

面向废旧电缆检测的轻量化网络模型

饶日昕¹, 王怡文¹, 曾砺志², 童心恬¹, 赵海涛^{1*}

(1. 华东理工大学信息科学与工程学院, 上海 200237; 2. 华东理工大学机械与动力工程学院, 上海 200237)

摘要: 目前我国废旧电缆回收主要依靠人工分拣, 存在耗时耗力、准确性低的问题。为了更好地将模型部署至小型化硬件设备并提高检测实时性, 提出基于改进 YOLOv5s 的废旧电缆检测网络模型。首先, 将主干网络中的标准卷积模块替换为轻量化的幻象(Ghost)模块以减小网络的复杂度, 并且在快速空间金字塔池化(SPPF)模块前引入卷积块注意力模块(CBAM), 提高了特征提取和融合的效率。其次, 将网络中 Neck 部分的 C3 模块结合有效通道注意力模块(ECA), 实现了跨通道的信息交互, 提高了网络特征融合能力。最后, 在损失函数的计算部分使用 Wise-交并比(WIoU)作为新的边界框损失函数以提升回归效果, 加快模型的收敛速度。实验结果表明: 改进模型的平均检测精度为 96.3%, 相比单点多框检测器(SSD)提高了 1.2 个百分点; 参数量为 5.15×10^6 , 相比 YOLOv5s 模型减少了 27.0%; 在小型嵌入式设备 LubanCat-1 上的推理速度达到 8.49 帧/s, 具有良好的实时性, 适用于废旧电缆的实时检测与分类。

关键词: 废旧电缆检测; YOLOv5s 模型; 轻量化; 注意力机制; 嵌入式设备

中图分类号: TP391.41

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0068537

Lightweight Network Model for Waste Cable Detection

RAO Rixin¹, WANG Yiwen¹, ZENG Lizhi², TONG Xintian¹, ZHAO Haitao^{1*}

(1. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China;

2. School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai 200237, China)

【Abstract】 Currently, waste cable recycling in China relies primarily on manual sorting, which is time-consuming, labor-intensive, and inaccurate. To deploy the model on small embedded devices more effectively and improve real-time detection, an improved lightweight network model based on YOLOv5s for waste cable detection is proposed herein. First, standard convolution modules in the Backbone of the network are replaced with lightweight Ghost modules to reduce network complexity, and a Convolutional Block Attention Module (CBAM) is introduced before a Fast Spatial Pyramid Pooling (SPPF) module to enhance feature extraction and fusion efficiency. Second, the C3 module in the neck of the network is combined with an Effective Channel Attention (ECA) module to facilitate inter-channel information interaction and enhance the network's feature fusion capability. Finally, Wise Intersection over Union (WIoU) is utilized as a new bounding box loss function to improve the regression effect and accelerate the model convergence speed. The experimental results demonstrate that the improved model achieves an average detection accuracy of 96.3%, which is 1.2 percentage points higher than that of the Single Shot multibox Detector (SSD). There are 5.15×10^6 parameters in the proposed model, which is a 27.0% reduction compared with the YOLOv5s model. The inference speed on the small embedded device, LubanCat-1, reaches 8.49 frame/s, indicating excellent real-time performance and suitability for real-time detection and classification of waste cables.

【Key words】 waste cable detection; YOLOv5s model; lightweight; attention mechanism; embedded device

0 引言

近年来,随着电力、通信等行业的不断发展,废旧电缆的数量日益增加。据统计数据表明,2021 年我国电力领域铜消费量为 605 万吨,占比达 40.70%,主要应用于电线电缆^[1]。长期以来,我国铜矿资源禀赋较差,铜资源主要依靠进口。废旧电线电缆中的铜是铜资源的重要来源之一,2021 年铜

二次资源回收量达到 438.6 万吨,约占总消费量的 31.6%^[2]。因此,对废旧电缆进行回收再利用可以大幅缓解我国铜资源供应压力,具有节约能源、保护环境的重大意义。

传统的电缆回收方法是将废旧电缆从管道(沟)中抽出后运至回收仓库或临时堆放处,再将所有旧电缆堆积在一起放置在场地上,按重量进行估价回收处理。然而,未进行任何分类就回收废旧电缆将

收稿日期: 2023-10-10 修回日期: 2023-11-05

基金项目: 国家自然科学基金(62173143)。

通信作者 E-mail: *haitaohao@ecust.edu.cn

造成巨大的资源浪费,且人工分拣耗时耗力,准确性低,受主观因素影响大。

在过去 10 多年中,基于卷积神经网络(CNN)的目标检测技术发展迅速,相比于传统方法,该技术具有速度快、精度高、客观性强等特点。随着目标检测技术在工业领域的应用需求日益突出,催生了各种类别的目标检测算法。目前,主流的目标检测算法有两类:一类是以更快速的区域卷积神经网络(Faster R-CNN)^[3]为代表的双阶段检测算法;另一类是以 YOLO 系列^[4-7]为代表的单阶段检测算法,无需独立地显示图片中的候选区域,直接从输入图像中提取特征即可实现目标检测,检测速度更快,更适合实时性高的检测任务。

随着时代的发展,人们更加关注神经网络的实际应用性能。有限的存储空间和算力资源制约了高精度模型在嵌入式平台上的部署。神经网络模型在移动端部署的需求迫切,如无人机^[8]、人脸识别系统^[9]等。为了解决这类问题,许多学者^[10-15]提出了 MobileNet、ShuffleNet、EfficientNet 等轻量级网络,以准确率的小幅度降低为代价大幅度减少了模型参数与运算量,节约了运算成本,提升了网络的实用性。

为克服传统回收方法的不足并兼顾废旧电缆检测过程的精确性和实时性,本文以 YOLOv5s 目标检测网络模型为基础,提出用于废旧电缆检测的改进模型,改进后的模型对电缆截面图像有更好的针对性,兼具效率和准确率。

1 YOLOv5s 网络模型

YOLO 算法是基于锚框回归的单阶段目标检测算法,核心思想是把目标检测问题处理成回归问题,用一个端到端的卷积神经网络模型直接从输入图像中提取相应的特征信息,预测目标物体所处的位置和所属类别的概率,具有速度快、效率高的特点。

YOLOv5 是 YOLO 算法系列的第 5 个版本, YOLOv5s 网络的深度与特征图的宽度较小,检测性能良好,相比体型较大、结构复杂的网络模型更适配于小型嵌入式设备。其由 4 个主要部分构成,分别为输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和检测网络(Prediction)。Input 部分通过 Mosaic 数据增强的方式对输入图像进行预处理; Backbone 部分先通过 Focus 结构对图像进行切片,在不丢失信息的情况下实现下采样,再通过跨阶段局部(CSP)^[16]结构和快速空间金字塔池化(SPPF)

结构生成新的特征图; Neck 部分采用特征金字塔网络(FPN)^[17]和路径聚合网络(PAN)^[18]进行特征融合; Prediction 部分生成检测框并预测目标的分类。

相较于 YOLOv4, YOLOv5 训练过程更为迅速,且能兼顾检测精度与检测速度,更适配于小型的嵌入式设备。因此,将 YOLOv5s 作为基础框架,提出更适配于废旧电缆识别功能的改进网络模型。

2 改进的 YOLOv5s 网络模型

由于废旧电缆的检测工作同时对精确性与实时性有一定的要求,且现存的深度网络模型大多属于泛化性的模型,即对多种物体都能达到一定的识别效果,但对特定类别目标的检测性能不够突出。因此,为了能针对性地用于废旧电缆识别,达到足够的检测精度,还需要根据实际情况进行具体的调整。

本文以 YOLOv5s 网络作为基础框架进行改进,将其主干部分中的基础卷积模块和 C3 模块分别替换为幻象(Ghost)模块与 C3Ghost 模块,减小网络参数量与计算量。针对电缆回收工作环境的特点,在 SPPF 模块前引入卷积块注意力模块(CBAM),聚焦空间与通道两种维度中的特征信息,加强网络对电缆图像的关注度。为减少电缆切割与摆放方式的不确定性对图像信息的不良影响,在特征融合层中将有效通道注意力(ECA)模块嵌入 C3 模块,提高网络对复杂情况的适应度,改善特征融合的效果。针对自制废旧电缆数据集存在的标注问题,采用新的边框损失函数 Wise-交并比(WIoU)以适配实验数据集,提升模型的学习效率。最终改进的 YOLOv5s 网络整体结构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。

2.1 Ghost 模块

为从图像中提取更丰富的信息,加深对数据的理解,人们通常倾向于构建更深更复杂的网络模型。然而,由于内存和计算资源有限,在嵌入式设备中部署大型网络是十分困难的。文献^[19]提出的 Ghost 模块能有效利用特征图中的冗余信息,以一系列低成本的操作产生新的特征图,在保证精度的同时降低模型计算量。

设输入数据 $\mathbf{X} \in \mathbb{R}^{c \times h \times w}$, c 为输入数据的通道数, h 和 w 分别为高和宽,则生成 n 个通道输出的普通卷积操作可以表示为 $\mathbf{Y} = \mathbf{X} \times f + b$ ($\mathbf{Y} \in \mathbb{R}^{h' \times w' \times n}$, $f \in \mathbb{R}^{c \times k \times k \times n}$), h' 和 w' 为输出特征图的高和宽, k 为卷积核的大小。在该普通卷积操作中,每秒浮点运算数(FLOPs)的计算公式可表示如下:

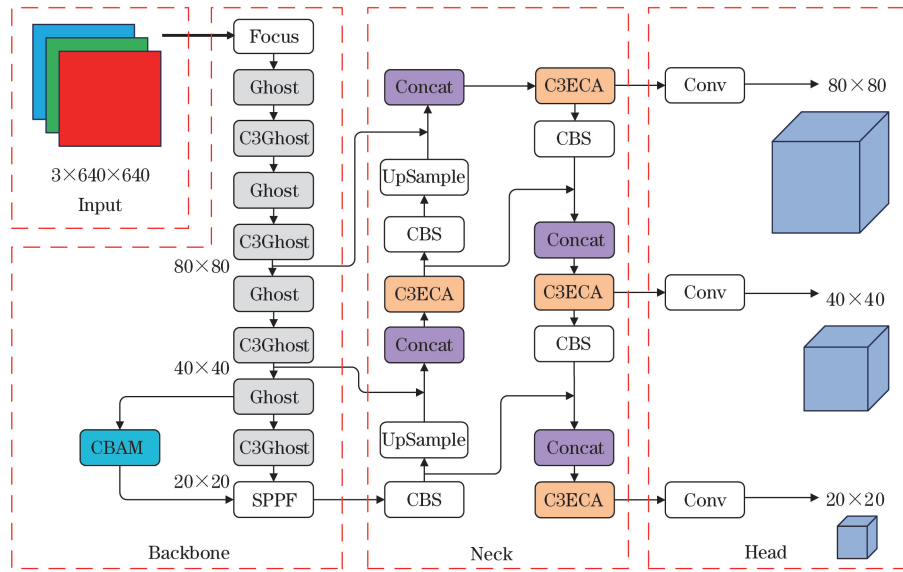


图 1 改进的 YOLOv5s 网络结构

Fig.1 Structure of improved YOLOv5s network

$$f_1 = n \times h' \times w' \times c \times k \times k \quad (1)$$

如图 2 所示,为优化参数量,Ghost 模块先经过普通卷积 $Y' = X \times f'$ ($f' \in \mathbb{R}^{c \times k \times k \times m}$, $m \leq n$) 获得原始特征图 $Y' \in \mathbb{R}^{h' \times w' \times m}$,再对 Y' 的每个通道进行更为低成本的线性运算 Φ ,生成通道数为 s 的幻象特征图,最终获得通道数为 $n = m \times s$ 的特征图 Y 。在该过程中 FLOPs 的计算公式可表示如下:

$$f_2 = \frac{n}{s} \times h' \times w' \times c \times k \times k + (s-1) \times \frac{n}{s} \times h' \times w' \times d \times d \quad (2)$$

式中: c 为输入数据的通道数; n 为输出特征图的通道数; h' 和 w' 为输出特征图的高和宽; k 为卷积核的大小; s 为每个幻象特征图的通道数; d 为线性运算过程的平均内核大小。

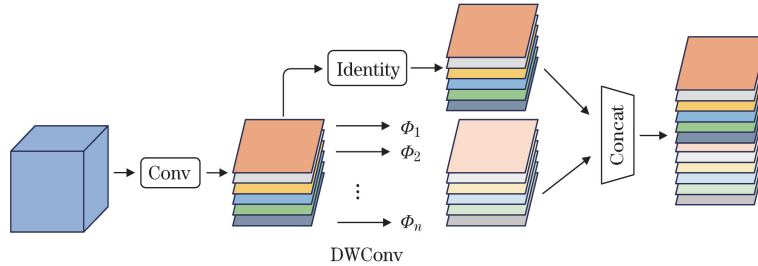


图 2 Ghost 模块结构

Fig.2 Structure of Ghost module

由此可计算得到使用 Ghost 模块代替普通卷积的理论加速比:

$$r_s = \frac{f_1}{f_2} = \frac{n \times h' \times w' \times c \times k \times k}{\frac{n}{s} \times h' \times w' \times c \times k \times k + (s-1) \times \frac{n}{s} \times h' \times w' \times d \times d} \approx \frac{s \times c}{s + c - 1} \approx s \quad (3)$$

在输出特征图大小相同的情况下,Ghost 模块所需的参数量与计算量为普通卷积模块的 $1/s$ 。

如图 3 所示,与 ResNet^[20] 中基本残差块的设计类似,GhostBottleneck 是由 Ghost 模块组成的瓶颈结构,主要由 2 个串联的 Ghost 模块组成:第 1 个用于扩展层,增加输入特征图的通道数;第 2 个用于减少通道数,以与 shortcut 尺寸匹配。使用 shortcut 连接这 2 个 Ghost 模块的输入和输出。借

鉴 MobileNetV2,第 2 个 Ghost 模块之后不使用 ReLU,其他每层后都应用批量归一化(BN)和 ReLU 非线性激活。上述描述适用于步长(stride)为 1 的情况,对于步长为 2 的情况,shortcut 路径由下采样层和深度卷积(DWConv)实现。

将原始网络 C3 模块中的基本 Bottleneck 模块更改为 GhostBottleneck,称为 C3Ghost 模块,用于替换原始主干网络中的 C3,从而在兼顾检测精度的同时减少网络的参数量与模型规模。

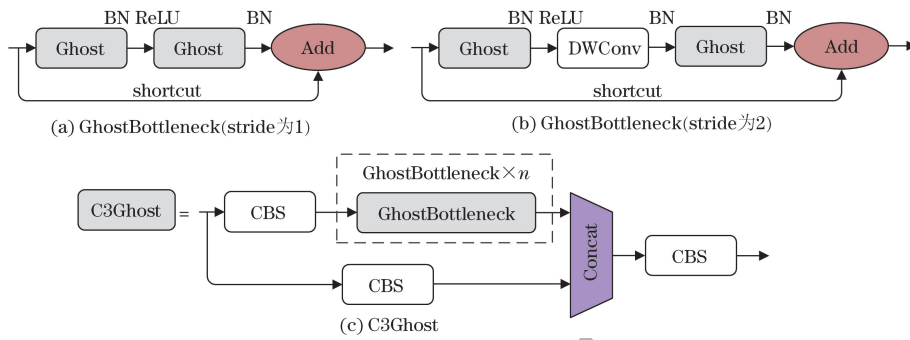


图 3 GhostBottleneck 与 C3Ghost 结构

Fig.3 Structure of GhostBottleneck and C3Ghost

2.2 CBAM

使用注意力机制的目的是将注意力集中在信息需要关注的部位,而过滤掉无关信息。在废旧电缆的回收仓库中,为了使模型能更加适应因不同光照、杂物、背景等导致的复杂多变的周围环境,更加关注图像中待测电缆的截面部分而尽可能地忽略其他无用信息,提高回收工作的效率,在模型中引入了 CBAM。

CBAM^[21]包含通道注意力模块(CAM)与空间注意力模块(SAM)两个子模块,其结构如图 4 所示。

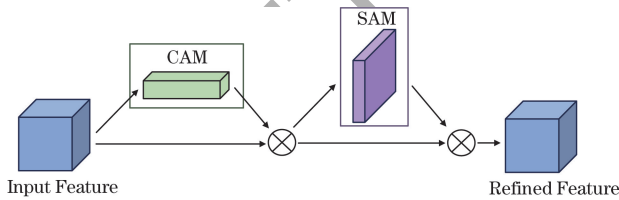


图 4 CBAM 模块结构

Fig.4 Structure of CBAM module

通道注意力可以让网络学习不同通道之间的联系与重要程度,更加关注对检测任务起更大作用的通道的特征。空间注意力可以让网络更多地提取图像中与分类有关的像素区域而过滤掉其他无关信息,从而提高检测网络对废旧电缆截面图像的敏感性,提高网络的学习能力。

2.3 ECA

在工作场景下的电缆检测环境相对复杂,不同废旧电缆的切割方式与摆放方式存在一定的偏差且电缆截面在图像中的位置与大小均易发生变化,为使网络模型在进行信息提取的过程中能有效地学习这些关键信息,提升对废旧电缆检测的泛化能力,同时平衡网络性能与复杂度之间的关系,本文引入 ECA 模块。

通过分析挤压和激励网络(SENNet)^[22]中维度缩减对通道注意力预测带来的负面影响,文献[23]

针对 SE 注意力模块的不足提出了一种轻量化的 ECA 模块,避免了降维,提高了通道注意力的学习效率。使用一维卷积层代替全连接层,减少了参数量,使每次卷积过程中只有部分通道起作用,更好地完成了跨通道的信息交互,使网络在只引入很少参数的情况下也能有较好的表现。

ECA 模块结构如图 5 所示,特征图首先进行全局平均池化(GAP)获得聚合特征,然后根据通道数自适应地计算一维卷积核的大小并进行一维卷积,接着通过 Sigmoid 激活函数学习每个通道的权重。自适应计算卷积核大小的公式可表示如下:

$$k = \left\lfloor \frac{\ln c + b}{\gamma} \right\rfloor_{\text{odd}} \quad (4)$$

式中: k 表示卷积核的大小; c 表示通道数; γ 和 b 为调整自适应计算的参数; $\lfloor \cdot \rfloor_{\text{odd}}$ 表示 k 取奇数。

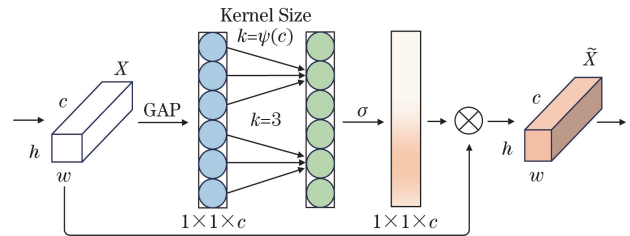


图 5 ECA 模块结构

Fig.5 Structure of ECA module

ECA 模块不仅可以作为独立的注意力层加入网络,也可以与 YOLOv5s 中的其他模块相结合。如图 6 所示,选择将 ECA 模块嵌入特征融合网络 Neck 部分的 C3 模块,加强通道合并后网络信息提取的能力,以增加少量参数为代价获得更高的检测性能。

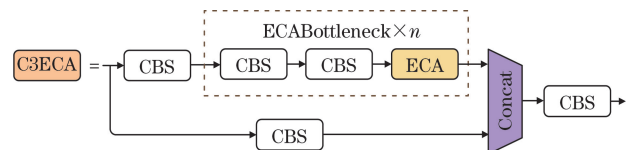


图 6 C3ECA 模块结构

Fig.6 Structure of C3ECA module

2.4 WIoU 损失函数

交并比(IoU)是指检测任务中预测框与真实框相交部分的面积与合并部分的面积之比,用于衡量预测结果的准确性,IoU 损失函数的计算可表示如下:

$$L_{IoU} = 1 - R_{IoU} \quad (5)$$

式中: R_{IoU} 为预测框与真实框的交并比。

YOLOv5s 算法默认采用完全交并比(CIoU)^[24]作为边框损失函数,CIoU 损失函数定义如下:

$$L_{CIoU} = 1 - R_{IoU} + R_{CIoU} \quad (6)$$

$$R_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (7)$$

式中: b 和 b^{gt} 为预测框和真实框的中心点; $\rho^2(b, b^{gt})$ 是两个中心点间的欧氏距离; c 是包含预测框和真实框的最小区域的对角线长度; R_{CIoU} 为惩罚项; $\alpha = \frac{v}{(1 - R_{IoU}) + v}$ 为权重函数, $v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2$ 用于度量长宽比, w^{gt} 、 h^{gt} 和 w 、 h 分别为真实框和预测框的宽、高。

目前,现有的公开数据集对废旧电缆截面的研究较为缺乏,故实验数据主要来源于自主采集与标注,其中不可避免地会包含一些低质量的标注边框。若继续使用 CIoU 损失函数,则长宽比、中心距等几何因素将增加低质量数据的训练惩罚而降低模型的鲁棒性,且训练使用的数据集中真实框的长宽比多接近 1,由于数据的长宽比过于单一会导致模型在不同情况下接收到的梯度信息较为相似,影响模型的学习能力和泛化性能,因此度量长宽比这一过程对加快模型收敛的贡献较小,本文决定改用 WIoU^[25]作为边框损失函数,加快模型收敛速度,削弱低质量样本几何度量的不良影响,提高模型整体性能。

在 WIoU v1 和 WIoU v2 的基础上,WIoU v3 通过计算梯度增益(聚焦系数)设计动态非单调的聚焦机制,使 WIoU 更适用于普通质量的锚框,计算公式可表示如下:

$$L_{WIoU v3} = r \times R_{WIoU} \times L_{IoU} \quad (8)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (9)$$

$$\beta = \frac{L_{IoU}^*}{L_{IoU}}, \beta \in [0, +\infty) \quad (10)$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (11)$$

式中: x_{gt} 、 y_{gt} 和 x 、 y 分别为真实框和预测框的中心点坐标; W_g 和 H_g 分别为包含真实框和预测框的最小区域的宽和高; R_{WIoU} 为惩罚项; β 为离群度,用来衡量锚框的质量,离群度越小边框质量越高; r 为非单调聚焦系数; α 和 δ 为超参数,可根据不同模型与数据集进行调整。

3 实验结果与分析

3.1 数据集准备

目前,国内对于电缆截面图像的相关研究较少,没有统一公开的电缆截面图像数据集,因此实验数据主要来源于自主采集。使用的相机为海康威视 TBS060-10GC,像素为 600 万。数据集共有 1 801 张图像,图像尺寸为 3 072×2 048 像素,将所有电缆图像分为 10 类,如图 7 所示,并将每类以 3:1:1 的比例分别划入训练集、验证集和测试集。为使数据集更接近于实际电缆截面检测场景,丰富数据集的多样性,便于后期模型训练和提高检测精度,考虑以下 3 个场景的拍摄并对图像进行数据加强处理(如图 8 所示):

- 1) 复杂环境拍摄,由于不同天气、光照条件可能会影响检测结果,因此通过设置不同亮度的方式进行拍摄。
- 2) 多样背景拍摄,由于实际背景的复杂程度不一致,因此通过添加遮挡物、更换背景等方式进行拍摄。
- 3) 不等距拍摄,由于待测电缆与相机的距离会影响图像清晰度,因此通过对焦和不对焦两种方式进行拍摄。

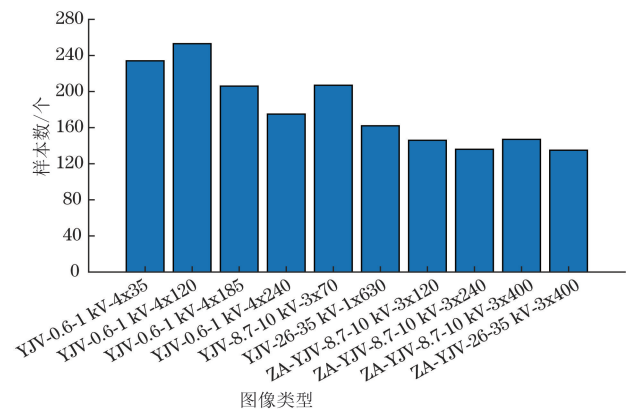


图 7 废旧电缆数据集分类

Fig. 7 Classification of waste cable dataset

3.2 实验环境

实验使用 Linux Ubuntu 20.04 操作系统,处理器为 15 核 CPU Intel® Xeon® Platinum 8358P CPU @ 2.60 GHz,90 GB 内存;NVIDIA GeForce RTX 3090 GPU,24 GB 显存。编程语言使用

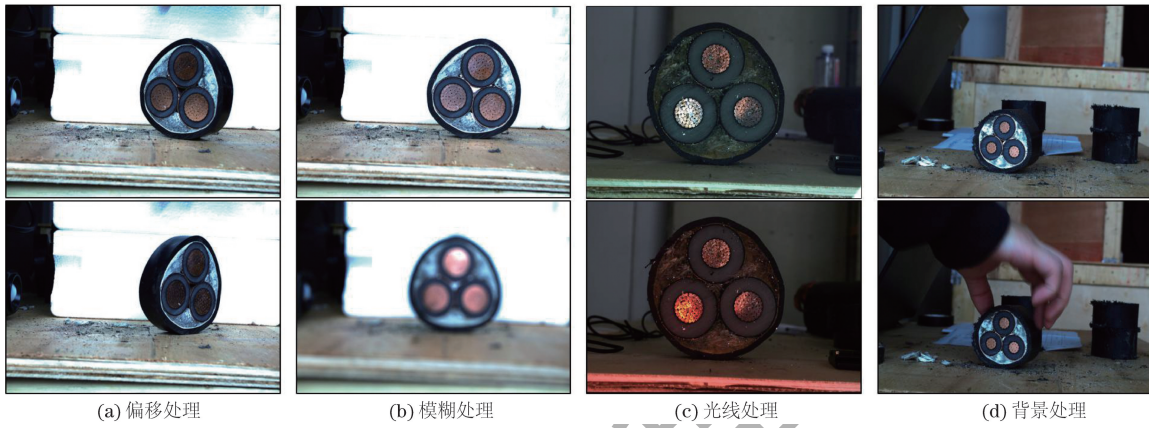


图 8 数据加强处理

Fig.8 Data augmentation processing

Python 3.8.10, 开发环境为 PyTorch 1.9.0、CUDA 11.1。基于 PyTorch 深度学习框架构建网络模型进行废旧电缆检测模型的训练和验证。在训练时采用的参数配置如表 1 所示。在训练完成后的模型在小型嵌入式开发板 LubanCat-1 上进行推理测试, 相应的配置如表 2 所示。

表 1 训练参数配置

Table 1 Configuration of training parameters

名称	配置
训练轮次/轮	300
批次大小	16
输入尺寸/像素	640
优化器	随机梯度下降 (SGD)
	[10, 13, 16, 30, 33, 23]
先验框尺寸/像素	[30, 61, 62, 45, 59, 119]
	[116, 90, 156, 198, 373, 326]

表 2 嵌入式开发板配置

Table 2 Configuration of embedded development board

名称	配置
主控	RK3566 4 核 A55, 1.8 GHz 1 TOPS NPU
eMMC 存储/GB	32
LPDDR4X 内存/GB	4

3.3 评估指标

主要使用平均精度均值 (mAP)、帧率 (FPS)、参数量、FLOPs、模型大小等客观指标来评价废旧电缆检测模型。mAP 是各废旧电缆类别平均精度 (AP) 的均值, 计算公式如式 (12) 所示, 其中, n 表示识别电缆类别的总数。AP 为精确率对召回率的积分, 计算公式如式 (13) 所示。精确率和召回率的计算公式如式 (14) 和式 (15) 所示。mAP@0.5 表示预测阈值为 0.5, 即当预测框与真实框的 IoU 大于 0.5 且预测类别正确时, 认为预测结果正确, 该次检测记为真正例 (N_{TP}), 否则记为假正例 (N_{FP}), 当图像中

存在废旧电缆但模型未检测出目标对象时该次检测记为假负例 (N_{FN})。帧率表示网络每秒可检测图片的数量, 单位为帧/s, 通常用来评估模型实时性。同时, 通过比较参数量和计算量, 能更全面地考量模型的综合性能。

$$P_{mAP} = \frac{1}{n} \sum_{i=1}^n P_{AP_i} \quad (12)$$

$$P_{AP} = \int_0^1 P(R) dR \quad (13)$$

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (14)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (15)$$

3.4 对比实验

为了更有效地评估改进模型的性能, 以 Faster R-CNN、单点多框检测器 (SSD)^[26] 与其他 YOLO 系列网络模型为对照组, 采用相同的数据集、训练参数及方法, 对改进模型与各种目标检测模型进行对比实验, 如表 3 所示, 其中, “—”表示未对相关数据进行测试。

相对于 YOLOv5s 模型, 改进模型大小减少了 26.1%, FLOPs 下降了 31.9%, 模型的复杂度得到了明显的优化, 而整体 mAP@0.5 仅下降了 1.7 个百分点。令两模型分别对部分具有多目标、电缆位置偏移、截面受损等特点的图像进行检测, 并标注出目标的位置和分类标签, 标签后的数字代表置信度。可视化的检测效果对比结果如图 9 所示。在图 9(a) 中, 对于含多个目标的图像, 由于 YOLOv5s 参数量较多, 对单目标的训练样本过拟合, 发生了分类错误。在图 9(b) 中, 电缆位置发生了偏移, 而改进模型对目标生成了两个识别框, 即认为其所属类别有两种可能性, 导致整体的 mAP 有所下降。这是由于目标发生偏移后, 图像特征也发生了一定的

表 3 不同模型对比实验结果

Table 3 Comparative experimental results of different models

模型	参数量/ 10^6	FLOPs/ 10^9	模型大小/MB	mAP@0.5/%	FPS/(帧·s ⁻¹)
Faster R-CNN	137.10	370.2	108.5	99.1	—
SSD	26.29	62.7	95.2	95.1	—
YOLOv3	61.58	155.4	117.0	92.2	2.34
YOLOv3-tiny	8.69	13.0	16.6	97.1	12.89
YOLOv4	64.36	60.5	244.6	95.0	—
YOLOv4-tiny	6.06	7.0	22.6	80.5	—
YOLOv5s	7.05	16.0	13.8	98.0	9.60
改进模型	5.15	10.9	10.2	96.3	8.49

变化,与识别框上标注的两种型号均有一定的相似性,但这也表明网络具有一定的泛化能力。若要识别出唯一确定的结果,则可以通过修改置信度的阈值、在发生重复识别时对电缆位置进行调整等。在图 9(c)中,对于截面受损严重的情况,YOLOv5s 的分类结果错误,而加入了相应注意力模块的改进模型具有更强的特征提取能力,对检测目标受损的图

像具有更强的鲁棒性,得到了正确的结果。Faster R-CNN 具有最高的识别精度,但其模型大小、参数量均不适用于嵌入式设备内存、算力有限的情况。相较于 YOLOv3,改进模型不仅在检测精度上获得了 4.1 个百分点的提升,且模型大小不足其 1/10。YOLOv3-tiny 虽然在检测精度上取得了不错的结果,但仍比 YOLOv5s 低 0.9 个百分点,且模型规模更大。

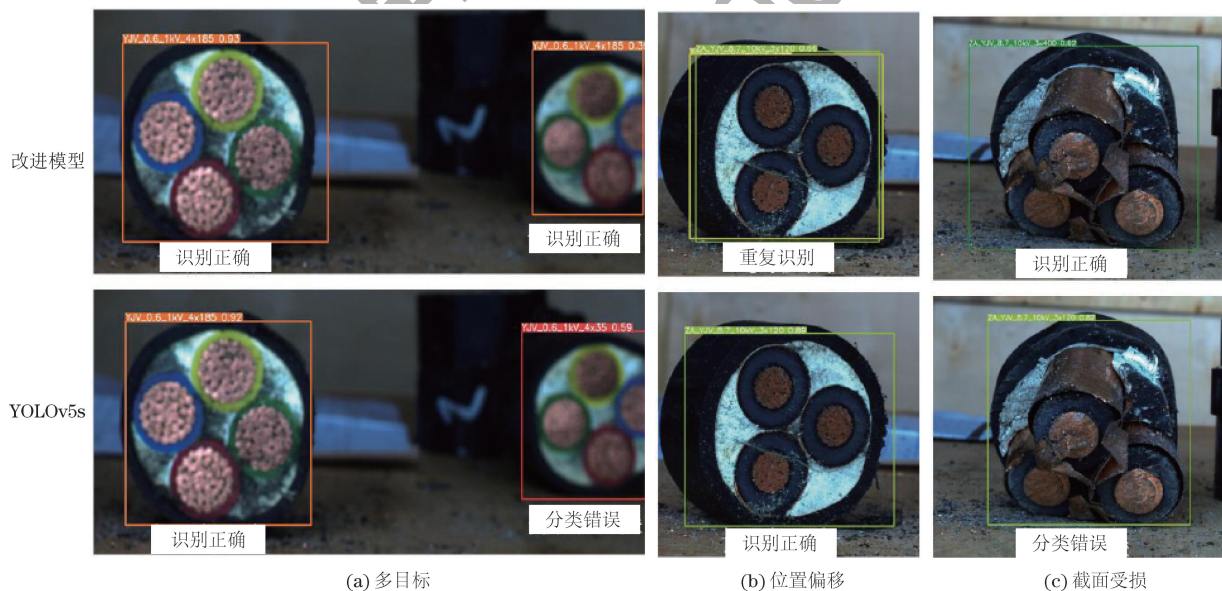


图 9 检测结果可视化对比

Fig. 9 Visual comparison of detection results

为了分析模型在不同环境下识别速度的变化,以不使用神经处理单元(NPU)加速的LubanCat-1与PC端为对照组,使用相同的权重转化的开放神经网络交换(ONNX)模型和Rockchip神经网络(RKNN)模型,将改进模型与各种目标识别模型进行对比实验,对比结果如图10所示。分析发现,由于在Ghost模块中加入了深度可分离卷积,虽然参数量有所下降,但是增加了卷积层数及串行运算量,导致在未经过

针对性优化的设备上使用改进模型进行推理时速度反而会略有下降。

本文在LubanCat-1设备上使用RKNN模型进行推理时,通过NPU加速,帧率可达到PC设备上ONNX模型的38.9%,相比于LubanCat-1设备上使用ONNX模型推理,即不使用NPU加速时,帧率有10倍左右的提升。由此可见,充分调用NPU的加速效果可以使模型在嵌入式设备上推理时达到较高的帧率,以满足实际需求。

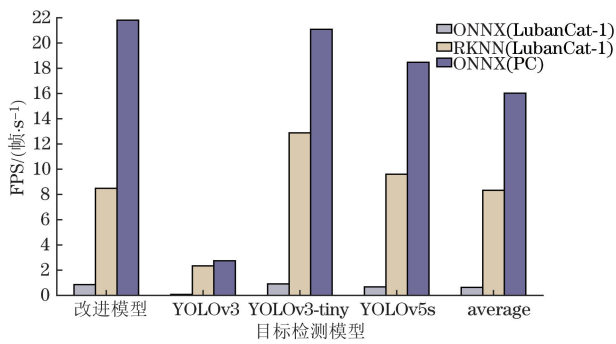


图 10 不同运行环境对比实验结果

Fig. 10 Comparative experimental results of different runtime environments

3.5 消融实验

为了进一步分析比较不同的改进策略对实验结果的影响,通过设置消融实验对比不同的改进策略。实验结果如表 4 所示,其中,“√”表示在 YOLOv5s 模型的基础上使用该模块进行改进,空白表示未使用该模块。

由表 4 可见,在主干网络中使用 Ghost 模块后,模型的参数量减少了 27.5%,但 mAP@0.5 也下降了 26.4 个百分点,模型性能发生了较严重的失真。分析原因可能为大幅减少参数后,模型的学习能力受到了一定程度的影响,而本文实验所使用的废旧

表 4 消融实验结果比较

Table 4 Comparison of ablation experimental results

Ghost	CBAM	C3ECA	WIoU	参数量/ 10^6	mAP@0.5/%	FPS/(帧·s ⁻¹)
				7.047	98.0	9.60
√				5.112	71.6	9.57
√	√			5.145	75.2	8.81
√	√	√		5.145	80.8	8.54
	√		√	7.079	99.4	9.10
		√	√	7.047	99.3	9.42
√			√	5.112	95.9	9.57
√	√	√	√	5.145	96.3	8.49

电缆数据集并非来源于公开数据集,标注质量相对较低,因此模型学习、收敛的速度明显下滑。将损失函数由 CIoU 更换为 WIoU 后,在相同的训练轮次下分别绘制训练损失曲线。如图 11 所示,当使用 WIoU 后,曲线的收敛速度得到了提升。在训练完成后模型的平均检测精度回升至 95.9%,在不额外增加参数量的情况下比使用 CIoU 增加了 24.3 个百分点。

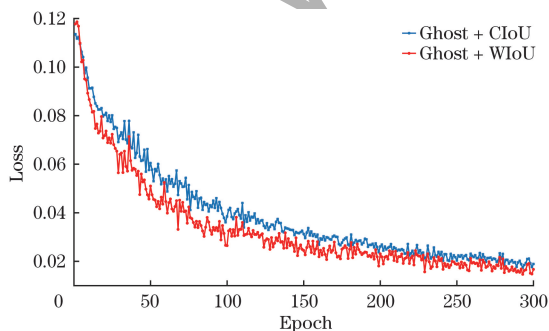


图 11 损失曲线对比

Fig. 11 Comparison of loss curves

为了权衡模型的复杂度与检测精度,在 YOLOv5s 模型的基础上单独增加 CBAM 与 C3ECA 模块,并使用 WIoU 损失函数,使 mAP@0.5 分别增加了 1.4 和 1.3 个百分点,而参

数量几乎没有增加。在使用 Ghost 模块与 CIoU 损失函数的基础上依次添加 CBAM 与 C3ECA 模块,模型的检测精度依次提升了 3.6 和 5.6 个百分点。在使用所有改进模块后,最终的改进模型的参数量为 5.145×10^6 ,为原模型的 73.0%,平均检测精度达到 96.3%,仅比原模型减少了 1.7 个百分点。

4 结束语

本文针对废旧电缆回收任务,提出一种改进的 YOLOv5s 网络模型,并在小型嵌入式设备 LubanCat-1 上实现了部署与推理。首先,将主干网络中的标准卷积模块换为 Ghost 模块以降低模型的复杂度,提高网络的鲁棒性。然后,在特征融合模块前加入 CBAM,分别从通道与空间两个维度出发提取更多关键信息,维持了检测性能的稳定。接着,在颈部网络的 C3 模块中融入 ECA 模块,在几乎不引入额外参数的前提下保证了较高的检测精度。最后,使用 WIoU 作为边框损失函数,提升了训练过程的回归效果。实验结果表明,相较于 YOLOv5s 模型,改进模型以 1.7 个百分点的小幅度精度损失为代价,使得参数量减少了 27.0%、网络规模更小,且在使用 NPU 加速的 LubanCat-1 上具有较快的

检测速度,能适用于嵌入式平台,具有一定的实际应用价值。在后续的研究中,将尝试使用不包含深度可分离卷积的轻量化网络,进一步提升推理效率。

参考文献

- [1] 张楠. 新能源产业发展背景下我国铜资源供需现状与趋势[J]. 中国矿业, 2023, 32(6): 2-9.
ZHANG N. Analysis of supply and demand status and trend of copper resources in China under development background of new energy industry[J]. China Mining Magazine, 2023, 32(6): 2-9. (in Chinese)
- [2] 文博杰,代涛,韩中奎,等. 中国铜资源在用存量与二次供应潜力[J]. 地球学报, 2023, 44(2): 325-332.
WEN B J, DAI T, HAN Z K, et al. Copper in-use stock and recycling potential in China[J]. Acta Geoscientica Sinica, 2023, 44(2): 325-332. (in Chinese)
- [3] RENS 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.
- [4] 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 (CVPR). Washington D. C., USA: IEEE Press, 2016: 779-788.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2017: 6517-6525.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2023-09-17]. <https://arxiv.org/abs/1804.02767>.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. [2023-09-17]. <http://arxiv.org/abs/2004.10934>.
- [8] 江波,屈若锟,李彦冬,等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4): 524519.
JIANG B, QU R K, LI Y D, et al. Object detection in UAV imagery based on deep learning: review [J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 524519. (in Chinese)
- [9] 柯鹏飞,蔡茂国,吴涛. 基于改进卷积神经网络与集成学习的人脸识别算法[J]. 计算机工程, 2020, 46(2): 262-267, 273.
KE P F, CAI M G, WU T. Face recognition algorithm based on improved convolutional neural network and ensemble learning[J]. Computer Engineering, 2020, 46(2): 262-267, 273. (in Chinese)
- [10] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. [2023-09-17]. <http://arxiv.org/abs/1704.04861>.
- [11] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 4510-4520.
- [12] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Washington D. C., USA: IEEE Press, 2019: 1314-1324.
- [13] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 6848-6856.
- [14] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M]. Berlin, Germany: Springer International Publishing, 2018.
- [15] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. [2023-09-17]. <http://arxiv.org/abs/1905.11946>.
- [16] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Washington D. C., USA: IEEE Press, 2020: 1571-1580.
- [17] 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 (CVPR). Washington D. C., USA: IEEE Press, 2017: 936-944.
- [18] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 8759-8768.
- [19] HANK, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 1577-1586.
- [20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2016: 770-778.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 3-19.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7132-7141.
- [23] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE Press, 2020: 11531-11539.
- [24] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 12993-13000.
- [25] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL]. [2023-09-17]. <http://arxiv.org/abs/2301.10051>.
- [26] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.