}}_{(d+1)(d+2)/2\text{项}} \tag{4}$$

可以发现,除 $\alpha(d+1,:)x^{d+1}x$ 外的每一项都与 x 成非线性相关,表示独立的隐式维度。因此在一个 d 维空间中使用星形运算可以有效地生成一个维数为 $(d+1)(d+2)/2\approx d^2/2$(考虑 $d\gg 2$)的隐式高维特征空间。因此当星形运算被整合到神经网络中并通过多层堆叠时,星形运算仅需少数几层,便能在紧凑的特征空间内实现近乎无限的维度扩展,从而显著提升网络的表达能力和计算效率。

2.5 LSCD 轻量化检测头

YOLOv8s 的检测头采用目前主流的将分类与检测头分离的解耦头结构,让模型在检测性能上得到了显著提升,但参数量的大幅增加可能限制模型在实际应用中的检测效果。因此借助轻量化共享卷积检测头 LSCD[30],使模型在保持较高检测精度的同时,减少参数量和计算量。

LSCD 检测头的结构如图 9 所示,在检测头部分,从颈部输出的 3 个特征层首先通过 1×1 卷积层进行通道数调整,将其统一为中间层通道数,以实现特征维度的标准化。

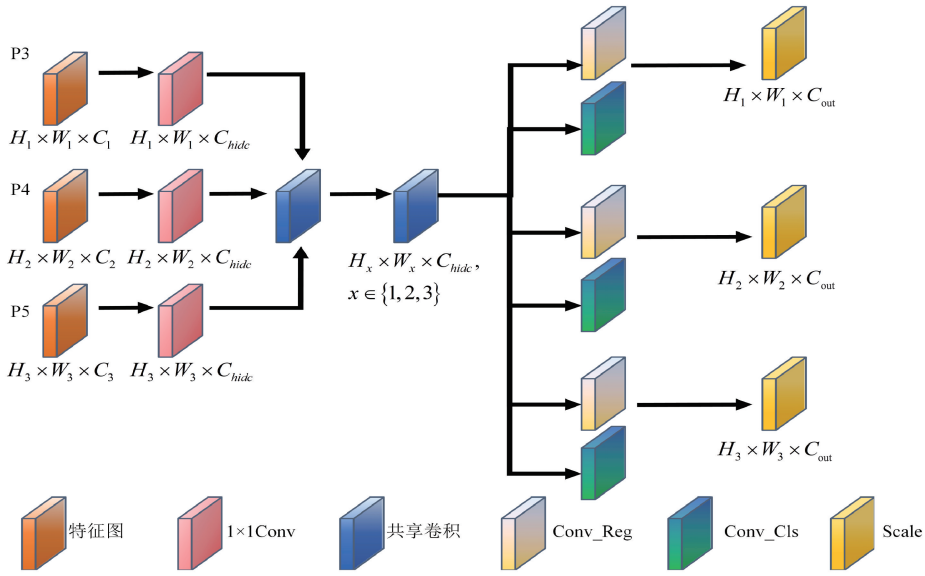


图 9　LSCD 结构

Fig. 9　LSCD structure

随后,所有特征层被输入到一个共享卷积模块中进行特征提取,该模块采用 3×3 卷积核,通过参数共享机制有效减少了模型的参数量和计算复杂度。最后检测头将回归分支和分类分支分离处理:在回归分支中,使用 1×1 卷积层预测边界框的坐标偏移量,并通过引入 Scale 层对输出特征进行动态缩放,以适配不同检测头所对应的目标尺度差异,从而实现对多尺度目标的精确定位;在分类分支中,同样采用 1×1 卷积层预测每个类别的概率分布。回归分支和分类分支的卷积层权重相互独立,使模型能够分别优化定位和分类任务,进一步提升检测性能。通过共享权重参数的设计,模型能够处理多尺度特征以增强对图像中物体间关系的理解能力,还能提高本身在复杂场景下的适用性。

2.6　Focaler-SIoU 损失函数

边界框回归在目标检测领域起着至关重要的作用,目标检测的定位精度很大程度上取决于边界框回归的损失函数。从目标尺度分析的角度来看,一般的检测目标可以视为简单样本,而极小的目标由于其精确定位困难,可以视为困难样本。现有研究利用边界框之间的几何关系来提高回归性能,而忽略了难易样本分布对边界框回归的影响。

YOLOv8 原本的损失函数为 CIoU[31],它的定义如式(5)所示。

$$CIoU = IoU - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v \tag{5}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{6}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{7}$$

式中:$\rho^2(b,b^{gt})$ 表示预测框和真实框的中心点的欧氏距离;c 为最小外接矩形的对角线距离;v 是度量长宽比相似性的修正因子;w^{gt} 和 h^{gt} 表示 GT box(ground truth box)的宽度和高度;w 和 h 表示锚框的宽度和高度,尽管 CIoU 具有较好的鲁棒性,但其却存在计算复杂、参数敏感等缺点。

为了能够关注不同的回归样本专注于不同的检测任务,Zhang 等[32]使用线性区间映射方法来重建 IoU 损失,从而改进边缘回归。分析了难易样本分布对回归结果的影响,然后提出了 Focaler-IoU,通过关注不同的回归样本,可以提高检测器在不同检测任务中的性能,如式(8)所示。通过动态调节参数 d 和 u 的取值,Focaler-IoU 能够自适应地聚焦于不同难易程度的回归样本,从而优化模型的回归性能。

$$IoU^{Focaler} = \begin{cases} 0, & IoU < d \\ \dfrac{IoU - d}{u - d}, & d \leqslant IoU \leqslant u \\ 1, & IoU > U \end{cases} \tag{8}$$

其损失定义如式(9)所示。

$$L_{Focaler-IoU} = 1 - IoU^{Focaler} \tag{9}$$

故借助 Focaler-SIoU 损失函数,并将用其替代了原本的 CIoU 损失函数。

$$L_{Focaler-SIoU} = L_{SIoU} + IoU - IoU^{Focaler} \tag{10}$$

2.7 基于 LAMP 方式的模型剪枝

在深度学习研究中,模型剪枝技术作为一种重要的优化手段,可以显著提升模型的效率和可部署性。通过对神经网络进行剪枝,可以有效去除结构中冗余且资源密集的部分,从而在维持模型精度的同时,大幅压缩模型规模并提升推理速度。

采用基于层自适应幅值的剪枝算法[33]来进行剪枝。与基于幅值的剪枝[34](magnitude pruning, MP)和许多传统剪枝方法不同,LAMP 是一种全局剪枝,它克服了以往全局剪枝时发生层崩溃的问题,LAMP 定义的权重重要性评估分数式如式(11)和(12)所示。

$$score(u;\boldsymbol{W}) = \frac{(\boldsymbol{W}[u])^2}{\sum_{v\geqslant u}(\boldsymbol{W}[v])^2} \tag{11}$$

$$(\boldsymbol{W}[u])^2 > (\boldsymbol{W}[v])^2 \Rightarrow score(u;\boldsymbol{W}) > score(v;\boldsymbol{W}) \tag{12}$$

式中:$\boldsymbol{W}$ 表示被展平为一维的权重张量,$\boldsymbol{W}[u]$ 就表示由索引 u 映射的 $\boldsymbol{W}$ 中的元素。LAMP 假设权重按照 u 和 v 的索引映射以升序排序。分子部分 $(\boldsymbol{W}[u])^2$ 表示目标连接权重的幅值平方;分母部分 $\sum_{v\geqslant u}(\boldsymbol{W}[v])^2$ 表示同一层中所有保留连接的权重幅值的平方和。

根据上述分析,权重项的数值与其对应的 LAMP 评分呈正相关,即权重值越大,LAMP 评分越高。因此,LAMP 评分较低的权重项通常被视为冗余部分,在剪枝过程中会被优先移除。基于层自适应幅值的剪枝算法 LAMP 的具体操作过程如图 10 所示。

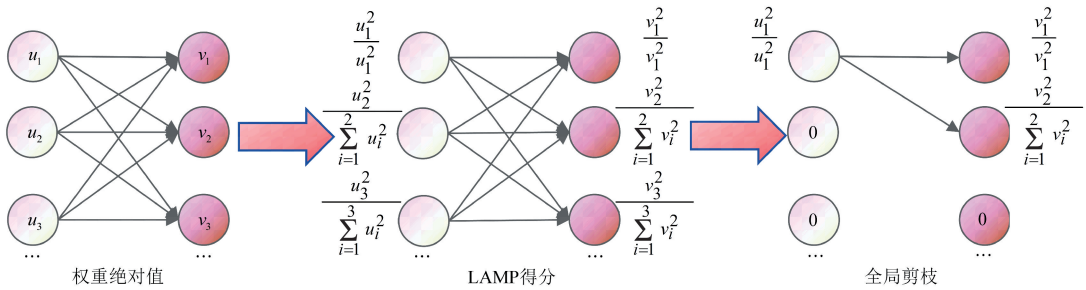


图 10 基于层自适应幅值的剪枝过程

Fig. 10 Pruning process based on layer-adaptive magnitude

3 实验结果与分析

3.1 实验环境与评价指标

实验环境为:操作系统采用 Ubuntu18.04 版本,硬件平台搭载 AMD EPYC 9754 128 核处理器(18vCPU)并配备 60 GB 运行内存,计算设备为 NVIDIA RTX 4090D 显卡(24 GB 显存)。编程环境基于 Python 3.8、Pytorch 2.0.0,并配置 CUDA 11.8 以支持 GPU 加速运算。在模型训练环节,基础训练与剪枝后微调均采用相同的超参数设置:训练轮次(epoch)固定为 120,批量大小(batch size)设为 16。优化器选用随机梯度下降法(stochastic gradient descent, SGD),动量参数(momentum)配置为 0.937。学习率调度策略采用周期性调整机制,并引入 Warm-Up 预热方法,初始学习率设定为 0.01,同时设置权重衰减系数为 0.000 5。

为全面评估模型性能,现从 3 个维度构建评价体系:首先,在检测精度方面,采用召回率(recall,R)、查准率(precision,P)、平均精度(average precision,AP)及其均值(mean average precision,mAP)作为量化指标;其次,通过帧率(frames per second, FPS)衡量模型的检测速度;最后,从网络复杂度角度,选取模型大小、参数量和浮点计算量作为评估指标,这些指标的数值与网络复杂度呈正相关关系。

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$AP = \int_0^1 Precision(r)\,\mathrm{d}r \tag{15}$$

$$mAP = \frac{\sum_{i=0}^{N} AP_i}{N} \tag{16}$$

3.2 不同模型检测结果对比

为了验证改进算法的有效性,在相同实验环境与参数设置的情况下,将 YOLO-MCSL 算法与 YOLOv8n、YOLOv8s、YOLOv8m、RT-DETR 和 Faster-RCNN 算法进行对比实验,实验结果如表 1 所示。

总体来看,YOLO-MCSL 算法在具有较高检测精度与速度的同时,模型大小显著降低,尤其是在处理智能电能表热缺陷这类复杂背景下的红外图像时,YOLO-MCSL 展现出更高的鲁棒性和适应性,有利于在实际工程中的应用。

表 1　常见模型对比

Table 1　Comparison of different models

模型	模型大小/MB	*mAP*50/%	*mAP*50:95/%	*FPS*
YOLOv8n	6.0	0.986 0	0.765	2 260.5
YOLOv8s	21.5	0.988 0	0.780	1 062.3
YOLOv8m	52.0	0.988 0	0.798	405.2
RT-DETR-l	56.3	0.990 0	0.748	179.8
RT-DETR-x	129.1	0.990 0	0.747	116.5
Faster-RCNN	108.3	0.990 1	0.610	7.2
YOLO-MCSL	7.5	0.988 0	0.771	1 451.0

3.3　消融实验

为了充分验证 YOLO-MCSL 的有效性,并系统评估各改进模块对基准模型的性能贡献,现采用模块化验证策略,将整体改进方案分解为 5 个组成部分进行对比实验:1)使用 MobileNetV4 替代 CSPDarkNet53 作为主干特征提取网络;2)引入 CCFF 特征融合模块;3)使用 C2f_Star 模块替换原有的 C2f 模块;4)使用 LSCD 检测头替代原来的检测头;5)引入 Focaler-SIoU 损失替换原损失 CIoU。由表 2 消融实验结果的可知,这些改进策略在模型架构设计上具有合理性和有效性,YOLO-MCSL 算法通过 5 个渐进式改进模块的引入,在目标检测性能上呈现出显著的提升趋势,验证了该算法在实际应用场景中的实用价值。

表 2　消融实验

Table 2　Ablation experiments

实验编号	改进策略					浮点计算量/GFLOPs	参数量/M	模型大小/MB	*mAP*50/%	*FPS*
	MobileNetV4	CCFF	C2f_Star	LSCD	Focaler-SIoU					
A						28.4	11.128	21.5	0.988	1 076.5
B	√					8.0	4.299	8.5	0.983	1 799.5
C		√				23.0	7.259	14.1	0.987	1 234.6
D	√	√				13.5	4.789	9.5	0.987	1 563.5
E	√	√	√			12.9	4.668	9.2	0.988	1 410.3
F	√	√	√	√		11.4	3.790	7.5	0.985	1 431.4
G	√	√	√	√	√	11.4	3.790	7.5	0.988	1 436.1

3.4　模型剪枝实验

剪枝率是模型压缩的关键参数,直接影响权重保留比例。虽然提高剪枝率能加速推理并压缩模型,但也会导致性能下降。故测试了不同剪枝率下的模型参数量和精度(如表 3 所示),剪枝策略后的数值(如 1.5、1.9)表示剪枝前后计算量的比值。以 LAMP2.0 为例,表示采用 LAMP 剪枝且剪枝率为 50%。实验显示,随着剪枝率增加,参数量减少,但 mAP 下降。经权衡,最终选择 50% 的剪枝率,在保证性能的同时显著压缩了模型体积。

表 3　不同剪枝策略对改进模型的效果对比

Table 3　Comparison of effects of different pruning strategies on the improved model

剪枝策略	浮点计算量/GFLOPs	参数量/M	模型大小/MB	*mAP*50/%	*FPS*
LAMP1.5	7.6	2.143	4.6	0.989	1 580.5
LAMP1.6	7.1	2.032	4.4	0.989	1 590.3
LAMP1.7	6.6	1.935	4.2	0.988	1 603.1
LAMP1.8	6.3	1.868	4.0	0.987	1 612.5
LAMP1.9	5.9	1.809	3.9	0.986	1 618.7
LAMP2.0	5.7	1.749	3.8	0.979	1 634.1
LAMP2.2	5.2	1.668	3.6	0.936	1 642.3
LAMP2.4	5.1	1.663	3.6	0.657	1 650.7

3.5　实验结果可视化

将 YOLO-MCSL 算法与目前流行的目标检测算法 YOLOv5s、YOLOv11、Faster R-CNN、RTDETR-l 进行对比实验。在相同条件下,各对比模型均通过自建的智能电能表热缺陷红外图像数据集进行训练。对比实验结果如表 4 所示。

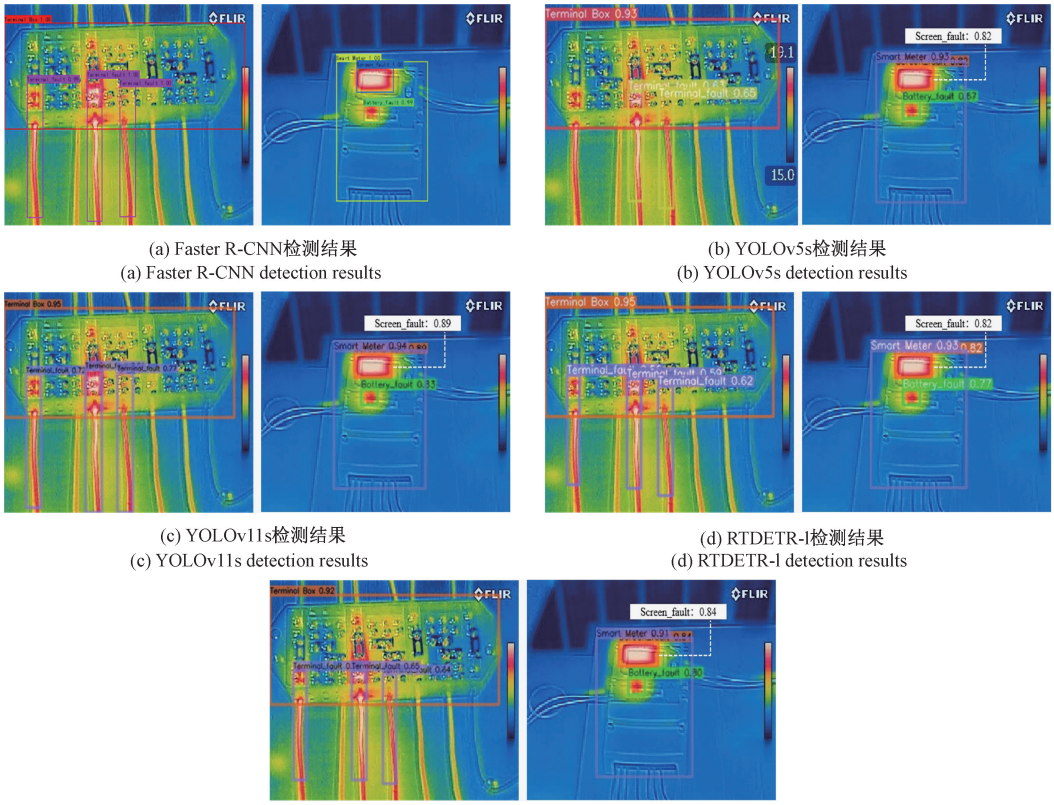
图 11 展示了各算法对测试图像的可视化检测效果,其中 YOLOv5s 存在漏检现象,其余算法均检测、分类准确。在密集目标场景中,YOLO-MCSL 相较于基准模型展现出更强的鲁棒性,能够有效区分重叠目标并保持较高的检测置信度。

对比实验结果表明,虽然 YOLO-MCSL 算法的检测精度略逊于其余算法,但检测速度、参数量、浮点运算数和模型大小指标上均显著优于 Faster R-CNN、YOLOv5s、YOLOv11s 和 RTDETR-l 算法,证明该方法具有良好的表现性能,使得模型更适合部署在算力有限的边缘设备上,能够满足实际智能电能表故障快速检测智能化发展的需求。

表 4 对比实验结果

Table 4 Results of comparative experiment

模型	*mAP*50:95/%	*mAP*50/%	*FPS*	参数量/M	浮点计算量/GFLOPs	模型大小/MB
Faster R-CNN	0.662	0.994	48.24	137.099	370.21	522.99
YOLOv5s	0.783	0.988	718.05	9.113	23.80	17.70
YOLOv11s	0.797	0.990	999.70	9.415	21.30	18.30
RTDETR-l	0.740	0.988	180.90	28.454	100.60	56.30
YOLO-MCSL	0.720	0.979	1 634.10	1.749	5.70	3.80



(a) Faster R-CNN检测结果
(a) Faster R-CNN detection results

(b) YOLOv5s检测结果
(b) YOLOv5s detection results

(c) YOLOv11s检测结果
(c) YOLOv11s detection results

(d) RTDETR-l检测结果
(d) RTDETR-l detection results

(e) YOLO-MCSL检测结果
(e) YOLO-MCSL detection results

图 11 检测结果可视化

Fig. 11 Visualization of detection results

4 结 论

1)针对目前未从智能电能表以及接线盒的热缺陷进行智能检测的角度开展相关研究,提出一种处理红外热成像图像的轻量级 YOLO-MCSL 目标检测方法。

2)通过 MobileNetV4 轻量级骨干网络提升模型对目标的特征提取能力;通过 CCFF 跨尺度特征融合模块来高效处理多尺度特征;通过 C2f_Star 模块降低网络的参数量与计算量;通过 LSCD 检测头利用共享卷积的思想在保持一定精度的情况下缩减网络大小;通过 Focaler-SIoU 损失函数提高检测器在不同检测任务中的性能;利用 LAMP 方式对 YOLO-MCSL 进行剪枝处理降低模型参数量。

3)实验结果表明,YOLO-MCSL 算法在检测精度方面达到 97.9%,与原始 YOLOv8s 网络相比,精度仅下降 0.9 个百分点。但网络参数量减少至原模型的 15.7%,计算量降低至 19.8%,同时推理速度提升 51.8%。该模型可部署于移动端设备,解决因算力不足而无法进行实时热缺陷检测的难题。

4)后期将研究如何通过自主化学习方法使该算法具备自主学习、自我更新的能力,同时丰富能够进行热缺陷

检测的电能计量设备种类,进一步实现电能计量设备的边端智能检测。

5)后续研究计划将本模型部署于移动端及嵌入式设备,实现智能电能表的实时热缺陷检测。在硬件选型方面,将综合考虑计算单元(CPU/GPU/TPU)性能、内存带宽和功耗等关键指标,优先选择已成功部署同类 YOLOv8 模型的嵌入式平台(如 Jetson 系列)。通过对比分析本模型与现有移动端检测模型的参数量、计算复杂度和推理速度等指标,验证本模型在同等硬件条件下的运行可行性。同时,将采用 TensorRT 量化压缩和 CUDA 并行计算等技术优化推理效率,确保满足实际应用场景的实时性要求。

参考文献

[1] ZHAO ZH Y, CHEN Y, LIU J N, et al. Evaluation of operating state for smart electricity meters based on Transformer-Encoder-BiLSTM[J]. IEEE Transactions on Industrial Informatics, 2023, 19(3): 2409-2420.

[2] 覃玉红, 唐求, 邱伟, 等. 多应力下电能计量设备基本误差预估[J]. 仪器仪表学报, 2022, 43(4): 18-25.
QIN Y H, TANG Q, QIU W, et al. Estimation of basic error of electric energy metering equipment under multiple stress[J]. Chinese Journal of Scientific Instrument, 2022, 43(4): 18-25.

[3] CHEN Y, ZHAO ZH Y, YU Y Q, et al. Improved interpretation of impulse frequency response analysis for synchronous machine using lifelong learning based on iCaRL[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-10.

[4] 仲林林, 吴奇, 叶俊杰, 等. 基于元学习的变电设备小样本缺陷图像检测[J]. 仪器仪表学报, 2024, 45(10): 154-167.
ZHONG L L, WU Q, YE J J, et al. Meta-learning-based few-shot image detection of defects in substation equipment[J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 154-167.

[5] 刘熹, 陈晨, 双丰. 基于改进 YOLOv7-tiny 的多种类绝缘子检测算法[J]. 仪器仪表学报, 2024,45(9): 101-110.
LIU X, CHEN CH, SHUANG F. Multi-type insulator detection algorithm based on improved YOLOv7-tiny[J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 101-110.

[6] 裴少通, 张行远, 胡晨龙, 等. 基于 ER-YOLO 算法的跨环境输电线路缺陷识别方法[J]. 电工技术学报, 2024, 39(9):2825-2840.
PEI SH T, ZHANG X Y, HU CH L, et al. The defect detection method for cross-environment power transmission line based on the ER-YOLO algorithm[J]. Transactions of China Electrotechnical Society, 2024, 39(9): 2825-2840.

[7] 马鹏, 樊艳芳. 基于深度迁移学习的小样本智能变电站电力设备部件检测[J]. 电网技术, 2020, 44(3): 1148-1159.
MA P, FAN Y F. Detection of power equipment components in small sample smart substation based on deep migration learning[J]. Power System Technology, 2020, 44(3): 1148-1159.

[8] OH S, PARK N, SAEL L, et al. Scalable tucker factorization for sparse tensors-algorithms and discoveries[C]. 2018 IEEE 34th International Conference on Data Engineering, 2018: 1120-1131.

[9] 崔昊杨, 张雨阁, 张驯, 等. 基于边端轻量级网络的电力仪表设备检测方法[J]. 电网技术, 2022, 46(3): 1186-1193.
CUI H Y, ZHANG Y G, ZHANG X, et al. Detection of power instruments equipment based on edge lightweight network[J]. Power System Technology, 2022, 46(3): 1186-1193.

[10] 邬开俊, 徐泽浩, 单宏全. 基于 FasterNet 和 YOLOv5 改进的玻璃绝缘子自爆缺陷快速检测方法[J]. 高电压技术, 2024, 50(5): 1865-1876.
WU K J, XU Z H, SHAN H Q. Improved rapid detection method for self-exploding defects in glass insulators based on FasterNet and YOLOv5[J]. High Voltage Engineering, 2024, 50(5): 1865-1876.

[11] 毛爱坤, 刘昕明, 陈文壮, 等. 改进 YOLOv5 算法的变电站仪表目标检测方法[J]. 图学学报, 2023, 44(3): 448-455.
MAO AI K, LIU X M, CHEN W ZH, et al. Improved substation instrument target detection method for YOLOv5 algorithm[J]. Journal of Graphics, 2023, 44(3): 448-455.

[12] 何永春, 申永伟, 吴涛, 等. 基于注意力机制的多尺度仪表检测[J]. 科学技术与工程, 2021, 21(31):

13430-13438.
HE Y CH, SHEN Y W, WU T, et al. Multi-scale instrument detection based on attention mechanism[J]. Science Technology and Engineering, 2021, 21(31): 13430-13438.

[13] 李帆, 王委, 郭寒, 等. 浅谈运行中智能电能表的常见故障及解决措施[J]. 电测与仪表, 2016,53(S1): 128-130.
LI F, WANG W, GUO H, et al. Common faults and solving measures of smart electricity meter in the field operation[J]. Electrical Measurement & Instrumentation, 2016, 53(S1): 128-130.

[14] 黄宗启. 用电信息采集系统电能计量数据异常的原因探讨[J]. 电子测试, 2019 (16): 125-126.
HUANG Z Q. Discussion on the causes of abnormal electric energy metering data in electric power information acquisition system[J]. Electronic Test, 2019(16): 125-126.

[15] 张天任, 孟强. 智能电表故障大数据分析探究[J]. 环球市场, 2020 (10): 128.
ZHANG T R, MENG Q. Exploration of big data analysis on smart meter faults[J]. Global Market, 2020 (10): 128.

[16] 刘振中. 智能电网建设中智能电能表的计量故障技术分析[J]. 电工技术, 2024 (S1): 330-332.
LIU ZH ZH. Analysis of measurement fault technology for intelligent energy meters in the construction of smart grid[J]. Electric Engineering, 2024 (S1): 330-332.

[17] 张静. 智能电能表及其计量故障处理研究[J]. 光源与照明, 2023 (1): 168-170.
ZHANG J. Research on smart electricity meters and their metering fault handling [J]. Lamps and Lighting, 2023 (1): 168-170.

[18] 郑含博, 李金恒, 刘洋, 等. 基于改进 YOLOv3 的电力设备红外目标检测模型[J]. 电工技术学报, 2021, 36(7): 1389-1398.
ZHENG H B, LI J H, LIU Y, et al. Infrared object detection model for power equipment based on improved YOLOv3 [J]. Transactions of China Electrotechnical Society, 2021, 36(7): 1389-1398.

[19] VARGHESE R, SAMBATH M. YOLOv8: A novel object detection algorithm with enhanced performance and robustness [C]. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, 2024: 1-6.

[20] QIN D F, LEICHNER C, DELAKIS M, et al. Mobile NetV4: Universal models for the mobile ecosystem [C]. Computer Vision-ECCV 2024, 2025: 78-96.

[21] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[22] ZHAO Y, LYU W Y, XU SH L, et al. DETRs beat YOLOs on real-time object detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[23] BISHOP C M. Pattern recognition and machine learning[M]. New York: Springer, 2006.

[24] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The elements of statistical learning: Data mining, inference, and prediction[M]. New York: Springer, 2009.

[25] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [C]. 2012 25th International Conference on Neural Information Processing Systems, 2012: 1097-1105.

[26] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[27] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv preprint arXiv: 1409. 1556, 2014.

[28] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]. Computer Vision-ECCV 2014, 2014: 818-833.

[29] MA X, DAI X Y, BAI Y, et al. Rewrite the stars[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5694-5703.

[30] 邓天民, 陈月田, 余洋, 等. 聚焦形状特征的路面病害检测算法[J]. 计算机工程与应用, 2024, 60(24): 291-305.
DENG T M, CHEN Y T, YU Y, et al. Pavement disease detection algorithm focusing on shape features[J]. Computer Engineering and Applications, 2024, 60(24): 291-305.

[31] ZHENG ZH H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box

regression[C]. 2020 34th AAAI Conference on Artificial Intelligence, 2020: 12993-13000.

[32] ZHANG H, ZHANG SH J. Focaler-IoU: More focused intersection over union loss[J]. ArXiv preprint arXiv: 2401.10525, 2024.

[33] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the magnitude-based pruning[J]. ArXiv preprint arXiv:2010.07611, 2020.

[34] HAN S, MAO H Z, DALLY W J. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding[J]. ArXiv preprint arXiv:1510.00149, 2015.

作者简介

陈方彬,现为西南大学工程技术学院 2022 级本科生,主要研究方向为电气设备的智能故障检测与诊断技术。

E-mail:cfb2713@ email. swu. edu. cn

Chen Fangbin is currently a 2022 undergraduate student at the College of Engineering and Technology, Southwest University. His main research interest includes intelligent fault detection and diagnosis technology for electrical equipment.

赵仲勇(通信作者),2011 年于重庆大学获得学士学位,2017 年于重庆大学获得博士学位,现为西南大学副教授,主要研究方向为为电气设备故障智能检测与诊断、脉冲功率技术及其应用。

E-mail:zhaozy1988@ swu. edu. cn

Zhao Zhongyong(Corresponding author) received his B. Sc. and Ph. D. degrees both from College of Chongqing University in 2011 and 2017, respectively. He is currently an associate professor at Southwest University. His main research interests include intelligent fault detection and diagnosis of electrical equipment, pulsed-power technology and its applications.

王建,2009 年于重庆大学获得学士学位,2012 年于重庆大学获得硕士学位,现为国网新疆电力有限公司电力科学研究院高级工程师,主要研究方向为电力设备故障诊断、输变电设备防灾减灾技术。

E-mail:yaowang360@ 163. com

Wang Jian received his B. Sc. and M. Sc. degrees both from Chongqing University in 2009 and 2012, respectively. He is currently a senior engineer at the Electric Power Research Institute of State Grid Xinjiang Electric Power Co, Ltd. His main research interests include fault diagnosis of power equipment and disaster prevention and mitigation technology for transmission and transformation equipment.

张开迪,2011 年于重庆大学获得学士学位,2014 年于重庆大学获得硕士学位,现为国网重庆市电力公司北碚供电分公司高级工程师,主要研究方向为配网智能化与数字化。

E-mail:zhangkaidi@ cq. sgcc. com. cn

Zhang Kaidi received her B. Sc. and M. Sc. degrees both from Chongqing University in 2011 and 2014, respectively. She is currently a senior engineer at the Beibei Power Supply Branch of State Grid Chongqing Electric Power Company. Her main research interests include intelligentization and digital distribution of network.