

DOI: 10.13382/j.jemi.B2306603

SGCNet: 一种轻量化的新能源汽车电池集流盘缺陷检测模型*

袁磊¹ 唐海^{1,2} 陈彦蓉¹ 高刃¹ 吴文欢¹

(1. 湖北汽车工业学院电气与信息工程学院 十堰 442002; 2. 湖北汽车工业学院汽车动力传动与电子控制湖北省重点实验室 十堰 442002)

摘要:集流盘作为新能源汽车电池的重要组成部分,其质量好坏关系到电池的性能,对车内人员的生命安全有着重要影响。实际工业应用中,在有限的计算资源下对电池集流盘缺陷进行实时检测是一项具有挑战性的任务。为了减小模型大小和计算量,降低应用成本,本文提出一种轻量化的新能源汽车电池集流盘缺陷检测模型(SGCNet)。首先,采用 ShuffleNet V2 作为主干特征提取网络,采用分组卷积和通道重排技术,在提取有效特征的同时降低计算复杂度,降低参数量。其次,设计了轻量化的特征融合网络 GC-FPN,采用轻量级 GhostNet 和 CARAFE 上采样算子,在减少参数冗余和保证检测精度的情况下充分保留特征图的语义信息,从而降低了计算成本。实验结果表明,SGCNet 模型检测准确率达到了 90.6%,模型大小为 3.2 M,GFLOPs 仅为 3.6,帧率达到了 178.6 fps。与目前先进的轻量化网络模型相比,具有较高检测精度和较低的计算量。最后,在嵌入式平台 NVIDIA Jetson Nano 上部署 SGCNet 模型,进行实时检测,每张图片的检测时间为 0.07 s,满足实际工业中电池集流盘缺陷检测任务对精度和实时性的要求。

关键词: 电池集流盘;轻量化;缺陷检测;Jetson Nano

中图分类号: TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.2

SGCNet: A lightweight defect detection model for new energy vehicle battery collector tray

Yuan Lei¹ Tang Hai^{1,2} Chen Yanrong¹ Gao Ren¹ Wu Wenhuan¹

(1. School of Electrical and Information Engineering, Hubei University of Automotive Technology, Shiyan 442002, China;

2. Key Laboratory of Automotive Power Train and Electronics, Hubei University of Automotive Technology, Shiyan 442002, China)

Abstract: As an important component of the new energy vehicle battery, the quality of the collector tray is related to the performance of the battery and has an important impact on the life safety of the vehicle occupants. In practical industrial applications, real-time detection of battery collector tray defects with limited computational resources is a challenging task. In order to reduce the model size and computational effort, and to reduce the application cost, this paper proposes a lightweight new energy vehicle battery collector tray defect detection model (SGCNet). First, ShuffleNet V2 is used as the backbone feature extraction network, and group convolution and channel shuffle techniques are adopted to reduce the computational complexity and the number of parameters while extracting effective features. Secondly, a lightweight feature fusion network GC-FPN is designed with lightweight GhostNet and CARAFE upsampling operators to fully retain the semantic information of the feature map while reducing parameter redundancy and ensuring detection accuracy, thus reducing the computational cost. The experimental results show that the SGCNet model achieves 90.6% detection accuracy, the model size is 3.2 M, the GFLOPs are only 3.6, and the FPS reaches 178.6. Compared with the current advanced lightweight network models, it has higher detection accuracy and lower computational effort. Finally, the SGCNet model is deployed on the embedded platform NVIDIA Jetson Nano for real-time detection, with a detection time of 0.07 s per image, meeting the requirements for accuracy and real-time performance for battery collector defect detection tasks in real industry.

Keywords: battery collector tray; lightweight; defect detection; Jetson Nano

收稿日期: 2023-06-06 Received Date: 2023-06-06

* 基金项目: 国家自然科学基金(52072116, 52075107)、湖北省自然科学基金项目(2022CFB53B)、湖北省教育厅科学技术研究项目(Q20201801)、湖北汽车工业学院博士科研启动基金(BK202004)、汽车动力传动与电子控制湖北省重点实验室(湖北汽车工业学院)(ZDK1201603)项目资助

0 引言

“绿色低碳”是当今时代发展的主要目标之一。在未来实现碳中和的发展过程中,新能源汽车的使用比例必然会逐渐升高,在长时间之后就会实现新能源汽车完全替代传统燃油汽车的局面。电池是新能源汽车的关键组件,我国的新能源汽车在发展方面的困境之一就是缺乏健全的新能源汽车电池检测及维护体系^[1]。电池的结构一般包括集流盘、外壳、电芯。相关技术中,集流盘与卷芯端面的焊接质量对全极耳电池的后续制备工作具有重要影响,集流盘需要分别与电芯和外壳进行焊接,在将电芯放入外壳时,可能因电芯与外壳间存在间隙使得电芯轴线偏离外壳轴线,容易导致偏心焊接(焊偏、重度不良)的情况发生,这样,焊接后的电芯在膨胀时,容易造成电芯与外壳的焊接处受力不均的情况发生;焊接时间过短没焊牢导致传输过程中集流盘脱落(无盖);焊接时间过长会造成集流盘产生不同程度损坏(坏点、焊穿)。这些缺陷会影响电池的可靠性和使用寿命^[2],严重的会造成电池短路进而引发爆炸,可能导致人员伤亡和财产损失。因此,必须仔细检测电池集流盘的表面缺陷以保证电池的安全性。然而,传统的人工检测劳动强度大、效率低并且误检率高。近些年来,高效、智能的目标检测技术逐渐取代了人工检测。

研究人员在 ILSVRC2012 图像分类项目中发现,卷积神经网络(convolutional neural network, CNN)能够学习具有鲁棒性和一定表达性的特征表示。因此, CNN 被应用于工业场景中的目标检测,但仍然存在一些局限性。大型 CNN 通常需要高性能的并行计算资源以提高检测性能,比如昂贵的 GPU;且网络模型的训练过程也需要消耗大量计算资源和时间。因此,需要开发轻量级的检测模型和框架,在有限的计算资源下进行部署应用。常见的设计轻量化网络模型的方法有:对原有的网络模型结构进行优化,以达到轻量化的目的;通过调整网络模型结构的宽度或深度缩放因子,重用相同的网络模型结构,实现不同的模型大小。对于轻量化的网络设计,目前经典且高效的轻量级卷积神经网络有 SqueezeNet^[3]、MobileNet^[4]、ShuffleNet^[5]、GhostNet^[6]、EfficientNet^[7]、Xception^[8]以及最新提出的 FasterNet^[9]等结构。

目前,基于深度学习的目标检测算法被广泛应用于各个领域,主流的目标检测算法分为基于回归的单阶段目标检测算法(如 YOLO 系列^[10-17])和基于候选区域的两阶段目标检测算法(如 R-CNN 系列^[18-21])两种,研究人员在这两种类型的目标检测算法上都取得了显著的成果。钱万明等^[22]以 YOLOv4 算法为基本框架,提出一种专门针对 PCB 裸板的缺陷检测算法,改进后的算法相较

Faster R-CNN,检测精度和速度都有大幅提升;刘艳菊等^[23]提出一种改进 YOLOv4 自动检测方法,用于热轧钢条的表面质量缺陷检测,改进后的方法平均精度值提高约 7.94%;李衍照等^[24]基于深度学习提出一种改进的 YOLOv5 焊缝缺陷检测方法,相较 Faster R-CNN,检测精度明显提高;钱承山等^[25]提出改进 YOLOv5 的遥感图像检测网络,改进后的网络相较原算法 mAP50 提高了 3.1%,但检测速度较慢;马燕婷等^[26]提出了一种基于 YOLOv5 网络改进的算法模型用于带钢缺陷检测,相较原算法 mAP50 提高 5.3%,但模型体积增加了 15.6 M。

以上方法的检测方法都取得了不错的效果,但仍存在如下问题:

- 1) 有些网络模型具有较高的检测精度,但参数数量和计算量给计算设备带来很大的负担,难以满足部署应用要求;
- 2) 有些轻量级的网络模型具有良好的检测速度,但检测精度仍有提升的空间;
- 3) 有些网络模型的检测目标为固定的尺寸和特征,但电池集流盘缺陷尺寸跨度大、特征种类多,应用于电池集流盘缺陷检测任务中普适性较差。

两阶段的目标检测算法检测精度高,但检测时间长、计算量较大,不满足工业实际需求。因此,本文使用单阶段目标检测算法,相比两阶段目标检测算法,它简单高效,适合工程应用,可以更好的部署在移动设备上,但在准确率上仍有差距。电池集流盘目标缺陷有着分布杂乱、尺寸差距大等问题,因此,检测电池集流盘表面缺陷具有挑战性。在之前的研究中^[27],同样对电池集流盘的缺陷检测进行了深入探讨,并取得了一定的成果,但模型的复杂度较高。为进一步实现模型的轻量化,本文提出了一个轻量化的电池集流盘缺陷检测网络模型(SGCNet),并在移动设备部署应用进行实时检测。本文的工作内容如下:

- 1) 新能源汽车电池集流盘还没有标准的缺陷数据集,本文实验所用的数据集为自行构建;
- 2) 受到轻量级网络 ShuffleNet V2^[28]和 YOLOv5 的启发,设计了一种更高效的轻量级骨干网络,在保证网络特征提取能力的同时降低计算复杂度;
- 3) 设计了 GC-FPN 作为特征融合网络,进一步减少模型参数量,加快特征融合;
- 4) 为检验 SGCNet 的检测性能,将最终检测模型成功部署到嵌入式平台 NVIDIA Jetson Nano 上,利用自制数据集,实现电池集流盘目标缺陷实时检测要求。

1 网络模型

为降低网络复杂性,本文设计了一种简单高效的轻

量化模型 SGCNet,用于电池集流盘的目标缺陷实时检测。图 1 为 SGCNet 模型的结构图。

1.1 轻量化主干特征提取网络

原 YOLOv5 在目标检测任务中都表现出了优异的性能。其主干特征提取网络分为 CBS 模块、C3 模块和 SPPF 模块。由于主干网络由大量的 CBS 模块和 CBS 模块堆叠的 C3 模块组成,计算开销成本很高,使得其在资源有限的移动设备上部署困难,这限制了其在实际工业中的应用。然而,本文提出的电池集流盘缺陷检测模型只针对坏点、焊偏等 5 种缺陷类型的目标对象且样本背景格式统一,其网络拟合能力不需要那么强。所以,本文

重构了 YOLOv5 的主干特征提取网络,在保持较高精度的同时减少计算量,使得模型更加适用于实际工业应用。

传统的特征提取方法会捕获大量的特征信息,产生冗余数据。通过叠加多个卷积层将生成许多冗余的特征映射,需要大量的参数和计算资源来处理。一些研究人员提出了对模型进行压缩的方法,如剪枝、量化、知识蒸馏等,可以有效地减少参数量,但这些方法存在损失检测精度和训练困难的问题。受到轻量化网络 ShuffleNet V2 的启发,本文将 ShuffleNet V2 作为主干网络,在提高检测性能的同时减少计算开销。如图 1 所示,SGCNet 模型的主干网络由 3 个堆叠的 Shuffle V2 Block 组成。

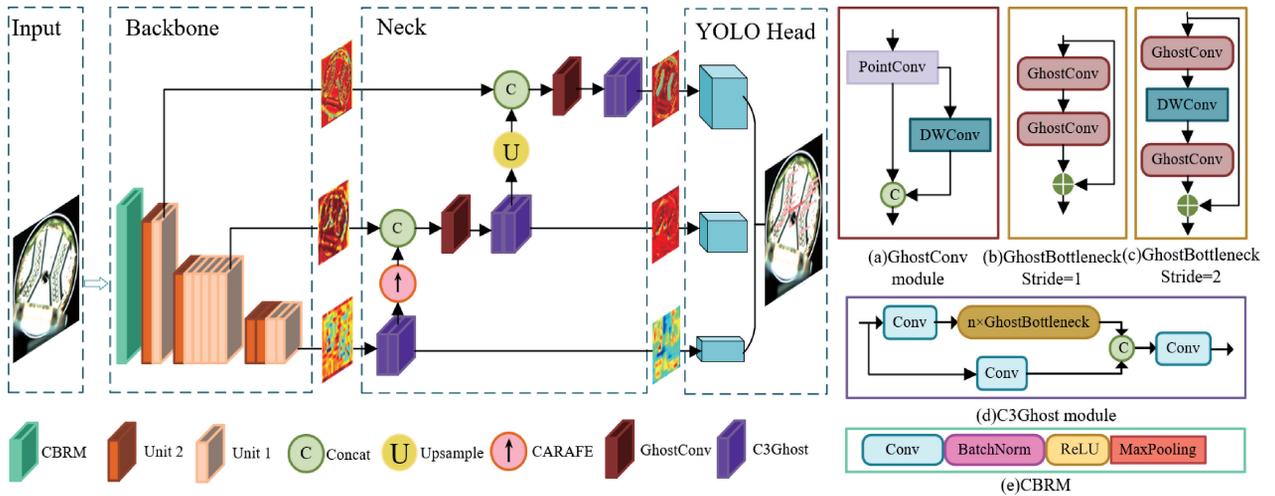


图 1 SGCNet 模型结构

Fig. 1 Model structure diagram of SGCNet

ShuffleNet V1 将组卷积和随机信道混洗两种操作相结合,形成了 ShuffleNet 的块网络模块。ShuffleNet V2 在 ShuffleNet V1 的基础上进一步完善,提出了 4 个操作原则,包括通道分离、输入和输出特征通道数量相等、在基本单元中使用常规卷积而不是分组卷积、使用 concat 代替元素级操作 add,以加快内存并行操作。ShuffleNet V2 网络结构如图 2 所示,主要由 Unit 1 和 Unit 2 两个单元模块组成。Unit 1 只提取特征,而不改变输入特征图的大小和通道数。Unit 2 将输入特征图的高度和宽度减半,并将通道数扩大到原来的 4 倍。

Unit 1 中,首先进行通道划分,左边的分支不做处理,右边的分支依次通过 1×1 的逐点卷积层、 3×3 的深度卷积层(Stride=1)、 1×1 的逐点卷积层。接着,左右分支进行特征融合。最后通过 Channel Shuffle 实现不同组之间的特征融合。Unit 2 中,首先将输入特征映射传输到两个支路,左边的分支经过 3×3 的深度卷积层、 1×1 的逐点卷积层,右边的分支与 Unit 1 中类似。其中,两个分支都采用 Stride = 2 的深度卷积(DWConv)提取语义信息,减小特征图的高度和宽度。接着,左右分支进行特征融

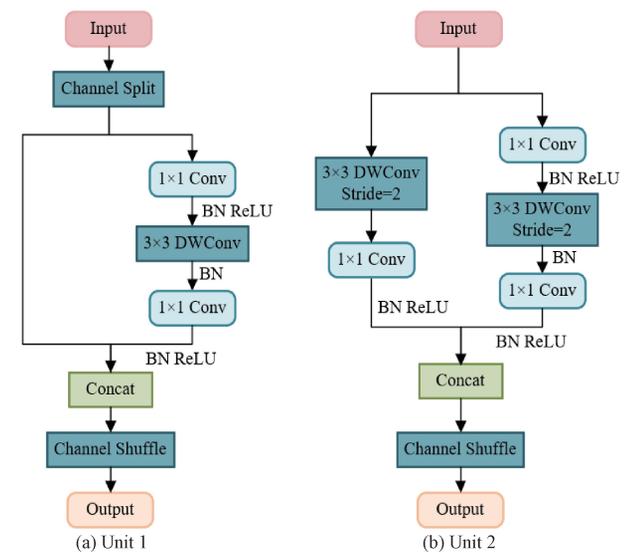


图 2 ShuffleNet V2 网络结构

Fig. 2 The Network Structure of ShuffleNet V2

合,得到通道数量是输入的 4 倍,大小减半的特征映射。

最后通过 Channel Shuffle 实现不同组之间的特征融合。

Channel Shuffle 操作原理如图 3 所示。通过对特征图的通道进行重塑和转置,将通道进行分解和重组,在不增加计算成本的情况下快速完成通道间的信息融合,释放了卷积操作带来的通道信息约束。因此,将 ShuffleNet V2 作为主干网络实现了特征的快速提取和传输,在保证网络特征提取能力的同时降低计算复杂度。

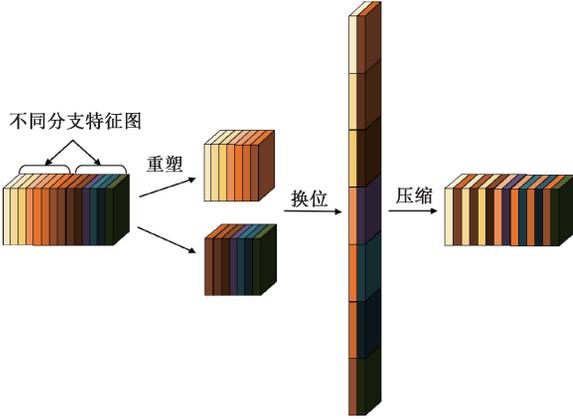


图 3 Channel Shuffle 操作原理

Fig. 3 Operating principle of Channel Shuffle

1.2 轻量化特征融合网络

CNN 的特征融合网络通过将不同层次的特征进行融合,可以更好地捕捉到输入数据中的多尺度、多层次信息,从而提高模型的性能。本文设计的轻量化特征融合网络 GC_FPN 如图 1 中所示。

(1) 首先,采用 FPN^[29] (feature pyramid network) 结构。FPN 通过自顶向下和自底向上的路径融合多尺度的特征,使得模型能够同时处理不同尺度的目标,提高了检测精度。其中,自底向上的路径进行特征提取,避免了计算冗余,提高了计算效率。

为了进一步减少模型参数量,更好地实现实时检测,引入 GhostConv 模块和 C3Ghost 模块,将 CBS 模块替换为 GhostConv 模块,将 C3 模块替换为 C3Ghost 模块。GhostConv 模块的 3 个阶段如下:

(1) 生成 N 个特征映射的标准卷积过程表示为:

$$r_s = \frac{N \times H' \times W' \times C \times K \times K}{N/S \times H' \times W' \times C \times K \times K + N/S \times (S-1) \times H' \times W' \times d \times d} = \frac{C \times K \times K}{1/S \times C \times K \times K + (S-1)/S \times d \times d} \approx \frac{S \times C}{S + C - 1} \approx S \quad (4)$$

参数的压缩比可计算为:

$$r_c = \frac{N \times C \times K \times K}{N/S \times C \times K \times K + N/S \times (S-1) \times d \times d} \approx \frac{S \times C}{S + C - 1} \approx S \quad (5)$$

$$Y = F * f + b \quad (1)$$

其中, F 为输入, $F \in R^{H \times W \times C}$ (H 为高, W 为宽, C 为通道数), $f \in R^{N \times K \times K \times C}$ ($K \times K$ 为卷积核大小), b 为偏置项, $Y \in R^{N \times H' \times W'}$ (H' 为输出高度, W' 为输出宽度);

首先使用标准卷积得到 M 个特征图 Y' (忽略偏置项), 如下:

$$Y' = F * f' \quad (2)$$

其中, $f' \in R^{M \times K \times K \times C}$ ($M < N$), $Y' \in R^{M \times H' \times W'}$;

(2) 然后对每个通道的固有特征图进行 Φ 操作处理,新的特征图 y_{ij} 被称为之前特征图的“Ghost”, 如下:

$$y_{ij} = \Phi_{i,j}(y'_i), i \in [1, M], j \in [1, S] \quad (3)$$

其中, y'_i 为 Y' 中第 i 个特征映射, Φ 为线性操作;

(3) 最后,将第 1 步得到的特征图和第 2 步得到的 Ghost 特征图连接起来,得到最终的结果 Output。

GhostConv 模块如图 4 所示。

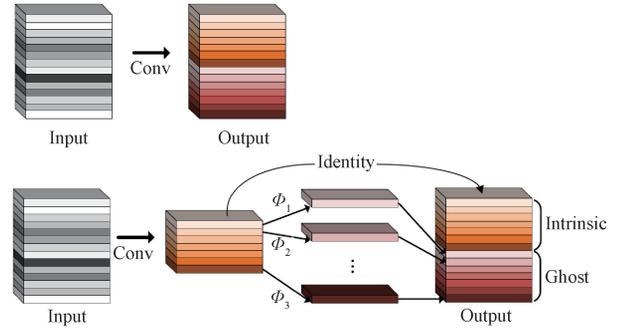


图 4 GhostConv 模块

Fig. 4 GhostConv module

接下来分析 GhostConv 模块理论上的速度。

单个卷积运算所需的 FLOPs 可以计算为 $N \times H' \times W' \times C \times K \times K$ 。在 GhostConv 模块中有 1 个恒等映射和 $N/S \times (S-1)$ 个线性运算 ($N = M \times S$), 每个卷积核的平均大小为 $d \times d$ (类似于 $K \times K$ 且 $S \ll C, M < S$)。GhostConv 模块的 FLOPs 个数可以表示为: $N/S \times H' \times W' \times C \times K \times K + N/S \times (S-1) \times H' \times W' \times d \times d$ 。使用 GhostConv 模块增强常规卷积的理论加速比可计算为:

根据表达式可以看出,标准卷积的计算量是 GhostConv 的 S 倍,显然 GhostConv 的计算成本更低。GhostBottleneck 是由 GhostConv 组成的轻量级 Bottleneck 结构,如图 1 所示,使用 GhostBottleneck 取代 C3 模块中的 Bottleneck 结构,形成新的模块 C3Ghost。这种改进不

降低检测精度的前提下减少了标准卷积带来的参数量,从而减小了模型的大小。

2) 最后,通常采用的最近邻插值方法对特征图的每个位置应用相同的上采样核。它直接将低分辨率特征图中的每个像素值复制 to 高分辨率特征图中的对应位置上。这个过程不涉及任何额外的计算,所以计算速度快,但是这种方法的主要问题在于输出结果缺乏平滑性,容

易产生锯齿状的边界,导致图像精度下降。为解决这一问题,本文采用轻量级 CARAFE^[30] 上采样算子,其原理如图 5 所示。CARAFE 在更大的接收域内感知和聚集上下文信息,它不是对所有特征应用一个固定的卷积核,而是动态地生成自适应上采样核,然后根据预测的上采样核重新组织特征。

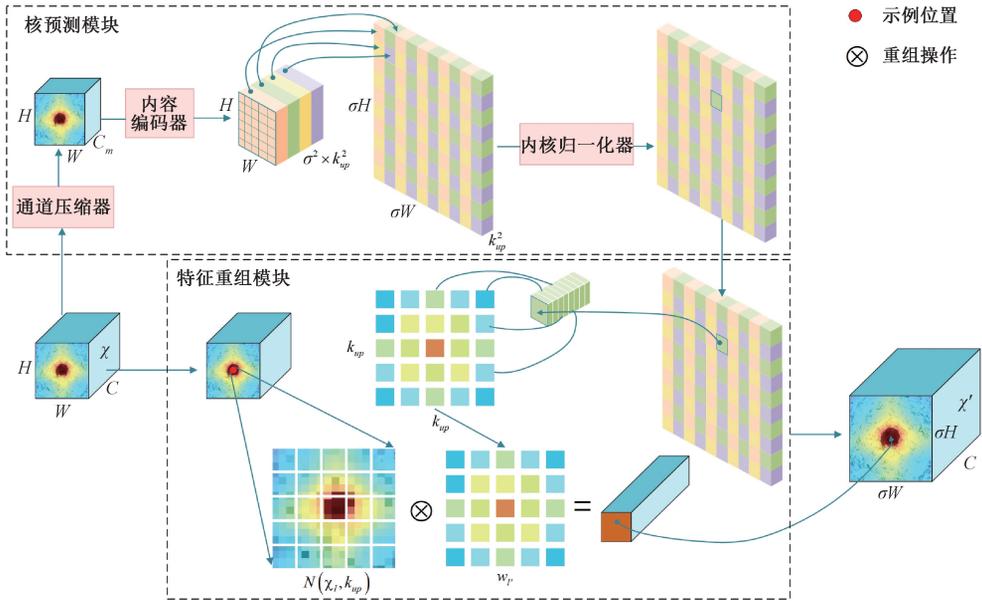


图 5 CARAFE 原理

Fig. 5 The principle of CARAFE

CARAFE 主要由两个模块组成:上采样核预测模块和特征重组模块。核预测模块的作用是生成一个重构的卷积核。上采样系数为 σ , 上采样核大小为 $k_{ap} \times k_{ap}$, 输入图像大小为 $W \times H \times C$ (W 为宽, H 为高, C 为通道数)。首先,为了降低复杂度,通过通道压缩器模块内的 1×1 卷积将输入通道数 C 压缩到 C_m 。然后,内容编码器模块利用大小为 $k_{encoder} \times k_{encoder}$ 的卷积层,对前一个模块输出的特征映射进行卷积运算,得到大小为 $W \times H \times k_{ap}^2 \times \sigma$ 的重组核。特征重组模块首先使用 softmax 对上一步得到的上采样核进行归一化,使卷积核的权重求和为 1。然后将重组内核进行加权和运算,对局部区域内的特征进行重组。接着,输入特征图的每个像素选择一个 $k_{ap} \times k_{ap}$ 的区域进行卷积,最后与重组核进行点积,得到大小为 $\sigma W \times \sigma H \times C$ 的新特征图。如式(6)所示:

$$w_r = \psi(N(x_l, k_{encoder}))$$

$$x'_r = \phi(N(x_l, k_{ap}), w_r) \tag{6}$$

其中, ψ 指上采样核预测模块; $k_{encoder}$ 为编码卷积的卷积核; ϕ 为特征重组模块,将 x_l 相邻位置的特征和内核 w_r 进行重组; $N(x_l, k)$ 为以位置 l 为中心的子区域 $k \times k$, 即 x_l 的邻接位置。

2 实验结果与分析

2.1 数据集

由于新能源汽车电池集流盘还没有标准的缺陷数据集,本文实验所用的数据集为自行构建,使用从生产线上收集的真实磷酸铁锂汽车电池集流盘缺陷数据,由高分辨率巴斯勒工业相机在光线良好的室内环境下进行拍摄采集。原始图像分辨率为 $2\,448 \times 2\,048$, 在预处理阶段通过 Python 以电极孔为中心对原始图像进行裁剪,去除无关背景,保留有用信息,截取之后的图像分辨率为 $1\,250 \times 1\,200$ 。使用 Labelme 数据标注工具将图像中所有的缺陷的位置和类别进行标注,标注后自动生成 JSON 格式文件,文件名与图片名始终保持一致。并且对标注结果进行反复验证和修正,避免主观因素对实验结果的影响。

图 6 为电池集流盘代表性的图像,将常见的 5 种缺陷类型定义为:焊穿 (weld through)、焊偏 (welding offset)、无盖 (no cover)、坏点 (bad point)、重度不良 (severely bad)。

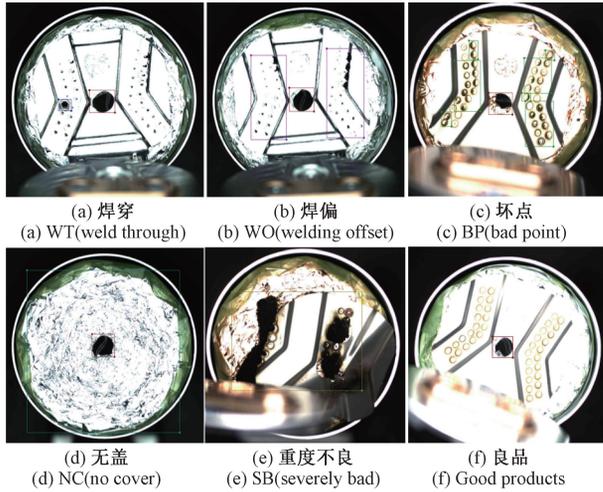


图 6 电池集流盘标签示例图像

Fig. 6 Example image of a label for a battery collector tray

2.2 实验环境及参数配置

实验环境及参数配置如表 1 所示。

表 1 实验环境及参数配置

Table 1 Experimental environment and parameter configuration

项目	内容
CPU	Intel(R) Core(TM) i5-12600KF
RAM	16 G
GPU	NVIDIA RTX3080×2
CUDA	Cuda11.3 and Cudnn10.1
深度学习框架	Pytorch 1.12.1
操作系统	Ubuntu 18.04.6
程序语言	Python 3.9
Batch-size	16
Epoch	300
Image Size	640

表 2 CARAFE 融合实验

Table 2 Fusion experiment of CARAFE

算法模型	1	2	体积/MB	计算量/GFLOPs	P	R	mAP50/%	mAP50-95/%
Shuffle_GFPN			2.7	3.5	84.9	89.3	89.3	64.4
CARAFE(1)	√		3.2	3.6	86.9	86.0	90.2	65.0
CARAFE(2)		√	3.0	4.7	86.5	86.2	89.4	64.1
CARAFE(1,2)	√	√	3.4	5.3	88.8	85.7	88.4	64.7

从表 2 可以看出, 相较 Shuffle_GFPN, 3 种嵌入 CARAFE 位置模型的 mAP50 分别提高 0.9%、0.1%、-0.9%, GFLOPs 分别增加 0.1、1.2、1.8, 模型体积分别增加 0.5、0.3、0.7 M。在两个上采样位置同时采用 CARAFE 效果不佳, 检测精度下降, 会使得边缘部分变得模糊, 并且容易将噪声进行扩散。结果表明, 在第 1 个上采样的位置使用 CARAFE 效果最佳。相比于最近邻插

2.3 评价指标

Precision、Recall 和 mAP50 被用来评价和比较 SGCNet 的检测性能, 分别列于式(7)~(10)中。

$$AP = \int_0^1 p dR \quad (7)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

其中, TP 为真正例, FP 为假正例, FN 为假反例。Precision 和 Recall 分别表示正确预测的样本数量占正类预测总数的比例, 而正类预测指的是正类样本数占正类样本总数的比例。P-R 曲线是由 Precision 和 Recall 之间的关系推导出来的。所有类别的平均精度 (AP) 是曲线和轴包围的区域的面积。mAP50 通过对数据集中所有类别的 AP 取平均值得到。

GFLOPs 用于衡量模型的复杂度。公式如下:

$$GFLOPs = \frac{\text{浮点运算次数}}{\text{以秒为单位的运行时间}} \quad (11)$$

2.4 实验结果及分析

1) CARAFE 融合实验

为进一步验证 CARAFE 上采样算子的有效性以及控制模型的参数量而寻找要嵌入的最佳位置, 基于 GhostNet 与 FPN 结构 (Shuffle_GFPN), 本文将特征融合网络的两个上采样依次采用 CARAFE, 其他部分不变。实验结果如表 2 所示。“1”和“2”分别表示两个上采样的位置, “√”表示此位置采用 CARAFE 上采样。

值上采样, 虽然参数量和计算量有微量增加, 但是它可以提供更高的上采样质量和更好的性能, 在不添加过多参数的情况下充分保留特征图的语义信息, 从而降低了计算成本, 并且检测精度也有所提高。

2) 消融实验

SGCNet 主干网络采用轻量化网络 ShuffleNet V2, 特征融合网络采用轻量化的 GC_FPN。为充分验证 SGCNet

的有效性,进行消融实验以验证不同部分的轻量级模块对网络性能的影响。每组实验使用了相同的训练参数和环境条件,实验结果如表 3 所示。以 YOLOv5m 为基准模型。将 ShuffleNet V2 作为主干网络的实验称为 YOLOv5m-A,在 YOLOv5m-A 的基础上采用 G_FPN (GhostNet 和 FPN 结构)作为特征融合网络的实验称为 YOLOv5m-B。

从表 3 可以看出:YOLOv5m 模型的检测效果较差,模型体积和参数量较大;当分别进行 ShuffleNet V2、G_FPN 和 CARAFE 改进时,相较基线网络,mAP50 值分别提升 1.6%、1.8%和 1.1%,模型大小分别减少 37.8、15.7 和 0 M,GFLOPs 分别降低 43.6、14.3、-0.1。说明采用 ShuffleNet V2 有效减少了模型的计算量和参数量;ShuffleNet V2 分别与 G_FPN 和 CARAFE 结合时,模型大小分别再减小 1.8、-13.8 M,GFLOPs 分别再下降 0.8、-13.3,mAP50 分别下降 0.6%、0.7%。虽检测精度略微下降,但 G_FPN 大幅降低了特征融合部分的参数量和计算量,仍在可接受范围内;仅 G_FPN 和 CARAFE 结合时,相较基线网络,模型大小、GFLOPs 和 mAP50 分别降低 13.2 M、11.9、1.2%。说明 GC_FPN 相比原始的 FPN+PANet 的特征融合网络能够大幅缩减模型参数冗余;当 3 种改进结合在一起时,相较基线网络,mAP50 值提高了 3.6%,体积减少 39.1 M,GFLOPs 降低 44.3,综合效果最佳。

结果证明了所提出的轻量化改进方法的有效性。

表 3 消融实验

Table 3 Ablation experiment

方法	ShuffleNet V2	G_FPN	CARAFE	体积/MB	计算量/GFLOPs	P	R	mAP50/%	mAP50-95/%
YOLOv5m				42.3	47.9	87.6	85.4	87.0	64.7
YOLOv5m-A	√			4.5	4.3	86.6	82.0	88.6	60.3
YOLOv5m-G_FPN		√		26.6	33.6	90.9	83.9	88.8	59.2
YOLOv5m-CARAFE			√	42.3	48.0	86.7	86.8	88.1	63.2
YOLOv5m-B	√	√		2.7	3.5	87.1	85.2	88.0	57.1
YOLOv5m-A_CARAFE			√	18.3	17.6	87.9	86.8	87.9	62.6
YOLOv5m-GC_FPN		√	√	29.1	36.0	83.0	82.1	85.8	57.4
SGCNet	√	√	√	3.2	3.6	88.2	89.1	90.6	64.4

表 4 主干、特征融合网络的对比实验

Table 4 Comparative experiments on backbone and feature fusion networks

Backbone	Neck	体积/MB	计算量/GFLOPs	P	R	mAP50/%	mAP50-95/%
Efficientnet_b0		9.3	8.7	86.6	82.6	85.7	59.7
Efficientnet V2_s		42.2	51.4	87.5	86.0	86.4	59.8
MobileNet V3_small_050		2.5	2.7	83.7	84.2	87.6	59.2
Xception	GC_FPN	45.3	74.4	90.5	82.5	86.2	57.7
Fasternet_t0		6.4	7.9	87.3	87.8	88.7	58.5
ShuffleNet V2		3.2	3.6	88.2	89.1	90.6	64.4
	FPN+PAN	4.5	4.3	86.6	82.0	88.6	60.3
ShuffleNet V2	BIFFN	18.4	17.8	85.4	86.3	88.7	60.3
	slimNeck	17.9	11.5	88.1	85.5	88.5	59.1
	GC_FPN	3.2	3.6	88.2	89.1	90.6	64.4

3) 对比实验

为验证 SGCNet 主干网络 (ShuffleNet V2) 和特征融合网络 (GC_FPN) 的有效性,选择不同的主干网络和特征融合网络在自制数据集上对比本文方法对电池集流盘的缺陷检测效果。实验结果如表 4 所示。由表 4 可知,特征融合网络为 GC_FPN,采用 ShuffleNet V2 为主干网络时,模型大小和 GFLOPs 略高于 Monilenet V3_small_050,但其他各项指标均为最佳;主干网络为 ShuffleNet V2,采用 GC_FPN 作为特征融合网络时,各项指标均为最佳。

为进一步验证所提出 SGCNet 的性能,选择多种轻量化模型在自制数据集上对比本文方法 SGCNet 对电池集流盘的缺陷检测效果。实验参数相同,实验结果如表 5 所示。由表 5 可知:SGCNet 的 mAP50 达到 90.6%,模型体积为 3.2 M,GFLOPs 为 3.6,FPS 为 178.6,均为最高值。相比两阶段目标检测算法 Faster R-CNN,mAP50 高 5.9%,体积小 79.8 M,GFLOPs 低 190.7,FPS 快 155.2;相比单阶段目标检测算法 SSD,mAP50 高 18.2%,体积小 49.4 M,GFLOPs 低 179.8,FPS 快 35.4;相比 Liu 等^[31]提出的 LF-YOLO,mAP50 高 1.2%,体积小 11.7 M,GFLOPs 低 12.7,FPS 快 56.6。由表可以看出,虽然一些轻量化模型在模型体积和推理速度方面具有优势,但在检测集流盘这种目标特征复杂的缺陷时没法达到和 SGCNet 相同的准确性水平,缺乏通用性。Efficientnet V2 虽然取得了 90.2%的 mAP50,但模型体积和 GFLOPs 较高。综上,

本文提出的 SGCNet 综合性能最佳。

表 6 为 YOLOv5m、YOLOv5m-A、YOLOv5m-B 和 SGCNet 对 5 个缺陷类型的 3 种对比指标检测结果。由

表 6 可知,SGCNet 较基准模型的 3 种评价指标均有不同程度的提高,有效解决了漏检问题。

表 5 轻量化模型的对比实验

Table 5 Comparative experiments on lightweight models

方法	Backbone	体积/MB	计算量/GFLOPs	P	R	mAP50/%	mAP50-95/%	FPS	mAP50 排序
SSD	VGGNet-16	52.6	183.4	89.7	64.4	72.4	42.8	143.2	13
Faster R-CNN	ResNet50	83.0	194.3	65.8	84.9	84.7	53.2	23.4	12
MobileNet V3	Monilenet V3_small_050	14.1	9.1	83.0	77.6	84.8	50.2	60.6	11
YOLOv5n	Modified CSP V5	3.9	4.2	83.2	84.2	85.3	62.8	166.7	10
Fasternet	Fasternet_t10	16.2	15.1	85.8	83.3	87.0	55.2	175.4	9
YOLOv5m	Modified CSP V5	42.3	47.9	87.6	85.4	87.0	64.7	70.4	8
LF-YOLO×0.5	—	4.0	4.6	76.9	86.1	87.3	59.3	135.1	7
Efficientnet	Efficientnet_b0	10.3	6.6	83.7	84.5	88.1	57.0	97.1	6
Xception	Xception	58.9	76.5	85.5	86.7	88.4	58.0	46.1	5
LF-YOLO	—	14.9	16.3	76.0	88.5	89.4	61.4	122.0	4
YOLOv4s	CSPDarknet-53-mish	18.7	20.6	85.5	88.6	89.5	64.6	166.7	3
Efficientnet V2	Efficientnet V2_s	58.5	65.7	88.2	87.6	90.2	64.4	77.0	2
本文	ShuffleNet V2	3.2	3.6	88.2	89.1	90.6	64.4	178.6	1

表 6 YOLOv5m、YOLOv5m-A、YOLOv5s-B 和 SGCNet 上 5 种缺陷的检测结果

Table 6 Detection results of five defects on YOLOv5m, YOLOv5m-A, YOLOv5s-B and SGCNet

类别	P				R				mAP50/%			
	5m	5m-A	5m-B	本文	5m	5m-A	5m-B	本文	5m	5m-A	5m-B	本文
Positioning hole	98	97.7	97.7	98.2	99.9	99.6	98.9	99.9	98	98	98.6	98.6
Bad point	67.2	79.2	75.5	70.4	63.7	46.9	61.4	65.4	65	70.9	72.6	70.4
Welding offset	95.9	94.8	91.7	94.1	95.1	88.9	94.9	96.5	96.4	96.1	96	97.4
No cover	92.3	78.8	90.3	97.4	100	100	100	100	99.5	99.5	99.5	99.5
Severely bad	100	98.6	100	100	80.5	88.9	82.9	94.3	88.7	95.1	88.7	99.5
Weld through	72.5	70.6	67.1	69.1	73	67.6	73	78.4	74.7	71.8	72.5	78.2
all	87.9	86.6	87.1	88.2	85.4	82	85.2	89.1	87	88.6	88	90.6

图 7 为多个模型和 SGCNet 的实验效果对比图。由图 7 可直观地看出,在第 1 行无盖检测中,YOLOv3-tiny 和 Xception 存在误检;第 2 行重度不良检测中,YOLOv3-tiny、Fasternet、YOLOv7-tiny、Monilenet V3 和 LF-YOLO×0.5 存在不同程度的误检;在第 3 行焊偏、焊穿的检测中,除了 SGCNet,其他模型对小的焊穿都存在漏检;在第 4 行焊穿的检测中,YOLOv3-tiny、YOLOv5m、YOLOv7-tiny、Monilenet V3 和 LF-YOLO×0.5 存在漏检;在第 5 行坏点的检测中,其他模型均存在不同程度的漏检和误检,SGCNet 识别的效果较好。综上说明 SGCNet 有效解决了集流盘缺陷检测任务中误检漏检的问题。

图 8 为模型综合性能对比散点图, X 轴为模型的推理时间, Y 轴为 mAP50, 点的大小表示模型 GFLOPs 的排名, 点越小位置越靠近左上角, 模型综合性能越好。由图 8 可知, SGCNet 在众多轻量化模型中, 模型计算量最低, 检测精度最高, 模型推理时间较短, 综合性能最佳。极大地提高了模型在移动设备等资源受限环境下的运行效率。

3 模型的部署

基于以上模型设计, 本文在嵌入式设备 NVIDIA Jetson Nano 上部署模型, 以实现移动端的电池集流盘目标缺陷实时检测。使用本文提出的 SGCNet 模型作为实际应用检测模型, 搭建 TensorRT 平台, 将模型转化为 ONNX 形式, 把模型读入 TensorRT 的加速引擎中, 部署到嵌入式设备 Jetson Nano 中。所述装置包括 NVIDIA Jetson nano、电源和单目摄像头。其中, NVIDIA Jetson Nano 是一种小型的嵌入式计算机, 采用了 64 位四核 ARM Cortex-A57 CPU 和 128 核 NVIDIA Maxwell GPU, 内置 4GB LPDDR4 RAM 和 16 GB eMMC 存储器; 为了避免受到线材损耗影响, 采用了 5 V 4 A 的 DC 电源供电; 单目摄像头为 200 万像素 USB 摄像头, 配备在开发板上, 对电池集流盘进行实时缺陷检测。图 9 为实时检测效果图。由图 9(a) 可以清楚的看出, 未经 TensorRT 加速的检测时间最高达到 0.111 s。从其他图可以看到, 使用

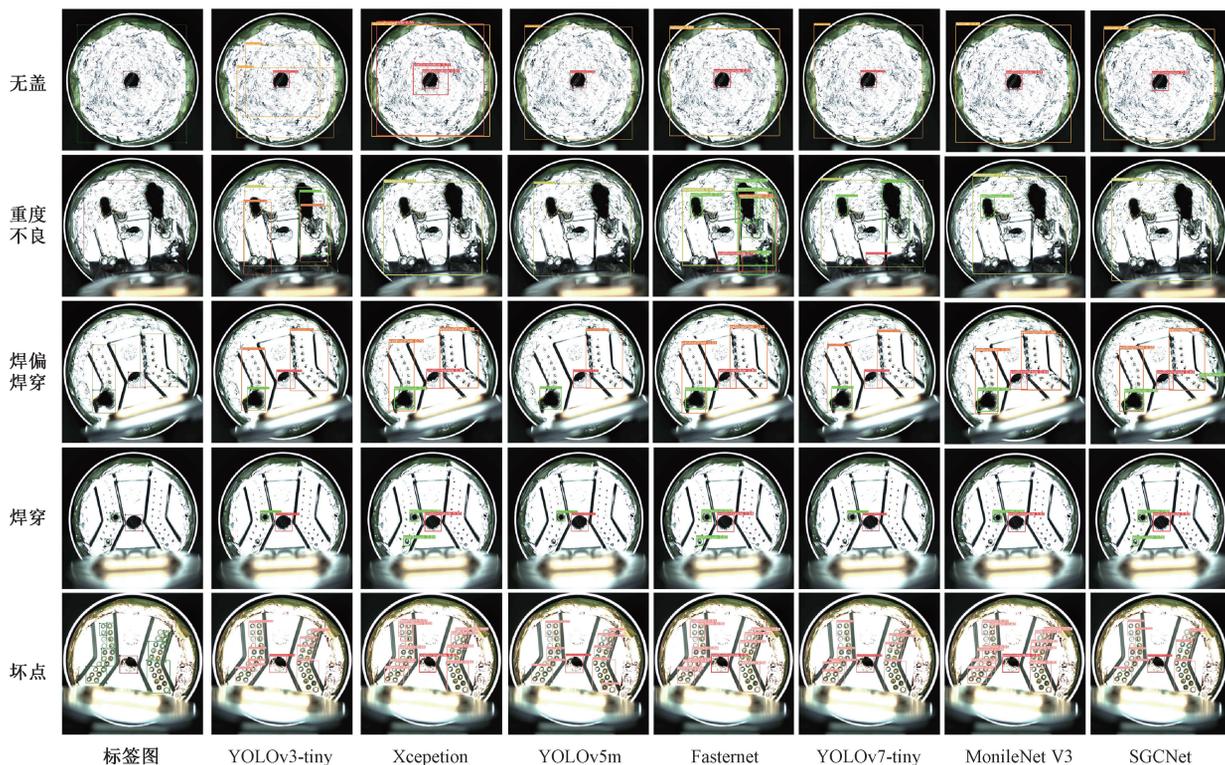


图 7 实验效果对比

Fig. 7 Comparison chart of experimental effects

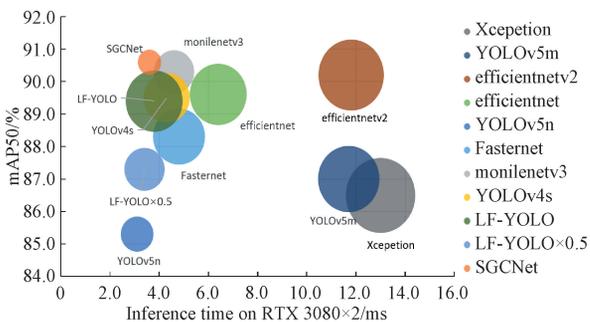


图 8 各模型综合性能对比

Fig. 8 Comprehensive performance comparison chart of each model

TensorRT 加速的检测时间可以稳定在 0.07 s, 且检测效果良好, 满足实际工业实时检测要求。

4 结 论

本文针对新能源汽车电池集流盘的 5 种缺陷检测问题, 提出 SGCNet 模型。将 ShuffleNet V2 作为主干网络, 在保证模型轻量化的同时, 仍能够获得接近于更深层次模型的精度表现; 设计 GC_FPN 作为特征融合网络, 使模型更好地关注图像中重要语义信息的同时, 减少了模型参数量, 降低了计算复杂度, 提高了训练和预测的速度。

通过实验结果可以看出, 本文提出的 SGCNet 模型在自制的新能源汽车电池集流盘缺陷数据集上的检测精度达到了 90.6%, 模型体积仅为 3.2 M, GFLOPs 为 3.6, 在精度方面领先于 SOTA 对象检测方法, 增强效果显著, 综合性能更强。将 SGCNet 模型部署到嵌入式设备 NVIDIA Jetson Nano 上进行实时检测, 每张图片检测时间为 0.07 s, 满足工业检测任务的实施要求。

未来将专注于更多工业产品的表面缺陷检测, 为本文框架中的每个模块设计更加轻量高效的网络结构。此外, 将继续优化本文中的方法, 通过模型剪枝或知识蒸馏等模型轻量化方法, 在保持检测精度的同时提高检测速度。

参考文献

[1] 王成健. 新能源汽车动力电池检测维护方法 [J]. 时代汽车, 2023 (8): 171-173.
WANG CH J. Mew mnergy vehicle power battery testing and maintenance methods [J]. Auto Time, 2023 (8): 171-173.

[2] 张男, 杨强. 用于电池的集流盘以及电池: 福建省, CN216488160U [P]. 2022-05-10.
ZHANG N, YANG Q. Collecting tray and battery for battery: FujianProvince, CN216488160U [P]. 2022-05-10.

[3] IANDOLA F N, HAN S, MOSKEWICZ M W, et al.

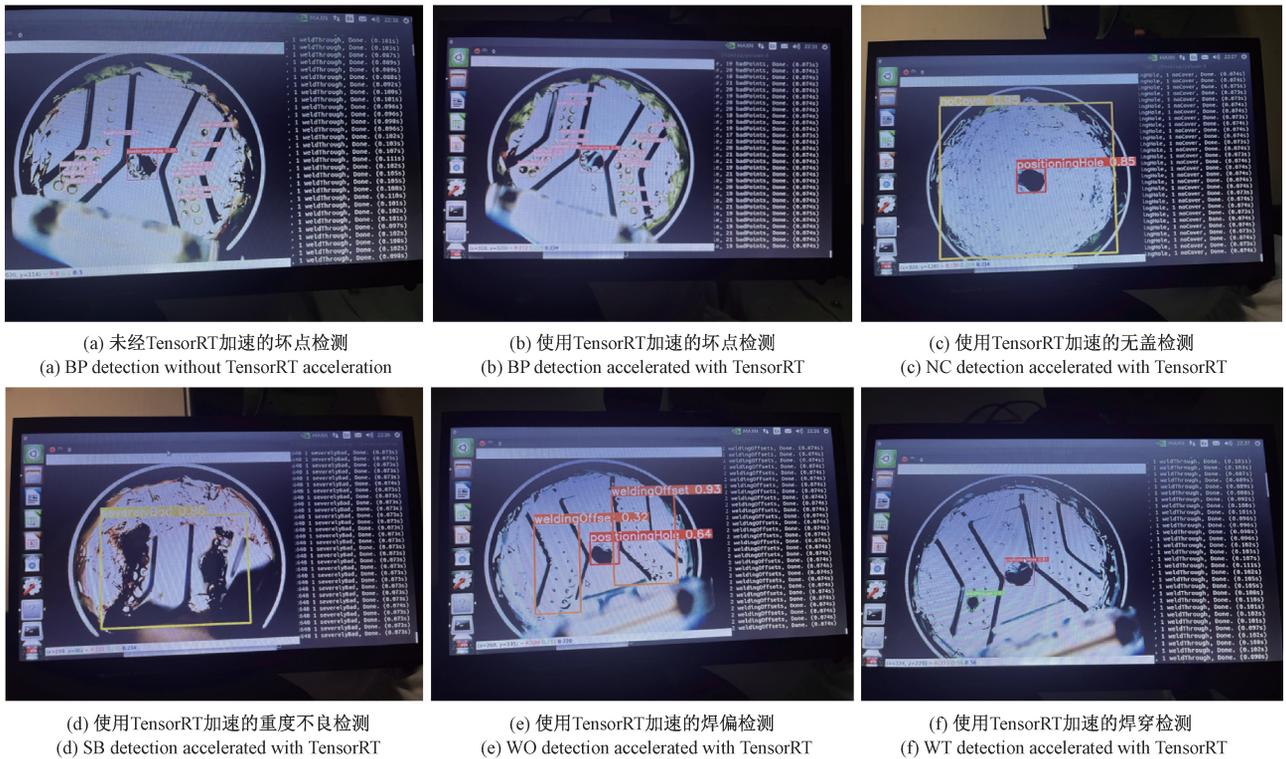


图 9 移动端实时检测效果图

Fig. 9 Real time detection rendering for mobile devices

SqueezeNet; AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size [J]. arXiv preprint arXiv:160207360, 2016.

- [4] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. arXiv preprint arXiv: 170404861, 2017.
- [5] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]. Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [6] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [7] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, PMLR, 2019: 6105-6114.
- [8] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [9] CHEN J, KAO S H, HE H, et al. Run, don't walk;

Chasing higher FLOPS for faster neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.

- [10] TAN M, LE Q. Efficientnetv2: Smaller models and faster training [C]. International Conference on Machine Learning. PMLR, 2021: 10096-10106.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [12] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [13] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv: 180402767, 2018.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv:200410934, 2020.
- [15] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. arXiv preprint arXiv:220902976, 2022.
- [16] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-

- art for real-time object detectors [J]. arXiv preprint arXiv:220702696, 2022.
- [17] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. arXiv preprint arXiv:210708430, 2021.
- [18] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [19] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [20] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28.
- [21] 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.
- [22] 钱万明, 朱红萍, 朱泓知, 等. 基于自适应加权特征融合的 PCB 裸板缺陷检测研究 [J]. 电子测量与仪器学报, 2022, 36(10): 92-99.
QIAN W M, ZHU H P, ZHU H ZH, et al. Research on defect detection of PCB bare board based on adaptive weighted feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(10): 92-99.
- [23] 刘艳菊, 王秋霖, 赵开峰, 等. 基于卷积神经网络的热轧钢条表面实时缺陷检测 [J]. 仪器仪表学报, 2021, 42(12): 211-219.
LIU Y J, WANG Q J, ZHAO K F, et al. Real-time defect detection of hot rolling steel bars based on convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2021, 42(12): 211-219.
- [24] 李衍照, 于镭, 田金文. 基于改进 YOLOv5 的金属焊缝缺陷检测 [J]. 电子测量技术, 2022, 45(19): 70-75.
LI Y ZH, YU L, TIAN J W. Metal weld defect detection based on improved YOLOv5 [J]. Electronic Measurement Technology, 2022, 45(19): 70-75.
- [25] 钱承山, 沈有为, 孙宁, 等. 改进 YOLOv5s 的遥感图像检测研究 [J]. 国外电子测量技术, 2022, 41(11): 57-66.
QIAN CH SH, SHEN Y W, SUN N, et al. Improved YOLOv5s remote sensing image detection research [J]. Foreign Electronic Measurement Technology, 2022, 41(11): 57-66.
- [26] 马燕婷, 赵红东, 阎超, 等. 改进 YOLOv5 网络的带钢表面缺陷检测方法 [J]. 电子测量与仪器学报, 2022, 36(8): 150-157.
MA Y T, ZHAO H D, YAN CH, et al. Strip steel surface defect detection method by improved YOLOv5 network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(8): 150-157.
- [27] 陈彦蓉, 高刃, 吴文欢, 等. 改进 YOLOv5 的新能源电池集流盘缺陷检测方法 [J]. 电子测量与仪器学报, 2023, 37(5): 58-67.
CHEN Y R, GAO R, WU W H, et al. A defect detection method for new energy battery collector disc based on improved YOLOv5 network [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(5): 58-67.
- [28] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116-131.
- [29] LIN T Y, DOLLÁR 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.
- [30] WANG J, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [31] LIU M, CHEN Y, XIE J, et al. LF-YOLO: A lighter and faster YOLO for weld defect detection of X-ray image [J]. IEEE Sensors Journal, 2023, 23(7): 7430-7439.

作者简介



袁磊, 现为湖北汽车工业学院硕士研究生, 主要研究方向为图像处理和目标检测。

E-mail: madness_lei@163.com

Yuan Lei is now a M. Sc. candidate of Hubei Institute of Automotive Technology. His main research interests include image

processing and object detection.



唐海(通信作者), 分别于 1998 年、2006 年和 2010 年获得武汉大学学士、硕士和博士学位, 现为湖北汽车工业学院电气与信息工程学院副教授, 主要研究方向为知识工程和图像处理。

E-mail: smile-tang@163.com

Tang Hai (Corresponding author) received his B. Sc., M. Sc. and Ph. D. degrees from Wuhan University in 1998, 2006, and 2010 respectively. He is now an associate professor at the School of Electrical and Information Engineering, Hubei University of Automotive Technology. His main research interests include the fields of knowledge engineering and image processing.