

DOI:10.19651/j.cnki.emt.2211394

基于通道剪枝的 ACAM-YOLOv5s 绝缘子缺陷检测*

赵立杰 袁昌彪 黄明忠 王国刚 张延华

(沈阳化工大学信息工程学院 沈阳 110142)

摘要: 针对现有的绝缘子缺陷检测深度神经网络模型规模大、计算资源消耗高、检测精度低,难以部署在边缘端,本文基于通道剪枝和 YOLOv5s 方法提出具有非对称卷积和注意力机制的轻量级绝缘子缺陷检测模型 ACAM-YOLOv5s。ACAM-YOLOv5s 模型采用非对称卷积模块 ACBlock 替换 YOLOv5s 骨干网络残差结构中的标准卷积,并结合通道和空间混合的注意力 CBAM 进行特征融合,以增强骨干网络的表达能力、特征提取能力以及鲁棒性。引入对边界框大小和位置灵敏性高的 PIoU 作为定位回归损失,解决绝缘子纵横比高导致缺陷检测定位准确率低的问题。基于 BN 层通道剪枝方法对 ACAM-YOLOv5s 模型进一步稀疏化训练、剪枝和微调,得到轻量化缺陷检测模型。实验结果表明,剪枝后的 ACAM-YOLOv5s 模型和原始 YOLOv5s 相比,在检测精度、计算量和模型体积方面,具有相对优势,能够满足边缘设备部署的需求,在无人机航拍绝缘子缺陷检测领域具有潜在价值。

关键词: 绝缘子缺陷检测;YOLOv5s;非对称卷积;注意力机制;PIoU;通道剪枝

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

ACAM-YOLOv5s insulator defect detection based on channel pruning

Zhao Lijie Yuan Changbiao Huang Mingzhong Wang Guogang Zhang Yanhua

(School of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China)

Abstract: To address the existing deep neural network models for insulator defect detection are large in size, high in computational resource consumption, low in detection accuracy and difficult to deploy at the edge end. In this paper, we propose a lightweight insulator defect detection model ACAM-YOLOv5s with asymmetric convolution and attention mechanism based on channel pruning and YOLOv5s method. The ACAM-YOLOv5s model uses the asymmetric convolution module ACBlock to replace the standard convolution in the residual structure of the YOLOv5s backbone network, combined with the attention CBAM of channel and spatial blending for feature fusion to enhance the expressiveness, feature extraction and robustness of the backbone network. PIoU, which is highly sensitive to the size and position of the bounding box, was introduced as a localisation regression loss to address the problem of low defect detection localisation accuracy due to high insulator aspect ratios. The experimental results show that the pruned ACAM-YOLOv5s model has relative advantages over the original YOLOv5s in terms of detection accuracy, computational volume and model size, which can meet the needs of edge device deployment and has potential value in the field of UAV aerial insulator defect detection.

Keywords: insulator defect detection; YOLOv5s; asymmetric convolution; attention mechanism; PIoU; channel pruning

0 引言

随着国内输电网络的高速建设,绝缘子的保护需求和巡检队伍力量之间的矛盾越来越突出^[1]。近些年,无人机航拍巡检取代了人力巡检,一定程度上减轻了巡检的困难与压力,但是无人机航拍的绝缘子图像需要回传到后台人工判断和统计,存在检测效率低、实时性难以保障^[2],因此

需要研究一种部署于无人机等边缘设备的绝缘子缺陷检测模型。

近年来,国内外专家对绝缘子故障的诊断进行了大量研究,早期人们利用图像处理技术提取绝缘子的边缘和纹理特征,以这些特征信息为基础再进行目标识别任务。韩正新等^[3]先用自适应阈值分割方法将绝缘子与背景分开,

收稿日期:2022-09-14

* 基金项目:国家重点研发计划(2018YB1700200)、2020年辽宁省高等学校创新人才支持计划(2020-94)、2021年度高等学校基本科研项目重点项目(LJKZ0442)资助

再利用形态学滤波处理消除噪声干扰得到绝缘子轮廓,最后使用距离差值、面积统计等特征对绝缘子缺陷定位。商俊平等^[4]通过计算绝缘子欧氏距离对绝缘子的自爆位置检测。王梦^[5]通过基于连通区域、直线拟合和形态学的图像处理方式,完成了对绝缘子精确定位和自爆缺陷的测量。上述采用传统方式的缘子定位和缺陷检测的主要优势是在技术背景简单、场景单一的环境中检测精度较高,但对于复杂的场景,绝缘子缺陷检测精度低、实时性差。随着深度学习理论不断完善以及硬件设备算力的提升,基于深度学习的绝缘子缺陷检测方法是今后的主要发展方向。Liao等^[6]结合了Resnet101和Soft-NMS的Faster R-CNN模型实现了多种绝缘子缺陷的检测,但是检测精度低、检测时间慢。王凯等^[7]通过轻量化Ghost Module模块改进YOLOv3,并且使用空间金字塔模块丰富特征图来提升模型的精度,最终模型的检测速度提升达到35 FPS,检测精度比Liao等人略微提高。Qiu等^[8]使用轻量级MobileNet来替代YOLOv4的特征提取网络CSPDarkNet53,并降低路径聚合网络(path aggregation network, PANet)的宽度,该模型的检测精度和速度分别达到了93.81%和53 FPS。Huang等^[9]通过裁剪通道删除YOLOv5冗余的卷积层实现模型轻量化,其次在网络的相邻残差模块之间加入自适应注意模块来增强网络的特征学习能力,与原模型相比模型大小减少了69.4%,但是mAP下降了0.52%。上述部分方法虽然实现了模型的轻量化,但是检测精度都是降低的。由于YOLOv5s拥有检测精度

高、检测速度快、占用内存少的特点,本文将以YOLOv5s模型为主体改进一个轻量级的绝缘子缺陷检测模型。

针对上述绝缘子缺陷检测模型计算量和参数量较大,难于部署于边缘设备,且检测精度降低的问题,提出了轻量化的非对称卷积和注意力机制YOLOv5s模型(asymmetric convolution and attention mechanism YOLOv5s model, ACAM-YOLOv5s)。首先,借鉴了非对称卷积^[10]的思想提出了ACBlock-C3(asymmetric convolution block C3)用于提高模型特征提取能力,同时在主干网络中采用注意力机制CBAM(convolutional block attention module)^[11]加强对目标特征的关注;其次,采用PIoU(Pixels-IoU)^[12]作为定位的回归损失函数来实现绝缘子缺陷的精确定位;最后对ACAM-YOLOv5s模型进行基于网络BN(batch normalization)层进行通道剪枝^[13],得到轻量级的ACAM-YOLOv5s模型。最终实验表明本文提出的轻量级的ACAM-YOLOv5s绝缘子缺陷检测模型和原始YOLOv5s相比,在降低模型计算量,减小模型体积的同时,还提高了绝缘子缺陷的检测精度。

1 ACAM-YOLOv5s 绝缘子缺陷检测模型

1.1 ACAM-YOLOv5s 模型结构

YOLOv5s模型是YOLOv5中的轻量的模型,拥有较高的检测速度和检测精度,因此本文以YOLOv5s模型为基础来设计绝缘子检测的轻量化模型。如图1所示,实线部分是YOLOv5s模型,虚线部分是改进的部分。ACAM-

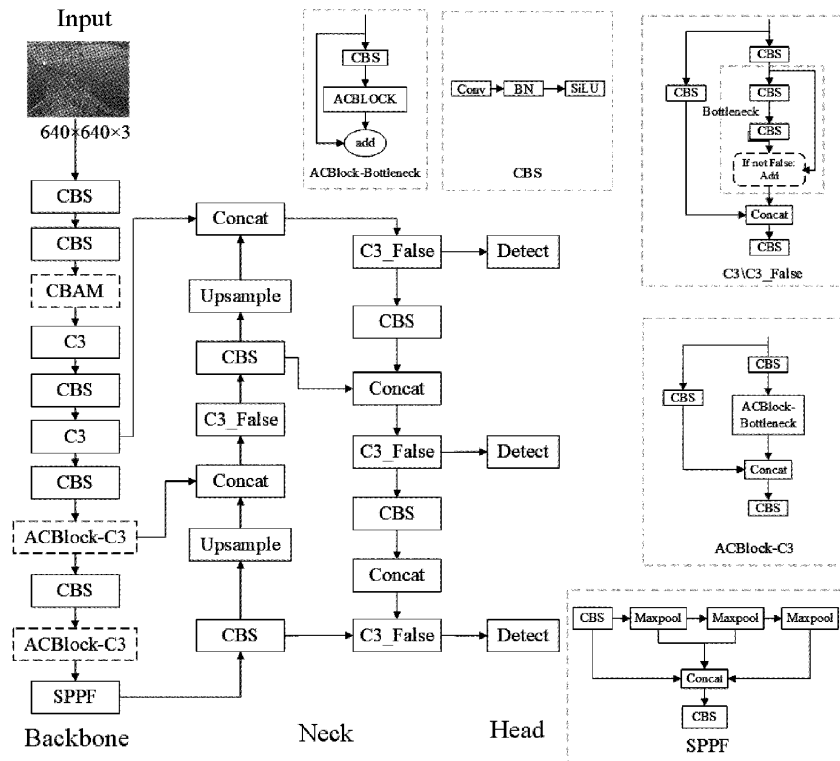


图1 改进的YOLOv5s模型结构

YOLOv5s 模型由 Input、Backbone、Neck 和 Head 构成。Input 是对输入图像进行预处理,主要涉及 Mosaic 图像增强、自适应图像缩放和自适应锚框计算。Backbone 主要包含了 CBS、CBAM、C3、ACBlock-C3 和 SPP (spatial pyramid pooling)^[14] 模块。其中 CBS 通过卷积步长实现特征图的降维,CBAM 是增强有效特征关注的注意力机制,C3 使用了残差结构来提取图像的深度语义信息,ACBlock-C3 是使用了非对称卷积的 C3,模型中的 SPPF 是更轻量的 SPP 实现。YOLOv5s 模型的 Neck 采用 PANet^[15] 结构,能够加强网络特征的融合和加快网络中推理信息的传输。Head 包含了检测头,损失函数和检测结果的优化,最终输出模型的检测结果。

1) ACBlock-C3 模块

YOLOv5s 的主干网络中大量使用 C3 模块,该模块主要依赖残差结构来提取特征。为了进一步加强主干网络的特征提取能力,把 C3 中残差单元的 3×3 卷积替换为非对称卷积 ACBlock,提出了 ACBlock-C3 模块,改进后的 C3 结构如图 2 所示。ACBlock 模块由 3 个卷积核并行组成,如图 3(a)所示,主要由 3×3、1×3 和 3×1 大小的卷积核构成。ACBlock 模块在利用方形卷积提取图像特征的基础上,还利用非对称卷积(水平卷积和垂直卷积)提取图像的横向和纵向特征,然后将得到的结果相加。与使用常规卷积的主干网络相比,改进后的主干网络可以获得更丰富的特征信息,还提升模型对图像反转和旋转的鲁棒性。

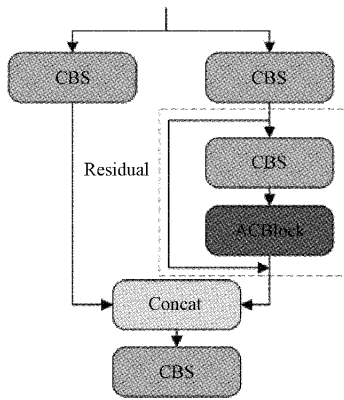


图 2 ACBlock-C3 结构

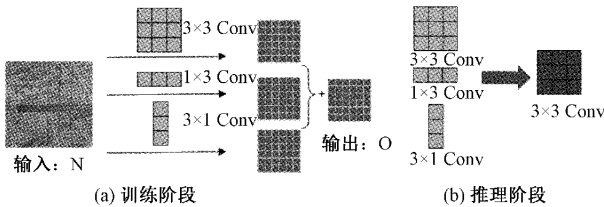


图 3 ACBlock 结构

在模型的训练阶段,ACBlock 模块中 3 个卷积核是同时且独立计算的,最终的输出是将这 3 个卷积层的计算结果进行叠加,然后继续进行前向传播。由于 ACBlock 模块

由 3 个卷积构成,所以模型参数量会增加,训练过程较长,但是并没有增加额外的参数。在训练完成后会把 ACBlock 的权重参数融合为标准卷积形式,用于模型的推理。融合的公式如下。

$$F^{(j)} = \frac{\theta_j}{\delta_j} F^{(j)} \oplus \frac{\bar{\theta}_j}{\bar{\delta}_j} F^{(j)} \oplus \frac{\hat{\theta}_j}{\hat{\delta}_j} F^{(j)} \quad (1)$$

$$b_j = -\frac{\mu_j \theta_j}{\delta_j} - \frac{\bar{\mu}_j \bar{\theta}_j}{\bar{\delta}_j} - \frac{\hat{\mu}_j \hat{\theta}_j}{\hat{\delta}_j} + \beta_j + \bar{\beta}_j + \hat{\beta}_j \quad (2)$$

$$O_{:,k} = \sum_{k=1}^C N_{:,k} \times F^{(j)} + b_j \quad (3)$$

式中: $F^{(j)}$ 、 b_j 是融合后的标准卷积的卷积核和偏置, N 是 ACBlock 模块的输入, O 是 ACBlock 模块的最终输出。 θ_j 、 $\bar{\theta}_j$ 和 $\hat{\theta}_j$ 是 3 个分支卷积核的权重, δ_j 、 $\bar{\delta}_j$ 和 $\hat{\delta}_j$ 是 3 个分支卷积核的方差, μ_j 、 $\bar{\mu}_j$ 和 $\hat{\mu}_j$ 是 3 个分支卷积核的批量标准化均值, β_j 、 $\bar{\beta}_j$ 和 $\hat{\beta}_j$ 是 3 个分支卷积核偏置。

在模型的推理阶段,模型初始化是使用融合后的 ACBlock 权重参数,由于融合后的 ACBlock 权重参数和相同的标准卷积层的参数结构一致,因此与 YOLOv5s 模型相比在推理阶段没有增加额外的计算,而是使用特征提取能力更强的融合后参数,最终的推理结果也会更加准确。

2) 融合注意力机制 CBAM

在 YOLOv5s 特征提取中,卷积层主要计算每个特征映射相邻位置的特征信息,而不考虑通道信息之间的相关映射^[16]。而 CBAM 注意力机制在训练学习的过程中对通道和空间的特征图的权重重新赋值^[17-18],给感兴趣的目标分配较大的权重,无兴趣的目标分配较小的权重,实现了特征信息的动态选择,进一步加强网络的特征提取能力。为了解决无人机航拍图像背景复杂、绝缘子缺陷目标小而导致检测精度低的问题,在使用了 ACBlock-C3 模型的主干网络中第一个 C3 模块前插入一个 CBAM 模块。CBAM 是通道和空间混合组成的注意力机制,其结构如图 4 所示。

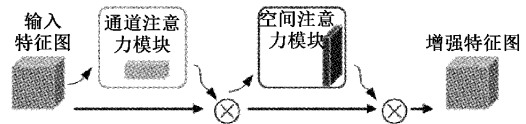


图 4 CBAM 注意力机制

通道注意力关注于特征图中有用的特征信息是什么,为了更加有效地计算通道注意力特征,采用平均池化与最大池化联合对特征的空间信息进行聚合,以降低特征的空间维数,然后由多层感知机为每个通道重新分配权重,产生最终的通道注意力特征图。空间注意力关注于特征图上有用特征信息的位置在哪里。首先,使用最大池化、平均池化将输入特征处理压缩为单通道,然后将特征图拼接。接着使用卷积对拼接后的特征图处理,产生最终的空间注意力特征图。通道注意力和空间注意力的计算公式

如下:

$$W_c(F1) = \sigma(\text{MLP}(\text{AvgPool}(F1)) + \text{MLP}(\text{maxPool}(F1))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (4)$$

$$W_s(F2) = \sigma(f^{7 \times 7}([\text{AvgPool}(F2); \text{MaxPool}(F2)])) = \sigma(f^{7 \times 7}(F_{avg}^s; F_{max}^s)) \quad (5)$$

式中: $F1$ 和 $F2$ 分别表示通道注意力和空间注意力的输入, σ 代表激活函数, MLP 代表一个共享的多层感知机, W_0 和 W_1 分别为感知机的两层参数, F_{avg}^c 和 F_{max}^c 分别表示全局平均池化和最大池化, $f^{7 \times 7}$ 表示 7×7 大小的卷积核操作。

1.2 损失函数的改进

损失函数是神经网络评估样本的重要基础, 由于损失函数的正确选择对模型收敛速度影响很大, 因此选择恰当的损失函数有助于得到更精准的定位。ACAM-YOLOv5s 中的损失函数 L 有三分: 分类损失 L_{cls} 、定位损失 L_{box} 、置信度损失 L_{obj} , 损失函数公式如下:

$$L = L_{cls} + L_{box} + L_{obj} \quad (6)$$

式中: 定位回归损失 L_{box} 是使用 CIoU Loss^[19] 函数, CIoU 是一个优秀的回归定位损失函数, 与 GIoU^[20], DIoU^[21] 相比不仅考虑了重叠面积和中心点距离, 还考虑了长宽比等因素的影响, 定位效果更准确。 L_{box} 计算公式如下:

$$L_{box} = 1 - CIoU(A, B) \quad (7)$$

式中: A 和 B 分别表示预测框和真实框, $CIoU$ 是计算的预测框和真实框的交并比。

但在实际绝缘子检测过程中发现, 对于具有高纵横比, 呈现细长倾斜的目标, CIoU 损失函数均会产生定位准确率较低的问题。为了解决这一问题, 本文将在损失函数中引入 PIoU, 该损失函数通过图像内部像素计数的方法来计算交并比, 而且该损失对边界框的大小、位置都是灵敏的。PIoU 损失函数定义为:

$$L_{PIoU} = - \left(\frac{1}{|M|} \right) \sum_{(b, b') \in M} \ln PIoU(b, b') \quad (8)$$

$$PIoU(b, b') = \frac{S_{b \cap b'}}{S_{b \cup b'}} \quad (9)$$

式中: M 为所有正样本的集合, $|M|$ 为正样本数, b 为真实目标, b' 为预测目标边框。 $PIoU(b, b')$ 是经二进制约束函数处理后真实目标 b 和预测目标框 b' 交集的像素点数和并集的像素点数比值。

1.3 基于通道剪枝的 ACAM-YOLOv5s 模型压缩

1) 稀疏化训练

ACAM-YOLOv5s 模型结构比较复杂, 其中存在一些冗余的结构, 模型的通道剪枝可以有效裁剪模型参数, 压缩模型体积, 降低计算量。通道剪枝常用的做法是为每个卷积层的通道设置一个评价因子, 通过该评价因子判断当前通道的重要性以此判定是否需要被优化。ACAM-YOLOv5s 模型中批量归一化(BN)^[22] 主要作用是使输出归一化, 防止过拟合, 提高训练效率, 在稀疏训练时选取

BN层的可训练参数 γ 作为通道评价因子, BN层具体操作如下:

$$Y = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (10)$$

式中: X 和 Y 是 BN 层的输入和输出, γ 可以评价通道的重要性, 作为通道剪枝的缩放因子, μ 和 σ^2 分别为均值和方差, ϵ 是一个极小的常值, β 是平移因子。BN 层中的缩放因子 γ 可以表示通道的重要性^[23], ACAM-YOLOv5s 模型训练后的 γ 呈正态分布, 值主要集中在 1 附近, 难以进行剪枝。需要将缩放因子 γ 添加到 ACAM-YOLOv5s 模型的损失函数中作为约束条件进行稀疏化训练, 并通过缩放因子施加 L1 正则约束, 使模型参数向结构化稀疏的方向调整, 帮我们找到重要的通道。其中模型新的损失函数如下:

$$L' = L + \lambda \sum_{r \in \Gamma} |\gamma| \quad (11)$$

式中: L 和 L' 分别表示 ACAM-YOLOv5s 模型常规训练的损失函数和稀疏训练的损失函数, $\lambda \sum_{r \in \Gamma} |\gamma|$ 为缩放因子的稀疏诱导惩罚, λ 为比例系数, $|\gamma|$ 为 L1 正则项, Γ 表示 BN 层缩放因子的集合。

2) 剪枝和微调

在使用 L1 约束的稀疏化训练后, 在得到的模型参数中包含了许多 BN 层的缩放因子, 其中缩放因子越接近 0 的通道越不重要, 即可剪去对应的通道, 最终剪去该通道对应的卷积层, 以此可以更好的降低模型的体积, 减少计算量。在模型剪枝过程中会根据设定的剪枝比例计算出模型 BN 层缩放因子对应的阈值 η , 其中缩放因子大于 η 的通道保存, 小于该阈值的通道则剪除。例如, 当 $\eta=0.4$ 时, 剪去 $\gamma < 0.4$ 通道后, 得到一个结构更紧凑的轻量模型, 如图 5 所示。一般情况下剪枝后的模型精度会下降, 需要进一步微调恢复模型精度。模型的剪枝率越高, 模型体积越小, 最终微调后的模型精度下降。因此, 需要根据实际应用选择合适的剪枝率来保证在检测精度和模型体积方面的平衡。

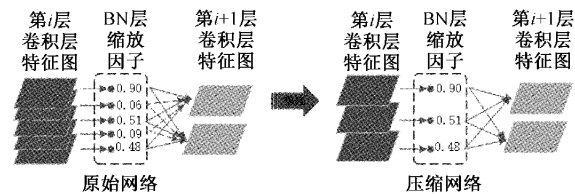


图 5 通道剪枝过程

2 实验结果与分析

2.1 实验平台与数据集

本实验是在 64 位 Windows10 操作系统计算机进行操作, 以 Pytorch 框架为模型的深度学习框架。实验环境具体配置如表 1 所示。

表 1 实验环境

实验环境	环境版本
操作系统	Windows 10 Professional Edition
CPU	Intel Core i9-9900K
GPU	NVIDIA GeForce RTX 2080Ti
内存	32G
加速框架	CUDA10.0
编程软件	Pycharm 2020.1.1
深度学习框架	Pytorch 1.7

本文的绝缘子数据来自 Tao 等^[24]公开的绝缘子数据集和第八届泰迪杯数据挖掘挑战赛电力巡检智能缺陷检测项目的数据。其中 Tao 的数据集共包括 848 张图像,包含 600 张正常绝缘子和 248 张爆裂的绝缘子。正常的绝缘子图像是在无人机巡检过程捕获的,缺陷的绝缘子图像则是通过图像分割的方法,从小部分原图分割出绝缘子,拼接在不同的背景下,形成缺陷的绝缘子图像。泰迪杯比赛的数据集是 40 张绝缘子数据集,其中的每一张图像中包含了多个正常的绝缘子和爆裂的缺陷绝缘子。因数据集中绝缘子数量少且类型单一,为了加强模型的鲁棒性,防止出现过拟合,通过对所有图像进行旋转、裁剪、平移、cutout、亮度、噪声的随机操作来丰富数据集。实验数据集总量为 3 700 张图像,按照 6 : 2 : 2 划分训练、验证和测试集,并将 LabelImg 标注后生成的 xml 标签转化为 txt 标签格式。

2.2 评价指标

为了验证所提绝缘子缺陷检测模型的可靠性,需要综合考虑检测精度,模型大小和检测速度,于是本实验选择平均精度均值(mAP),模型大小,BFLOPs(billion floating

point operations)和 FPS(frame per second)作为模型的评价指标。其中 BFLOPs 表示模型训练时需要的浮点运算量,代表着模型的复杂程度和占用内存。FPS 是每秒钟模型检测的图像帧数,是模型推理速度的重要指标。mAP 指标的计算公式如下:

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

$$AP = \int_0^1 P(R) dR \tag{14}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{15}$$

其中, P 和 R 分别表示精确度和召回率; TP 表示正样本的数量; FP 表示正样本被误认为负样本的数量; FN 表示负样本被误认为正样本的数量; $P(R)$ 表示不同召回率 R 下对应的精确率 P ; AP 表示平均精度; mAP 表示平均精度的均值; N 为类别总数。

2.3 ACAM-YOLOv5s 模型训练和测试

本文所研究的 YOLOv5s 目标检测算法是一种有监督的深度学习算法,需要大量的绝缘子数据集数据来训练模型。对原始的数据集扩充后达到了 3 700 张图像,其中 2 220 张训练图像、740 张验证图像和 740 张测试图像,并且输入图像分辨率重为 640×640 。为了在训练期间加快模型的收敛速度,模型训练使用预训练权重的方式和 SGD 优化器进行训练,并在训练之前设置部分超参数,epoch 为 200,学习率为 0.01,batch 尺寸为 8,热身训练为 3 轮。模型训练的 200 轮过程中损失和准确率变化,如图 6 所示。

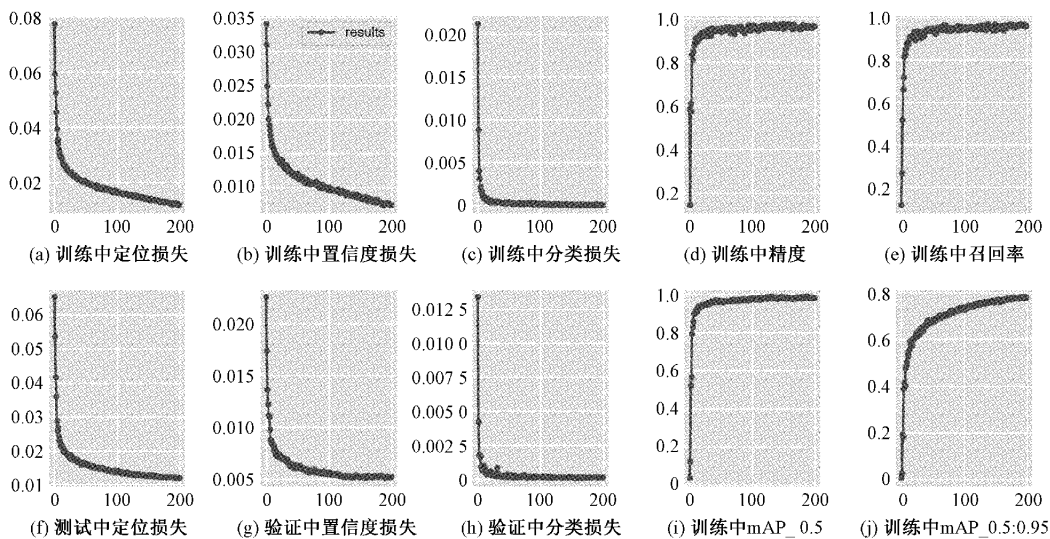


图 6 ACAM-YOLOv5s 训练中的损失和精度

基于在实验数据集上使用不同改进策略训练得到的改进模型,测试结果如表 2 所示。从原始网络 YOLOv5s

到使用不同改进策略的模型分别命名模型 A 到模型 D。从表 2 中可以发现,模型 B 使用了 ACBlock-C3 模块的骨

干网络,与原始 YOLOv5s 的模型 A 相比,mAP 提高了 0.5%,这是因为使用了 ACBlock-C3 模块的主干网络加强了对目标纵向和横向信息的提取,从而提取更多的目标特征。模型 C 在模型 B 的基础上使用了注意力机制 CBAM,mAP 比模型 B 相比提高了 0.2%,这是因为 CBAM 注意力机制增强了重要的特征,抑制了不重要的特征,使主干网络更关注于重要的特性信息。模型 D 在模型 C 的基础上把 YOLOv5s 的定位损失函数 CIoU 替换成了 PIoU,最

终实验的平均检测精度达到了 97.9%,证明了 PIoU 损失可以更好的解决细长物体和倾斜物体的回归损失,使检测精度更优。

综上所述,原始 YOLOv5s 算法的 mAP 为 96.8%,使用了 ACBlock-C3、CBAM 注意力机制和 PIoU 损失函数等方法的 ACAM-YOLOv5s 模型的 mAP 达到了 97.9%,而模型的体积和计算量只是略微增加,检测速度仍然达到了 116 fps,因此证明了本文改进方法的有效性。

表 2 YOLOv5s 改进过程实验对比

模型	ACBlock-C3	CBAM	PIoU	mAP/%	模型大小/M	模型权重/BFLOPs	检测速度/FPS
A				96.8	13.6	15.9	130
B	✓			97.3	14.9	17.1	120
C	✓	✓		97.5	15.0	17.1	116
D	✓	✓	✓	97.9	15.0	17.2	116

2.4 ACAM-YOLOv5s 模型剪枝实验结果

在稀疏训练过程中,根据式(10)的损失函数对模型约束,而比例因子 λ 影响模型的稀疏化程度。 λ 过大会带来损失函数的升高,导致模型性能急剧下降,而 λ 过小会导致模型稀疏化程度不够,不能进行有效的剪枝。本文选择了 $\lambda=0.0004$,图 7 是稀疏训练过程中损失函数和精度的变化曲线,由图可知在稀疏训练过程中损失下降速度较快,精度也得到了较大的恢复。

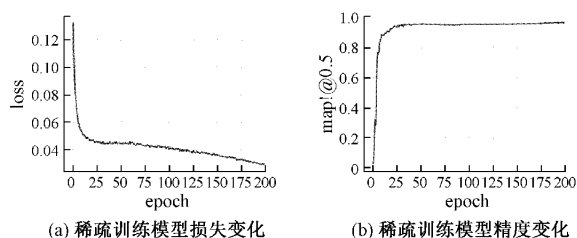
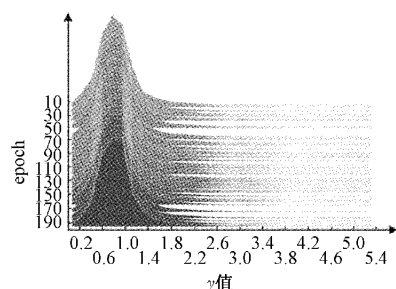


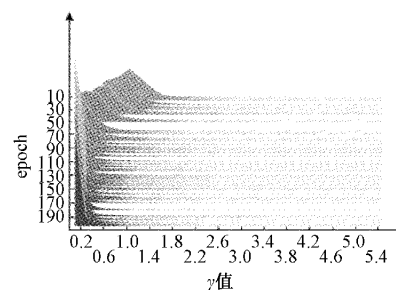
图 7 ACAM-YOLOv5s 稀疏训练中损失与精度

图 8 所示为 200 轮稀疏训练前后的 BN 层 γ 值分布,图 8(a)表示常规训练过程中的 γ 分布情况,每一轮的 γ 值分布是比较广泛的,但大部分值是集中在 1 附近,且随着训练轮数的增加, γ 值的分布没有明显变化,因此无法判断哪些通道的作用更大,无法进行剪枝。稀疏训练后的 γ 分布情况如图 8(b), γ 值稀疏的分布在 0~1,且随着训练轮数的增加, γ 值的分布的分布逐渐向 0 靠近,说明此时模型稀疏化程度较好,这些 γ 值对应的通道权重小,可以进行模型的剪枝。

在使用 λ 为 0.0004 进行稀疏化训练后,需要选择一



(a) 稀疏训练前BN层 γ 值分布



(b) 稀疏训练后BN层 γ 值分布

图 8 ACAM-YOLOv5s 稀疏训练前后 γ 值分布

定的剪枝率对模型进行剪枝。而在剪枝时为了保证网络结构的完整性,每一个 BN 层中的通道至少保留一个,也就是裁剪阈值不能超过每一层网络中 γ 参数最大值的最小值,因此模型的剪枝率也不能超过最大剪枝率。使用了 3 种剪枝率对模型剪枝,其中 70% 是最大剪枝率。如表 3 所示,剪枝之后的模型体积,计算量 BFLOPs 都下降了,经过最终微调后的 mAP 分别为 97.8%,97.6% 和 97.1%。

表 3 不同剪枝率效果对比

模型	剪枝后 map/%	微调后 map/%	模型大小/M	模型权重/BFLOPs	检测速度/FPS
ACAM-YOLOv5s(剪枝率 50%)	0.87	97.8	6.70	7.5	111
ACAM-YOLOv5s(剪枝率 60%)	0.76	97.6	5.72	6.4	112
ACAM-YOLOv5s(剪枝率 70%)	0.60	97.1	5.06	5.6	114

其中模型 1(Prune-0.5)的剪枝率最低,虽然微调后的 mAP 最高,但是参数量仍然比较多;模型 2(Prune-0.6)的经过微调后 mAP 下降的较少,模型体积进一步减少。模型 3(Prune-0.7)的体积最小,但检测精度也是最低的,表明此时剪枝率过高,丢失了重要的特征通道。综上所述,模型 2 的检测精度下降最少,同时模型的参数量和模型大小也有一定程度的减少,因此本文选择模型 2(Prune-0.6)作为最终的轻量化 ACAM-YOLOv5s 模型。

2.5 ACAM-YOLOv5s 模型性能对比分析

为了验证最终轻量化的 ACAM-YOLOv5s 模型的检测效果,使用相同的数据集训练,将本文模型的测试结果与 SSD、CenterNet、YOLOv3、YOLOv4、YOLOv5s、YOLOv5n 以及使用 EfficientNet-lite、GhostNet、MobileNetV3、PP-LCNet 和 ShuffleNet 轻量化骨干网络的 YOLOv5s 轻量化模型进行比较,具体结果如表 4 所示。

表 4 不同目标检测算法的检测效果

模型	骨干网络	map/%	检测速度/FPS	模型大小/M
SSD	VGG16	86.5	19	92.1
CenterNet	Resnet50	94.5	83	150.0
YOLOv3	Darknet53	86.4	35	240.0
YOLOv4	CSPDarknet53	94.6	45	244.5
YOLOv5s	CSPDarknet53	96.8	130	13.6
YOLOv5n	CSPDarknet53	96.4	130	3.6
EfficientNetLite-YOLOv5s	EfficientNetLite	95.6	112	7.5
GhostNet-YOLOv5s	GhostNet	96.1	78	10.6
MobileNetV3-YOLOv5s	MobileNetV3-small	94.5	111	7.0
PP-LCNet-YOLOv5s	PP-LCNet	94.3	141	6.6
ShuffleNetV2-YOLOv5s	ShuffleNetV2	94.7	136	6.3
ACAM-YOLOv5s(Light)	Improved CSPDarknet53	97.6	112	5.7

基于上述一系列实验对比与分析,与当下主流的一阶段目标检测模型相比,轻量化 ACAM-YOLOv5s 模型在检测精度上是最高的,mAP 达到了 97.6%,相比于 SSD、CenterNet、YOLOv3、YOLOv4、YOLOv5s、YOLOv5n 分别提高了 11.1%、3.1%、11.2%、3.0%、0.8%、1.3%。在模型大小方面,轻量化 ACAM-YOLOv5s 的模型体积只有 5.7 M,仅次于 YOLOv5n,但是 YOLOv5n 的检测精度较低,难以满足实际需求。在检测速度方面,轻量化 ACAM-YOLOv5s 达到了 112 fps,与原始 YOLOv5s 模型相比略微下降,但比其它一阶段模型都要快。与使用了主流轻量化网络改进的 YOLOv5s 模型相比,轻量化 ACAM-YOLOv5s 在模型大小与检测速度与上述轻量化模型相当的情况下,检测精度最高,分别提高了 2.0%、1.5%、3.1%、3.3%、2.9%。综上所述,在综合考虑模型的检测精度,模型大小和检测速度后,本文的轻量化 ACAM-YOLOv5s 模型与主流一阶段算

法以及使用轻量化主干网络改进的 YOLOv5s 系列算法相比,模型性能得到很大提升,本文模型凭借着高精度、轻量化的性能可以在实际航拍任务中进行绝缘子的缺陷检测。为了更直观的观察本文模型的检测效果,使用了原始 YOLOv5s 和轻量化 ACAM-YOLOv5s 模型其在测试数据集上的检测效果如图 9 所示。

图 9(a)~(c)依次是测试图像,YOLOv5s 和轻量化 ACAM-YOLOv5s 的检测结果。图 9(b)中 YOLOv5s 模型检测缺陷的 mAP 是 0.95 和 0.96,图 9(c)中轻量化 ACAM-YOLOv5s 模型检测缺陷的 mAP 是 0.98 和 0.97,检测精度都有提升。图 9(b)中 YOLOv5s 对绝缘子的检测出现了漏检,而图 9(c)中轻量化 ACAM-YOLOv5s 检测到了 YOLOv5s 模型未检测到的绝缘子。轻量化 ACAM-YOLOv5s 模型拥有更高的检测精度,减少了漏检绝缘子,因此可知本文的轻量化 ACAM-YOLOv5s 模型拥有更好的检测效果。

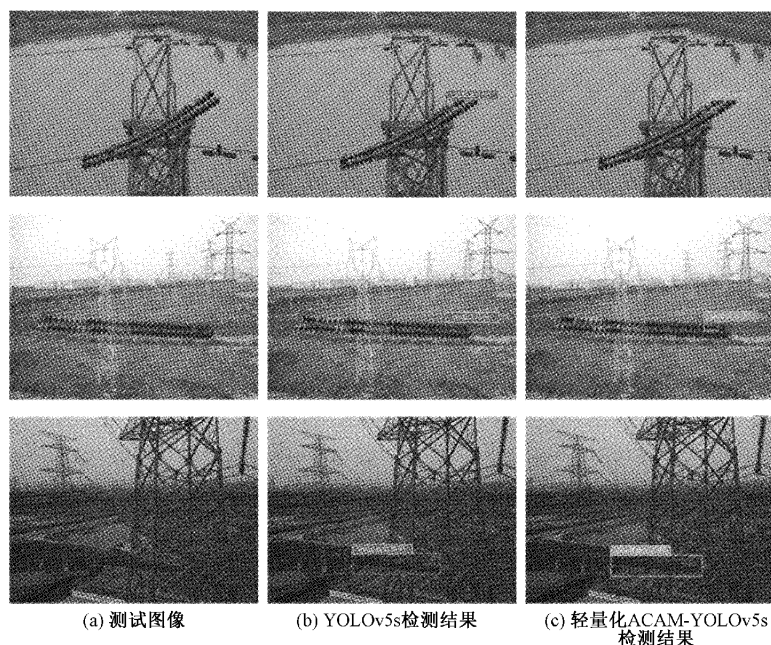


图 9 YOLOv5s 和轻量化 ACAM-YOLOv5s 算法的检测结果

3 结 论

针对绝缘子缺陷检测精度低及模型难于部署于边缘设备等问题,本文设计并实现了一种轻量化的绝缘子缺陷检测模型 ACAM-YOLOv5s。该模型使用了非对称卷积模块 ACBlock, CBAM 注意力机制和 PIoU 损失函数提高模型检测精度。接着对 ACAM-YOLOv5s 模型进行通道剪枝,在保持较高检测精度的情况下压缩模型体积,减少模型的参数,降低了计算量,使其有利于部署于边缘设备。在实验平台上的 mAP 达到了 97.6%,模型体积仅有 5.72 M,推理速度达到了 112 fps。虽然本文提出的轻量化 ACAM-YOLOv5s 模型拥有更高的检测精度和更小的模型体积,但是检测速度没有提升,因此下一步研究是进一步提升模型的检测速度。

参考文献

- [1] MIAO X, LIU X, CHEN J. Insulator detection in aerial images for transmission line inspection using single shot multibox detector[J]. IEEE Access, 2019, 7(18): 9945-9956.
- [2] LI X, SU H, LIU G. Insulator defect recognition based on global detection and local segmentation[J]. IEEE Access, 2020, 8(19): 59934-59946.
- [3] 韩正新, 乔耀华, 孙杨, 等. 基于图像识别的无人机输电线路绝缘子故障检测方法研究[J]. 现代电子技术, 2017, 40(22): 179-186.
- [4] 商俊平, 李储欣, 陈亮, 等. 基于视觉的绝缘子定位与自爆缺陷检测[J]. 电子测量与仪器学报, 2017, 31(6): 844-849.
- [5] 王梦. 基于绝缘子图像的缺陷检测方法研究[D]. 武汉: 华中科技大学, 2019.
- [6] LIAO G P, YANG G J, TONG W T. Study on power line insulator defect detection via improved faster region-based convolutional neural network[C]. 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), IEEE, 2019: 262-266.
- [7] 王凯, 徐敏, 孙翔, 等. 改进 YOLOv3 的绝缘子自爆缺陷检测方法[J/OL]. 小型微型计算机系统. (2021-11-03)[2022-07-06]. <http://kns.cnki.net/kcms/detail/21.1106.TP.20211102.1606.048.html>.
- [8] QIU Z, ZHU X, LIAO C, et al. Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model[J]. Applied Sciences, 2020, 12(3): 1207.
- [9] HUANG S Y, DONG X J, WANG Y. Detection of insulator burst position of lightweight YOLOv5[C]. ICCAI' 22: Proceedings of the 8th International Conference on Computing and Artificial Intelligence, 2022: 573-578.
- [10] DING X, PARK J, DING G. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1911-1920.
- [11] WOO S, LYU S, LEE J Y. CBAM: convolutional block attention module[J]. European Conference on Computer Vision, 2018, 11(6): 3-19.

- [12] CHEN Z, CHEN K, LIN W. Piou loss: Towards accurate oriented object detection in complex environments[C]. European Conference on Computer Vision. Springer, Cham, 2020: 195-211.
- [13] LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2736-2744.
- [14] HE K, ZHANG X, REN S. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1906.
- [15] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [16] 陈文豪,何敬,刘刚,等. 引入注意力机制的卷积神经网络高光谱图像分类[J/OL]. 激光与光电子学进展. (2021-08-03) [2022-07-06]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20210802.1730.070.html>.
- [17] JIE H, LI S, GANG S. Squeeze-and-excitation networks[J]. Transactions on Pattern Analysis and Machine Intelligence, 2020, 48(2): 2011-2023.
- [18] 任丰仪,裴信彪,乔正,等. 融合 CBAM 的 YOLOv4 轻量化检测方法[J/OL]. 小型微型计算机系统. (2022-03-02) [2022-07-06]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20210802.1730.070.html>.
- [19] ZHENG Z, WANG P, REN D. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 11(6): 3-19.
- [20] REZATOFIGHI H, TSOI N, GWAK J Y. Generalized intersection over union: A metric and a loss for bounding box regression[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [21] ZHENG Z, WANG P, LIU W. Distance-IoU loss: Faster and better learning for bounding box regression [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 12993-13000.
- [22] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. Proceedings of the 32nd International Conference on Machine Learning, 2015: 448-456.
- [23] HE Y, DONG X, KANG G, et al. Asymptotic soft filter pruning for deep convolutional neural networks[J]. IEEE Transactions on Cybernetics, 2019, 50(8): 3594-3604.
- [24] TAO X, ZHANG D, WANG Z, et al. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks[J]. IEEE Transactions on Systems, 2018, 50(4): 1486-1498.

作者简介

赵立杰, 博士, 教授, 主要从事智能信息处理方面的研究。
E-mail: zlj_lunlun@163.com

袁昌彪, 硕士研究生, 主要从事深度学习与目标检测方面的研究。
E-mail: 2072793140@qq.com

黄明忠(通信作者), 硕士, 副教授, 主要从事智能信息处理方面的研究。
E-mail: syhmz75111@163.com

王国刚, 博士, 教授, 主要从事图像信息处理方面的研究。
E-mail: 14950938@qq.com

张延华, 硕士, 副教授, 主要从事电力系统信息化方面的研究。
E-mail: 1315007672@qq.com