

DOI:10.19651/j.cnki.emt.2417530

基于改进 YOLOv8 的绝缘子缺陷检测方法<sup>\*</sup>杨 茜<sup>1</sup> 熊 炜<sup>1,2</sup> 孟圣哲<sup>1</sup> 黄玉谦<sup>1</sup>

(1. 湖北工业大学电气与电子工程学院 武汉 430068; 2. 南卡罗来纳大学计算机科学与工程系 南卡哥伦比亚 29201)

**摘 要:** 准确检测出绝缘子缺陷是电网维护的主要任务之一, 针对目前绝缘子缺陷检测算法识别精度不高和模型过大而难以部署至无人机等移动端设备的问题, 提出了以 YOLOv8 算法为基础网络进行改进以实现提高检测精度并将模型轻量化的方法。该方法采用多尺度融合网络 BiFPN 中的特征融合模式充分融合多尺度特征, 然后在原算法中融入可变形注意力机制 DAttention 以较低复杂度提取特征, 此外引入融合平均池化和最大池化的坐标注意力 DAF-CA 增强关键信息, 最后改用最小点距损失函数 MPDIoU 作为损失函数改善边界框回归的训练效果, 从而提高算法精度。在数据集上进行了多组对比实验, 结果表明该方法提出的方法均值平均精度可达约 91.0%, 模型浮点数和参数量分别为 7.2 G 和 2.07 M, 各项性能指标均优于目前常用的检测算法。该方法可为电网智能巡检提供参考。

**关键词:** YOLOv8; BiFPN; DAttention; DAF-CA; MPDIoU

**中图分类号:** TN606      **文献标识码:** A      **国家标准学科分类代码:** 520.6040

## Insulator defect detection method based on improved YOLOv8

Yang Qian<sup>1</sup> Xiong Wei<sup>1,2</sup> Meng Shengzhe<sup>1</sup> Huang Yuqian<sup>1</sup>

(1. College of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China;

2. Department of Computer Science and Engineering, University of South Carolina, Columbia 29201, America)

**Abstract:** Accurately detecting insulator defects is one of the main tasks of power grid maintenance. In response to the problems of low recognition accuracy of current insulator defect detection algorithms and large models that are difficult to deploy to mobile devices such as drones, a method based on YOLOv8 is proposed to improve the detection accuracy and lightweight the model. This method uses the feature fusion mode in a bi directional feature pyramid network BiFPN to fully fuse multi-scale features, and then integrates a deformable attention mechanism DAttention into the original algorithm to extract features with lower complexity. In addition, it introduces a fusion of average pooling and maximum pooling coordinate attention DAF-CA to enhance key information, and finally uses the minimum point distance based Intersection over Union MPDIoU as the loss function to improve the training effect of bounding box regression, thereby improving the accuracy of the algorithm. Multiple comparative experiments were conducted on the dataset, and the results showed that the proposed method achieved an average accuracy of about 91.0%. The model had a floating point count of 7.2 G and a parameter count of 2.07 M, respectively, and all performance indicators were superior to commonly used detection algorithms. This method can provide reference for intelligent inspection of power grids.

**Keywords:** YOLOv8; BiFPN; DAttention; DAF-CA; MPDIoU

## 0 引 言

电力是不可或缺的基础能源之一, 绝缘子作为高压输电线路电气绝缘和机械支撑的重要部分, 长期暴露在户外受到恶劣天气、机械负荷等因素破坏会产生缺陷, 其缺陷会

直接导致输电线路中断, 影响电力系统的稳定性以及个人用户的输配电。其中, 绝缘子缺陷主要分为两大类: 破损和闪络。为保证输电质量, 需定期巡检输电线路以排除故障。传统的维护大多使用人工巡检方式, 但其受制于环境和人为因素, 精确度和效率较低, 近年来随着技术发展, 逐渐使

收稿日期: 2024-12-03

<sup>\*</sup> 基金项目: 国家自然科学基金(61571182, 61601177)、国家留学基金(201808420418)、湖北省自然科学基金(2019CFB530)、湖北省科技厅重大专项(2019ZYYD020)项目资助

用无人机进行巡检,通过深度学习和计算机视觉技术对无人机拍摄到的图像和视频进行目标检测,识别出绝缘子缺陷,从而辅助巡检工作人员进行检测和维护。这种方法不仅提高了检测效率,还减少了人为因素对检测结果的影响,从而提高效率。

传统的图像处理方法适合在简单背景下做特征提取和阈值分割等,但面对复杂场景时,检测精度和鲁棒性都比较差,新型辅助巡检方式利用目标检测的方法,该方法检测效率较高且误检率较低。目标检测算法目的是在图像中找到所有感兴趣的物体,并为每个物体提供准确的边界框和类别标签。目前目标检测的主流算法分成两类,第 1 类分为两阶段:提取候选区域框特征,并对相应区域框进行分类预测,如区域卷积神经网络(region-based convolutional neural networks, Faster R-CNN)<sup>[1]</sup>,检测精度相对较高但其检测速度较为缓慢;第 2 类是基于端到端的一阶段回归方法,如实时物体检测算法(you only look once, YOLO)系列,检测速度相对较快但其检测精度相对不高。然而随着 YOLO 系列模型版本的进步,目前实时物体检测算法第 8 个版本(you only look once v8, YOLOv8)<sup>[2]</sup>在检测精度和速度方面都达到了较高水平。

文献[3]在 Faster R-CNN 算法的基础上进行改进,利用检测效果更好、性能更优的残差网络(residual network50, ResNet50)<sup>[4]</sup>代替原始视觉几何组网络(visual geometry group network, VGG)<sup>[5]</sup>进行缺陷识别,从而提高绝缘子检测精度。尽管检测精度有所提高,但是网络参数量与计算量增加,检测速度大幅下降,代价较大。

YOLO 系列网络近几年发展不断更新新版本,以 YOLOv5 为例,文献[6]对定向识别航拍绝缘子及其缺陷检测方法研究,在 YOLOv5 的主干特征提取网络引入轻量化注意力机制模块(convolutional block attention module, CBAM)<sup>[7]</sup>,对全局进行特征提取,使得复杂度较高;在特征融合阶段使用改进的空间金字塔池化结构(spatial pyramid pooling feature, SPPF)<sup>[8]</sup>,但是没有很好地利用显著信息从而漏掉了部分原始信息,损失了一定精度;使用完整损失函数(complete intersection over union, CIoU)<sup>[9]</sup>添加角度损失分类,更加关注边界框的形状导致其计算复杂度较大。

基于此,针对绝缘子缺陷检测这一应用场景存在的问题,未充分对有用特征进行采样提取,无法兼顾检测精度、参数量与计算量参数,为构建更轻量级且检测精度更高的网络模型,本文提出一种基于 BDDM-YOLOv8 特征融合的绝缘子缺陷检测方法,该方法在 YOLOv8 算法的基础上引入了多尺度融合网络(bi-directional feature pyramid network, BiFPN)<sup>[10]</sup>中的特征融合模式,应用自上而下和自下而上的多尺度特征并将其加权融合,选择性融合更多有用特征,提高精度并显著降低参数量,然后在骨干网络的第四个 C2F 模块后引入注意力机制(deformable attention, DAttention)<sup>[11]</sup>,关键采样点特征提取的方式使得复杂度较

低,此外引入注意力模块(distinctive and average features-coordinate attention, DAF-CA)<sup>[12]</sup>,参考平均信息和显著信息,从而增强关键信息以提取更准确的特征,最后改用最小点距损失函数(minimum point distance based intersection over union, MPDIoU)<sup>[13]</sup>作为新的损失度量,改善边界框回归的训练效果,提高网络的收敛速度和回归精度。本文的改进算法可以融合不同尺度特征信息进行信息交互以实现更高效的特征提取,进一步提高模型的检测精度,对有用特征进行采样提取有效降低参数量以及计算复杂度,能基本满足无人机远程部署要求,实现无人机巡检任务目标。

## 1 YOLOv8 原理及模型架构

YOLOv8 有 5 个模型, YOLOv8n 基准网络在保证较高检测精度的同时检测速度最快,参数量最低,因此本文选择 YOLOv8n 模型为基准,该模型主要包括输入层、主干网络、颈部网络和输出层。

YOLOv8 模型的输入层由 Mosaic 图像增强、自适应锚框计算和自适应图片缩放部分组成。Mosaic 图像增强将多个图像拼接在一起,从而增加训练数据的多样性。自适应锚框计算训练数据,找到适合目标检测的锚框大小和比例。自适应图片缩放是根据输入图像的尺寸自动调整网络的输入大小。

主干网络由卷积归一激活组合模块(conv batch normalization SiLU, CBS)、跨阶段局部网络(cross stage partial network, C2F)和空间金字塔池化融合网络(spatial pyramid pooling fusion, SPPF)等模块组成。CBS 模块是 YOLOv8 中的主要特征提取模块,批归一化卷积和 Sigmoid 门控线性单元(sigmoid linear unit, SiLU)结合可以有效地提取图像特征。C2F 模块用于进一步提取特征并增加网络的感受野。SPPF 模块是一种空间金字塔池化模块,可以在不同尺度上提取特征。

颈部网络采用特征金字塔网络(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN)。特征金字塔网络用于融合不同层次的特征信息,以便在不同尺度上进行目标检测。路径聚合网络用于进一步整合特征信息,提高检测的准确性和鲁棒性。

输出层使用解耦头结构,将分类和检测两个任务分离开。解耦头结构可以提高模型的灵活性和可扩展性,同时减少了计算量。通过输出层,可以得到每个目标的位置和类别信息。YOLOv8n 网络的主要结构如图 1 所示。

## 2 改进后的 YOLOv8 算法

本文提出的绝缘子检测整体网络如图 2 所示,在 YOLOv8 算法的基础上引入了 BiFPN 网络中的特征融合模式,应用自上而下和自下而上的多尺度特征并将其融合,然后在骨干网络的第 4 个 C2F 模块后引入注意力机制 DAttention,在提取重要特征的同时减少计算量,此外引入

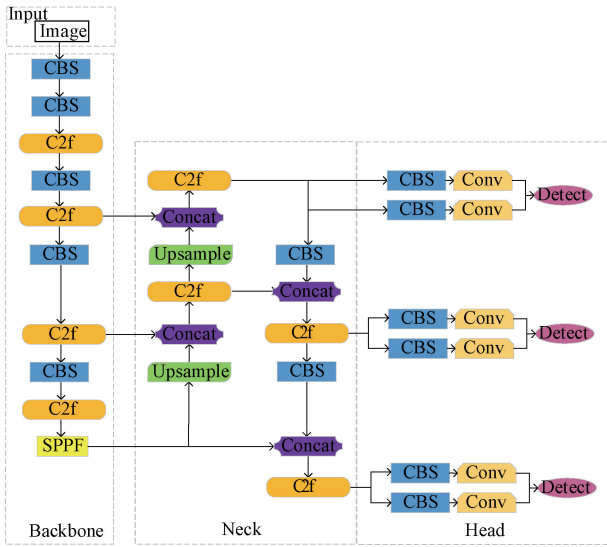


图 1 YOLOv8n 算法的主要结构

Fig. 1 The main structure of YOLOv8n algorithm

注意力模块 DAF-CA,参考平均信息和显著信息,从而增强关键信息以提取更准确的特征,最后改用 MPDIoU 作为新的损失度量,改善边界框回归的训练效果,提高网络的收敛速度和回归精度。激活函数仍使用 SiLU 并结合批归一化卷积有效地提取图像特征。

### 2.1 主干网络部分的改进

#### 1)可变形注意力机制(DAttention)

传统的注意力机制是对全局进行特征提取,使得复杂度较高,传统的注意力模块公式如式(1)所示,模块的每个 Query 和所有的 Key 计算注意力权重,所以收敛时间较长。

$$Att(z_q, x_k) = W \left[ \sum_{k \in \Omega_k} A_{qk} \times W'x_k \right] \quad (1)$$

式中:可变形注意力模块 DAttention (deformable attention)是一种用于计算机视觉任务的注意力机制,它在关注区域内对参考点周围的一小部分关键采样点进行特征提取,使得复杂度较低。可变形注意力模块如式(2)所示,

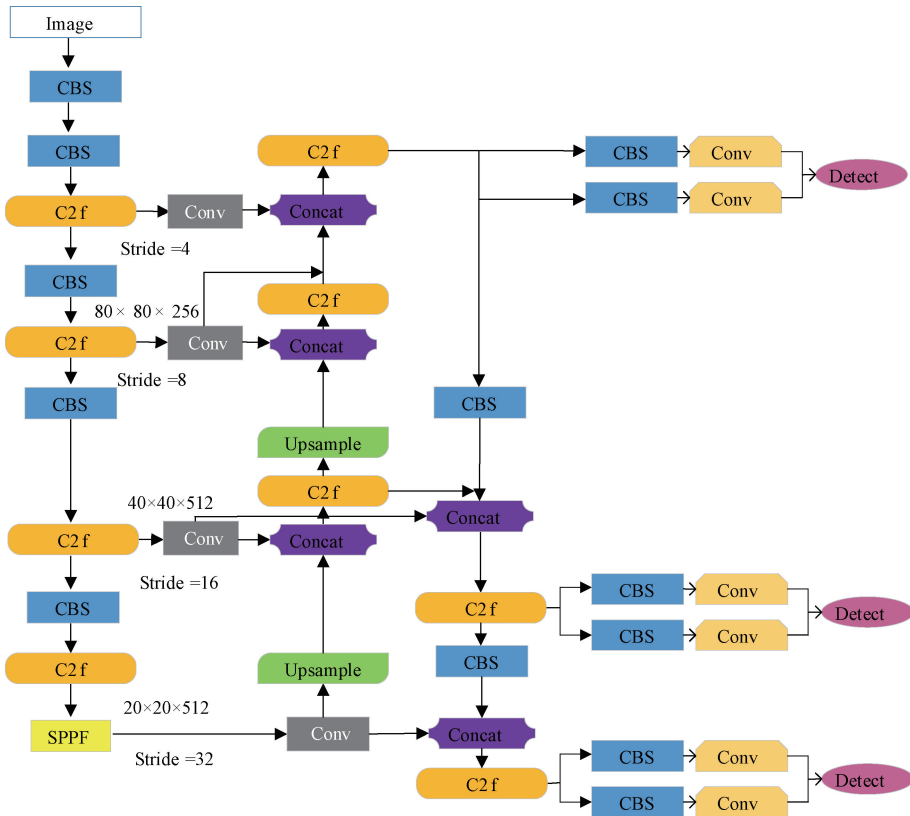


图 2 改进后的网络

Fig. 2 Improved network

每个 Query 未和所有位置的 Key 计算注意力权重,而是对于每个 Query,仅在部分位置中采样 Key,所以减少了收敛时间,并且 Value 也是基于这些位置进行采样插值得到的,最后将这个局部的注意力权重施加在相对应的 Value 上。

$$DAtt(z_q, p_q, x) = W \left[ \sum_{k=1}^K A_{qk} \times W'x(p_q + \Delta_{qk}) \right] \quad (2)$$

式中:W'是输入特征到 value 的转移矩阵,用于将  $x_k$  变换成 value;注意力施加在 value 后,进行线性变换从而得到不同头部的输出结果 W。  $A_{qk}$  是由 Query 和 Key 矩阵做

内积得到的归一化注意力权重。 $p_q$  代表参考点  $z_q$  的位置坐标。 $\Delta p_{qk}$  为第  $k$  个采样点与参考点的位置偏移,采样的  $key$  总数为  $k$ 。每个  $query$  只需和这采样的  $k$  个位置进行特征交互,  $(x(p_q + \Delta p_{qk}))$  为采样点插值出来的  $value$ ,不必像常规卷积注意力先从全局位置开始学习从而过渡到局部真正有意义的位置。位置偏移  $\Delta p_{qk}$  是可学习的,由  $Query$  特征  $z_q$  经过全连接层得到。标量注意力权重  $A_{qk}$  由  $Query$  特征  $z_q$  经过全连接层得到,同时在  $k$  个采样点之间归一化,且  $\sum_{k=1}^K A_{qk} = 1$ 。

传统注意力机制的计算要考虑所有的  $key$ ,而在可变形注意力模块中只需考虑一小部分的  $key$ ,每个  $Query$  只采样  $K$  次,从而减少了计算量。可变形注意力机制的实现过程如图 3 所示。可变形注意力模块分为两部分:偏移模块和注意力模块。首先将输入向量转换成特征图然后生成  $Query$  向量,同时考虑参考点的坐标,接着将输入特征图线性变换,与  $Query$  向量并行输入到偏移模块中,然后将偏移模块的产出与  $Query$  向量输入到注意力模块,从而得到最终结果。可变形注意力模块加入了采样,不管输入图像多大,只关注目标点附近的几个点,不需要全局位置学习从而有效降低了复杂度。

## 2) 融合平均池化和最大池化的坐标注意力 (DAF-CA)

本文受 CBAM 注意力的启发,发现坐标 CA 注意力机制<sup>[14]</sup> (coordinate attention for efficient mobile network design, CA) 并没有很好地利用显著信息,因此采用了一种融合平均池化和最大池化的坐标注意力 DAF-CA (distinctive and average features-coordinate attention)。DAF-CA 包含了 CA 未利用的显著信息,使得注意力权重分配更加合理以提高模型的性能。它不仅参考平均信息,而且参考显著信息,从而可以更准确地找到并增强关键信息。DAF-CA 注意力机制构建了保持原始信息深度可分离层,以不降级的方式保护了通道之间的丰富信息,为接下来的特征提取提供了更好的操作空间。深度可分离层通过提取更准确的特征,大大提高了模型的检测性能。

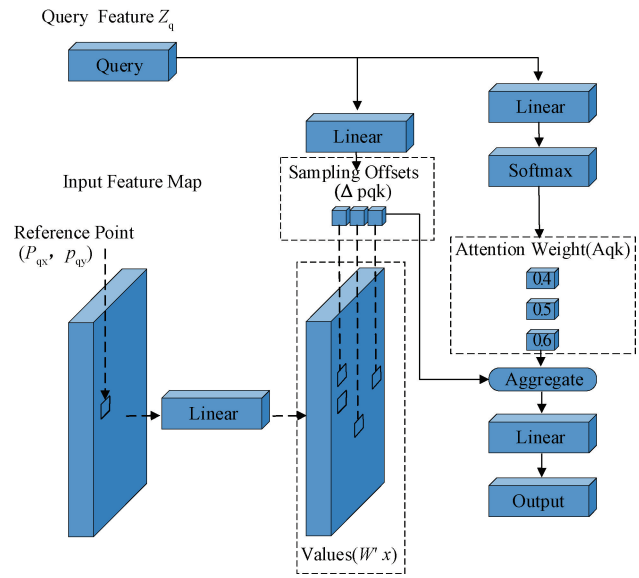


图 3 可变形注意力机制 Deformable Attention 的过程

Fig. 3 The process of Deformable Attention mechanism

DAF-CA 注意力机制如图 4 所示。其上半部分空间维度不变,压缩通道维度,关注空间维度,主要用于获取目标的位置信息。输入分别通过最大池化和平均池化分别得到两个  $C \times H \times 1$  的特征图,然后经过拼接操作对两个特征图进行堆叠。下半部分通道维度不变,压缩空间维度主要关注输入图片中重要的信息。输入分别经过两个并行的最大池化层和平均池化层,将特征图从  $C \times H \times W$  变为  $C \times 1 \times W$  的大小,将这两个特征图通过拼接操作进行合并,形成一个更完整的特征表示。然后通过维度变换和维度叠加将上下部分所得拼接生成新特征层  $X'$ 。为了降低计算复杂度,通过衰减率  $r$ ,卷积核大小为  $1 \times 2$  或  $2 \times 1$ ,填充为 0 的卷积运算提取新的特征层  $X''$ 。接着使用 split 函数将其划分为高度和宽度特征层,通过  $1 \times 1$  卷积将其通道维度恢复到  $C$  维。最后通过  $\delta$  函数计算每个特征的权值,然后将权值乘以原始特征层,最后得到特征层输出。

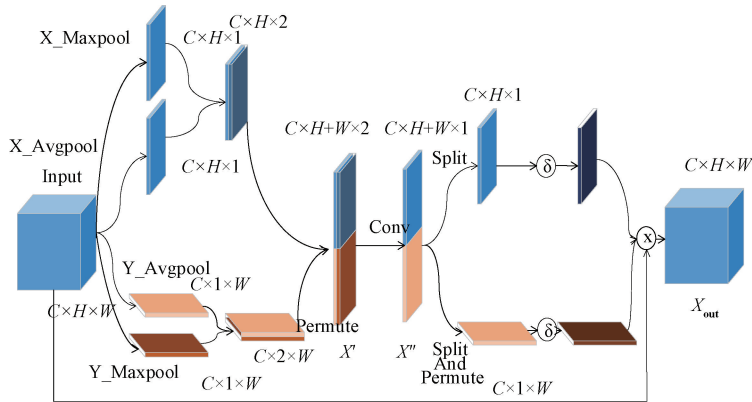


图 4 DAF-CA 机制

Fig. 4 DAF-CA mechanism



$$\begin{cases} X'_h = c(P_{ha}(X), P_{hm}(X)) \\ X'_w = c(P_{wa}(X), P_{wm}(X)) \\ X' = c(X'_h, X'_w) \\ X'' = F(X') \\ X_{out} = X \times \delta(f(f_1(X')_{split})) \end{cases} \quad (3)$$

式中:  $P_{ha}$  和  $P_{hm}$  表示  $h$  维的平均池化和最大池化,  $P_{wa}$  和  $P_{wm}$  表示  $w$  维的平均池化和最大池化。  $c$  函数表示特征层的叠加,  $f_1$  表示衰减率为  $r$  的卷积操作,  $F$  表示卷积核大小为  $1 \times 2$  或  $2 \times 1$ , 填充为 0。  $split$  函数将特征层在高度或宽度维度上分成两部分。  $f$  表示使用卷积核大小为  $1 \times 1$  的卷积层和原始维度的输出通道进行卷积操作。  $\delta$  表示特征的归一化。  $\times$  表示逐点乘法。  $X_{out}$  是原始特征层  $X$  乘以权值后的增强特征层。 DAF-CA 由于包含了原始 CA 未利用的显著信息, 使得注意力权重分配更加合理, 在参数量与原始 CA 几乎相同的情况下, 检测精度得到了显著提高。

## 2.2 颈部网络加入改进的多尺度融合网络 (BiFPN)

在传统的特征融合方法中, 通常会将不同尺度的特征图以相同的权重进行加权, 这种方法在处理分辨率相同的特征图时效果较好。 但是当输入的特征图分辨率不同时, 以相同的权重进行加权可能会导致输出的特征图对不同尺度的特征图的高度重视程度不平等。 BiFPN (bi-directional feature pyramid network) 引入了一种新的特征融合方法, 它根据不同输入特征的重要性设置不同的权重。 这种方法使得模型在进行特征融合时, 更好地考虑到不同尺度特征图的重要性, 引入可学习的权重来学习不同的输入特征, 并重复应用自上而下和自下而上的多尺度融合, 从而提高模型的性能。 颈部网络借鉴了 BiFPN 网络, 并对其进行了改进。

BiFPN 是一种多尺度融合网络, 它将双向跨尺度连接和加权特征进行融合, BiFPN 网络如图 5 所示, 首先在自上而下的输入边的前提下增加了一个额外的自下而上的路径聚合网络, 形成双向网络从而对来自所有层次的特征进行融合解除了单向信息流限制; 删除只有一个输入边的节点, 这些节点只有一条输入边, 对网络的特征融合贡献较小, 删除可以简化双向网络从而提高效率; 在输入节点到输出节点中增加一条额外的边, 在不增加过多成本的前提下融合更多特征。

BiFPN 结构中的加权融合方式采用快速归一化融合, 如式(4)所示, 该融合方式解决了训练速度慢的问题。 该方式将权重放缩至  $0 \sim 1$ , 并在每个  $W_i$  后加入 Relu 函数从而保证  $W_i \geq 0$ , 使  $\epsilon = 0.0001$ , 将  $\epsilon$  设置为较小的值从而避免了数值不稳定的问题, 每个归一化权重的值保持在  $0 \sim 1$ , 因未使用 Softmax 方式, 所以训练速度很快。

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \times I_i \quad (4)$$

式中:  $I_i$  为输入,  $O$  为输出,  $W_i$  为  $I_i$  对应的可学习的权重,  $\sum_i W_j$  为所有权重值之和。

BiFPN 融合了双向跨尺度连接和快速归一化, 式(5)描述了第 6 级的两个融合特征。  $P_6^{id}$  是自上而下路径上第 6 级的中间特征,  $P_6^{out}$  是自下而上路径上第六级的输出特征。 Conv 是深度可分离卷积, 并且每个卷积后都添加了批量归一化和激活函数。 所有其他特征都以类似的方式构建。

$$\begin{cases} P_6^{id} = \text{Conv}(\frac{w_1 \times P_6^{in} + w_2 \times \text{Resize}(P_7^{in})}{w_1 + w_2 + \epsilon}) \\ P_6^{out} = \text{Conv}(\frac{w'_1 \times P_6^{in} + w'_2 \times P_6^{id} + w'_3 \times \text{Resize}(P_5^{out})}{w'_1 + w'_2 + w'_3 + \epsilon}) \end{cases} \quad (5)$$

式中:  $P_6^{in}$  是第 6 层输入,  $W_1$  是第 6 层输入对应的权重,  $P_7^{in}$  是第 7 层输入,  $W_2$  是第 7 层输入对应的权重,  $W'_1$ 、 $W'_2$  和  $W'_3$  为中间层对应的权重。

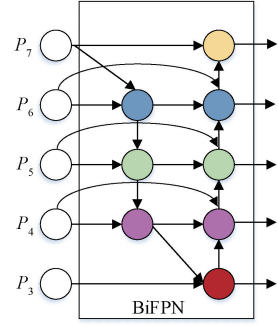


图 5 BiFPN 网络

Fig. 5 BiFPN network

颈部网络借鉴了 BiFPN 网络, 并对其连接方式进行改进, 在主干网络的第 1 个、第 2 个和第 3 个 C2f 网络以及 SPPF 网络连接卷积核  $1 \times 1$  卷积后使用 BiFPN 网络的连接模式, 实现双向跨尺度连接, 并在输入节点到输出节点增加一条额外的连接从而融合更多特征。 改进后的颈部网络如图 6 颈部框中所示, 该结构是加权且双向连接的, 即自顶向下和自底向上结构, 通过构造双向通道实现跨尺度连接, 将特征提取网络中的特征直接与自下而上路径中的相对大小特征融合, 保留了更浅的语义信息, 而不会丢失太多的深层语义信息。 通过添加一个跳跃连接和双向路径来实现加权融合和双向跨尺度连接。

## 2.3 损失函数的改进

最小点距损失函数 (MPDIoU)

边界框回归 (bounding box regression, BBR) 在目标检测中被广泛应用, 是定位目标的重要步骤。 大多数现有的边界框回归损失函数在预测框与实际标注框具有相同的宽高比, 但宽度和高度的值完全不同时, 无法进行有效优化。 原有 YOLOv8 算法使用 CIoU 损失函数更加关注边界框的形状导致其计算复杂度较大。 而 MPDIoU

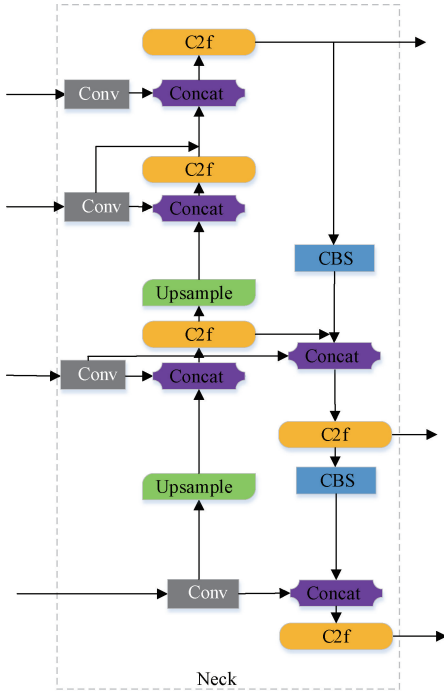


图 6 引入 BiFPN 网络改进后的网络

Fig. 6 Improved network by introducing BiFPN network

(minimum point distance based intersection over union) 基于最小点距离的新型边界框相似度比较度量可以简化两个边界框之间的相似性比较,改善边界框回归的训练效果,有效提高收敛速度和回归精度。该度量包含了现有损失函数中考虑的所有相关因素,即重叠或非重叠区域、中心点距离以及宽度和高度的偏差,同时简化计算过程。

为了提高边界框回归的准确性和效率,通常情况下,利用左上角和右下角点的坐标来定义一个唯一的矩形。基于交并比的度量标准,MPDIoU 最小化预测边界框与实际标注边界框之间的左上角和右下角点距离。

$$\begin{cases}
 B_{prd} = (x_1^{prd}, y_1^{prd}, x_2^{prd}, y_2^{prd}) \\
 B_{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt}) \\
 d_1^2 = (x_1^{prd} - x_1^{gt})^2 + (y_1^{prd} - y_1^{gt})^2 \\
 d_2^2 = (x_2^{prd} - x_2^{gt})^2 + (y_2^{prd} - y_2^{gt})^2 \\
 A^{gt} = (x_2^{gt} - x_1^{gt}) \times (y_2^{gt} - y_1^{gt}) \\
 A^{prd} = (x_2^{prd} - x_1^{prd}) \times (y_2^{prd} - y_1^{prd}) \\
 x_1^I = \max(x_1^{prd}, x_1^{gt}), x_2^I = \min(x_2^{prd}, x_2^{gt}) \\
 y_1^I = \max(y_1^{prd}, y_1^{gt}), y_2^I = \min(y_2^{prd}, y_2^{gt}) \\
 I = \begin{cases} (x_2^I - x_1^I) \times (y_2^I - y_1^I), & x_2 > x_1^I, y_2 > y_1^I \\ 0, & \text{其他} \end{cases} \\
 IoU = \frac{I}{u}, \text{ where } u = A^{gt} + A^{prd} - I \\
 MPDIoU = IoU - \frac{d_1^2}{h^2 + w^2} - \frac{d_2^2}{h^2 + w^2} \\
 L_{IoU} = 1 - IoU, L_{MPDIoU} = 1 - MPDIoU
 \end{cases} \quad (6)$$

式中:  $B_{prd}$  和  $B_{gt}$  分别为预测边界框和实际边界框左上角点和右下角点的坐标值。 $d_1$  和  $d_2$  分别为左上角点和右下角点的预测边界框和实际边界框之间的距离。式(6)中  $A_{prd}$  和  $A_{gt}$  分别为预测边界框和实际边界框的面积,  $I$  为预测边界框和实际边界框的交集面积,  $U$  为预测边界框和实际边界框的并集面积,  $L_{IoU}$  为 IoU 的损失函数表达式,  $L_{MPDIoU}$  为 MPDIoU 的损失函数表达式。与 IoU 不同的地方是增加了关于  $d_1$  和  $d_2$  的两项,如图 7 所示为预测边界框与实际标注边界框左上角和右下角点之间的距离。

通过使用 MPDIoU 作为新的损失度量,从而改善边界框回归的训练效果,提高收敛速度和回归精度。MPDIoU 简化了两个边界框之间的相似性比较,同时适用于重叠或非重叠的边界框回归。因此,MPDIoU 可以很好地替代交并比作为所有性能指标的度量,从而改善边界框回归的训练效果。



图 7 MPDIoU 计算图示

Fig. 7 MPDIoU calculation diagram

### 3 实验与结果分析

#### 3.1 数据集制作及其划分

本文的实验数据集来自 roboflow 由无人机采集的输电线路绝缘子缺陷图像,数据集一共包含 520 张,包含破损绝缘子(broken shell)、闪络绝缘子(flashover damage shell)和正常绝缘子(good shell) 3 类目标,通过图像增强调整图像亮度、对比度、饱和度、增加噪声等方式使其扩张至 1 580 张,随机将训练集、验证集、测试集的比例划分为 7:2:1,其中训练集 1 106 张,验证集 316 张,测试集 158 张,如图 8 所示用 LabelImg 对所有绝缘子图片进行标注。如图 9 所示,整体数据集共包含 8 281 个标签,其中破损绝缘子 1 005 个,闪络绝缘子 2 211 个,正常绝缘子 5 065 个,破损绝缘子、闪络绝缘子和正常绝缘子的比例约为 12:27:61。

为保证实验的合理性,使得目标检测通用评价指标全类平均精度均值的均值大小能够准确评估缺陷绝缘子的检测效果,在数据集制作中由于正常绝缘子(good shell)精度过高接近 1,会对实验训练结果产生重大影响导致均值

失衡,全类平均精度均值无法准确表示缺陷绝缘子的检测精度,并且实验只关注绝缘子缺陷部分,不需要检测出没有缺陷的正常绝缘子,因此采用删除正常绝缘子类别的技术方法,以满足模型训练需要。



图 8 LabellImg 标注图片

Fig. 8 LabellImg annotation image

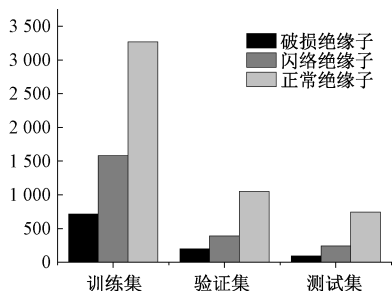


图 9 数据标签统计结果图

Fig. 9 Data label statistics results chart

本次实验中使用的 YOLOv8 模型是在 Ubuntu18.04.6 操作系统下配置,CPU 型号为 Intel i7-11700F@2.50 GHz,运行内存为 16 GB,GPU 型号为 NVIDIA GeForce RTX 3060,深度学习框架为 Pytorch2.1.2+cu121,运算架构为 Cuda11.1,开发环境是 Python3.8。训练过程中将输入图像尺寸处理为  $640 \times 640$ ,实验设置的批量大小(Batchsize)为 16,初始学习率设置为 0.01,动量设置为 0.937,训练周期设置为 300 轮<sup>[15]</sup>。

### 3.2 模型评价指标

文中研究采用  $mAP@.5$ 、 $mAP@.5:.95$ 、精确率(Precision)和召回率(Recall)对模型检测效果进行评估, $mAP@.5$ 指  $IoU=0.5$  时的全类平均精度均值, $mAP@.5:.95$ 表示在不同  $IoU$  阈值,从 0.5 到 0.95,步长 0.05(0.5、0.55、0.6、0.65、0.7、0.75、0.8、0.85、0.9、0.95)上的平均  $mAP$ 。

$$\begin{cases} Precision = \frac{TP}{TP + FP} \times 100\% \\ Recall = \frac{TP}{TP + FN} \times 100\% \end{cases} \quad (7)$$

式中:  $TP$  (true positive) 为模型正确检测到的目标数量;  $FP$  (false positive) 为模型错误地将非目标识别为目标的数量;  $FN$  (false negative) 为模型漏检的目标数量。

Recall-Precision 曲线以及  $AP$  (average precision) 值是评估目标检测模型性能常用的指标。Recall 通常作为横坐标, Precision 作为纵坐标,绘制出的曲线可以展示模型在

不同阈值下的性能表现。对于多类别检测任务,可以计算每个类别的  $AP$  值,然后对所有类别的  $AP$  值取平均得到  $mAP$  (mean average precision) 值,用于综合评估模型在所有类别上的检测效果,计算公式如式(8)所示。

$$\begin{cases} AP = \int_0^1 P dR \\ mAP = \frac{\sum_{i=0}^N AP_i}{N} \end{cases} \quad (8)$$

式中:  $N$  为数据集中检测目标的类别数;  $i$  是当前类别的编号。本次实验中,  $N=2$ , 即破损绝缘子和闪络绝缘子。

浮点数(Flops)和参数量(parameters)作为模型空间复杂度的评价指标即模型大小,如式(9)所示。

$$\begin{cases} Flops = h \times w \times k^2 \times c^2 \\ parameters = C_{in} \times (C_{out} \times k^2 + 1) \end{cases} \quad (9)$$

式中:  $h$  表示特征图高度,  $w$  表示特征图宽度,  $k$  表示卷积核大小,  $c$  表示特征图通道数,  $C_{in}$  表示输入通道数,  $C_{out}$  表示输出通道数。

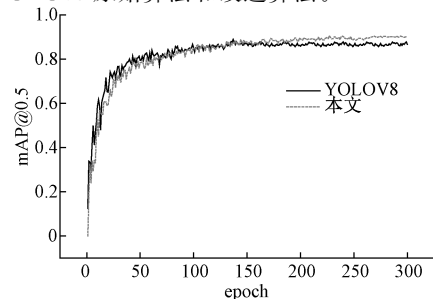
帧率(frames per second, FPS)指模型每秒能够处理的帧数,推理每张图片的平均时间是模型处理单张图片所花费的时间,计算公式如式(10)所示。

$$FPS = 1 / \text{平均推理时间} \quad (10)$$

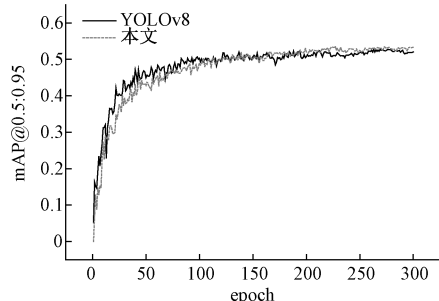
## 4 实验结果分析

### 4.1 本文算法训练结果

使用 YOLOv8 改进算法在数据集上进行训练和测试,训练过程的模型评价指标如图 10。线条颜色由深至浅分别代表 YOLOv8 原始算法和改进算法。



(a) 平均检测精度  $mAP@.5$   
(a) Average detection accuracy  $mAP@.5$



(b) 不同阈值的平均检测精度  $mAP@.5:.95$   
(b) Average detection accuracy at different thresholds  $mAP@.5:.95$

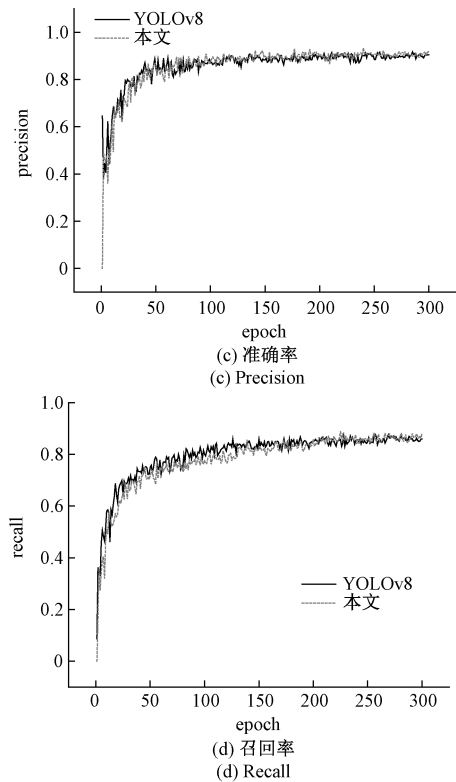


图 10 模型评价指标

Fig. 10 Model evaluation indicators

由图 10 可知,随着训练轮次的增加  $mAP@.5$  曲线逐渐上升,缺陷检测的  $mAP@.5$  达到 88.9%, $mAP@.5:@.95$  达到 52.0%,准确率和召回率分别达到了 88.6%和 82.7%。YOLOv8 改进算法缺陷检测的  $mAP@.5$  达到 91.0%, $mAP@.5:@.95$  达到 53.6%,准确率和召回率分别达到 90.3%和 87.8%。

4.2 不同注意力机制对比实验

为验证本文注意力机制的优越性,本文在改进网络的基础上同时使用注意力机制 DAttention 和 DAF-CA 与在同位置使用以下 4 种注意力模块进行比较,实验结果如表 1 所示。I 代表破损绝缘子,II 代表闪络绝缘子,分别为破损绝缘子和闪络绝缘子的检测精度。

本文将经过最后的注意力层处理的特征图进行可视化处理,其结果如图 11 所示,其中颜色越深代表权重越大,是检测任务所需关注目标的可能性越大。由于本文在样本空间分布不均,其中特征通道 SE 注意力模块(squeeze-and-excitation,SE)<sup>[16]</sup>只考虑了通道维度上的注意力,无法捕捉空间维度上的注意力,适用于通道数较多的场景,但对于通道数较少的情况可能不如其他注意力机制;高效通道 ECA 注意力模块(efficientchannel attention,ECA)<sup>[17]</sup>在处理全局上下文依赖性和通道空间关系方面存在一定的限制;全局机制 GAM 注意力模块(global attention with multi-head,GAM)存在计算复杂度、内存

表 1 注意力模块对比实验

Table 1 Comparison experiment of attention module

注意力模块	Parameter/M	GFlops/G	I/%	II/%	$mAP@0.5/%$	$mAP@.5:@.95/%$
SE	2.99	7.4	92.5	86.3	89.4	53.1
ECA	2.99	7.4	90.4	86.8	88.6	53.2
CBAM	2.12	7.4	92.9	86.5	89.7	53.4
GAM	2.86	7.8	93.6	87.0	90.3	53.0
CA	2.01	7.1	92.7	86.1	89.4	53.1
DAttention-DAF_CA	2.07	7.2	94.2	87.8	91.0	53.6

占用大的问题,对于大规模网络和高分辨率图像可能会带来性能瓶颈;轻量化 CBAM 注意力模块需要更多的计算资源,计算复杂度较高;坐标 CA 注意力模块由于需要对整个特征图进行注意力权重的计算,因此无法捕捉长距离的依赖关系。

本文使用的注意力机制 DAttention 只关注目标点附近的几个点,不需要全局位置学习从而降低了复杂度。注意力机制 DAF-CA 参考平均信息和显著信息,从而更准确地找到和增强关键信息,两个注意力机制结合使得网络有

效提高检测精度。

4.3 不同损失函数对比实验

为验证 MPDIoU 损失函数的优越性,本文在改进网络的基础上同时使用以下 3 种损失函数进行比较,实验结果如表 2 所示。I 代表破损绝缘子,II 代表闪络绝缘子,分别为破损绝缘子和闪络绝缘子的检测精度。300 轮次的  $mAP@.5$ 、 $mAP@.5:@.95$ 、精确率和召回率曲线如图 12 所示,线条颜色由深至浅分别代表 CIoU、GIoU、DIoU、MPDIoU。



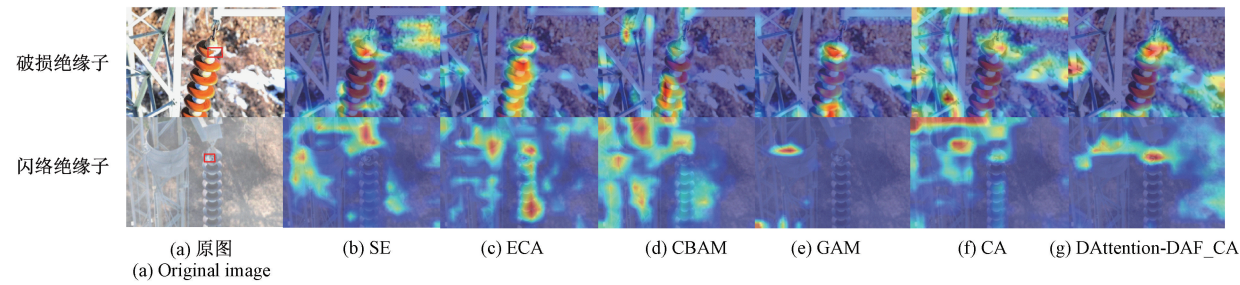


图 11 不同注意力机制特征可视化结果图

Fig. 11 Visualization results of different attention mechanism features

表 2 不同损失函数对比实验

Table 2 Comparative experiments on different loss functions

损失函数	Parameter/M	GFlops/G	I/%	II/%	mAP@0.5/%	mAP@.5:@.95/%
CIoU	2.07	7.2	93.5	86.1	89.8	53.2
GIoU	2.07	7.2	93.8	86.0	89.9	52.9
DIoU	2.07	7.2	92.8	85.6	89.2	53.1
MPDIoU	2.07	7.2	94.2	87.8	91.0	53.6

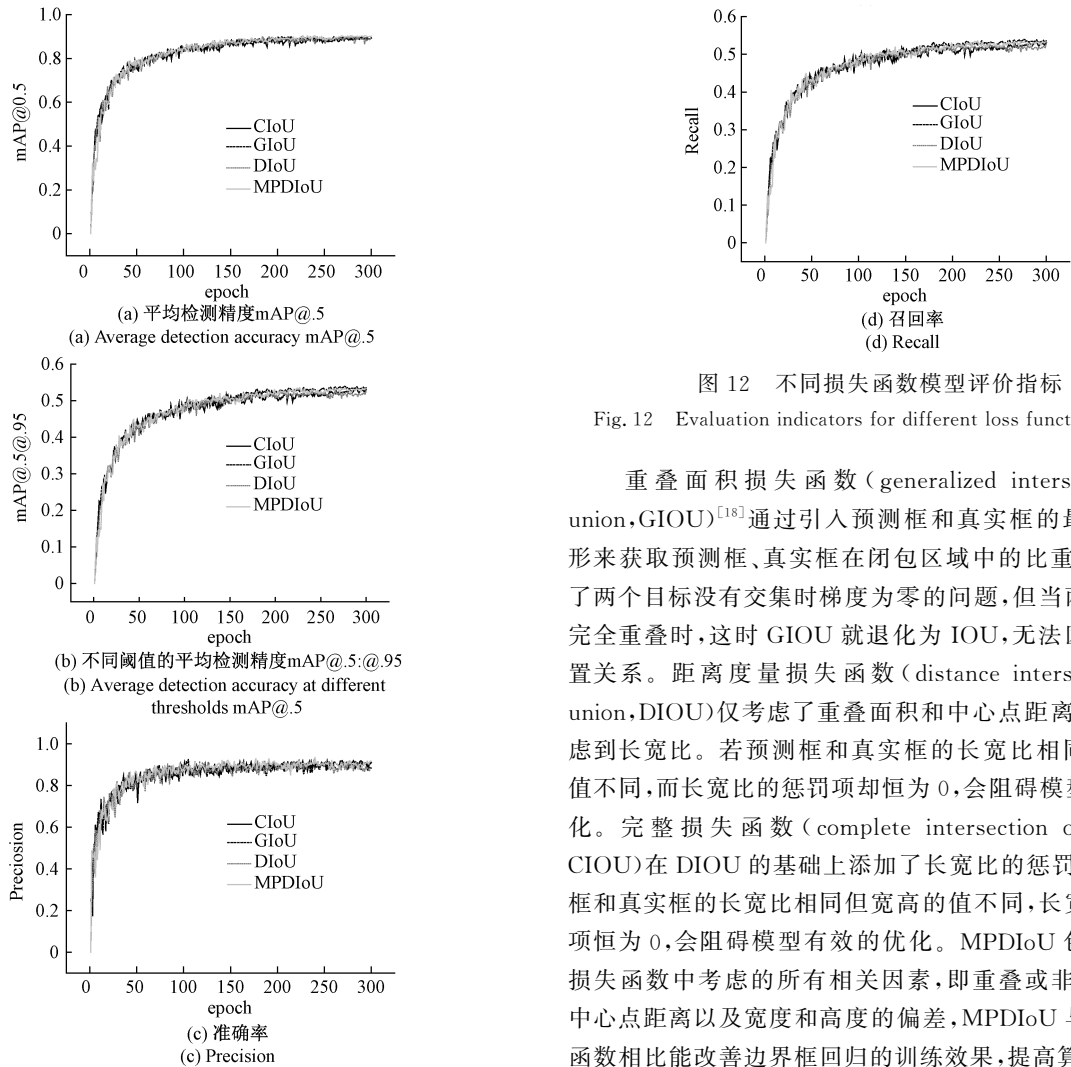


图 12 不同损失函数模型评价指标

Fig. 12 Evaluation indicators for different loss function models

重叠面积损失函数 (generalized intersection over union, GIoU)<sup>[18]</sup>通过引入预测框和真实框的最小外接矩形来获取预测框、真实框在闭包区域中的比重,从而解决了两个目标没有交集时梯度为零的问题,但当两个预测框完全重叠时,这时 GIoU 就退化为 IOU,无法区分相对位置关系。距离度量损失函数 (distance intersection over union, DIoU)仅考虑了重叠面积和中心点距离,并没有考虑到长宽比。若预测框和真实框的长宽比相同但宽高的值不同,而长宽比的惩罚项却恒为 0,会阻碍模型有效的优化。完整损失函数 (complete intersection over union, CIOU)在 DIoU 的基础上添加了长宽比的惩罚项,若预测框和真实框的长宽比相同但宽高的值不同,长宽比的惩罚项恒为 0,会阻碍模型有效的优化。MPDIoU 包含了现有损失函数中考虑的所有相关因素,即重叠或非重叠区域、中心点距离以及宽度和高度的偏差,MPDIoU 与其他损失函数相比能改善边界框回归的训练效果,提高算法精度。

4.4  消融实验

本文采用消融实验验证算法模型各种模块的有效性,分别从模型的准确率、召回率、 $mAP@.5$ 、 $mAP@.5:@.95$ 、计算量、参数量、精确率、召回率和  $FPS$  进行比较<sup>[19]</sup>。实验结果如表 3 所示,第一行表示原始的 YOLOv8 网络,a 代表在骨干网络中引入 BiFPN 网络的中的特征融合模块,b 代表在 C2F 模块中添加 DAttention 注意力机制,c 代表在 SPPF 模块后添加 DAF-CA 注意力模块,d 代表使用 MPDIoU 损失函数。I 代表破损绝缘子,II 代表闪络绝缘子,分别为破损绝缘子和闪络绝缘子的检测精度。

由消融实验表 3 可知,本文所提算法的破损绝缘子检测精度提高了 2.2%,闪络绝缘子检测精度提高了 2.0%,

$mAP@.5$  提高了 2.1%, $mAP@.5:@.95$  提高了 1.6%,参数量降低了 31.2%,浮点数减少了 11.1%,精确率提高了 1.9%,召回率提高了 6.2%, $FPS$  提高了 2.3%。综上得出本文提出的改进算法在绝缘子缺陷检测上的效果更好,以较低的计算资源实现了对绝缘子各类别缺陷的高精度识别,改进算法的各项指标均有所提升。

4.5  不同检测算法对比实验及分析

为验证本文算法的优势,将本文方法与目前其他目标检测算法进行比较,表 4 为不同算法的  $mAP@.5/(%)$ 、 $mAP@.5:@.95/(%)$ 、计算量、参数量、精确率、召回率和  $FPS$ ,其中 I 代表破损绝缘子,II 代表闪络绝缘子,分别为破损绝缘子和闪络绝缘子的检测精度。

表 3  YOLOv8-improve 模型消融实验  
Table 3  YOLOv8 improve model ablation experiment

a	b	c	d	I/%	II/%	$mAP@.5/%$	$mAP@.5:@.95/%$	Parameter/M	GFlops/G	Precision	Recall	FPS
—	—	—	—	92.0	85.8	88.9	52.0	3.01	8.1	0.886	0.827	227.3
✓	—	—	—	92.9	85.5	89.2	53.9	1.99	7.1	0.901	0.859	228.0
✓	✓	—	—	93.1	86.2	89.7	53.2	2.06	7.2	0.907	0.889	231.9
✓	✓	✓	—	94.0	85.6	89.8	53.2	2.07	7.2	0.902	0.864	231.9
✓	✓	✓	✓	94.2	87.8	91.0	53.6	2.07	7.2	0.903	0.878	232.5

表 4  不同目标检测算法对比实验  
Table 4  Comparative experiment of different object detection algorithms

算法	I/%	II/%	$mAP@0.5/%$	$mAP@.5:@.95/%$	Parameter/M	GFlops/G	Precision	Recall	FPS
Faster R-CNN	74.7	70.3	72.5	40.9	4.04	24.4	0.461	0.863	37.8
YOLOv5-s	89.5	82.1	85.8	44.3	7.46	17.4	0.856	0.794	178.4
YOLOv6-s	89.9	82.3	86.1	45.6	4.30	11.0	0.879	0.690	188.2
YOLOv7-tiny	91.0	84.8	87.9	49.0	6.21	13.1	0.878	0.802	190.5
YOLOv8-n	92.7	85.1	88.9	52.0	3.01	8.1	0.886	0.806	227.3
YOLOv9	93.5	86.5	90.0	52.5	2.77	7.9	0.891	0.850	235.4
YOLOv10	93.8	86.6	90.2	53.1	2.32	7.5	0.897	0.843	240.7
DETR	95.1	89.5	92.3	55.2	12.44	57.3	0.911	0.881	63.7
本文	94.2	87.8	91.0	53.6	2.07	7.2	0.903	0.878	232.5

从实验结果来看,本文算法的  $mAP@.5/(%)$ 、 $mAP@.5:@.95/(%)$ 、计算量、参数量、精确率、召回率,在检测精度和计算复杂度上综合性能优于现有的 Faster R-CNN、YOLOv5-s<sup>[20]</sup>、YOLOv6-s<sup>[21]</sup>、YOLOv7-tiny<sup>[22]</sup>、YOLOv8-n、YOLOv9<sup>[23]</sup>、YOLOv10<sup>[24]</sup>,以较低的计算资源实现了对绝缘子各类别缺陷的高精度识别,更适合电力巡检过程中对绝缘子缺陷的检测。

此外,本文研究了目前备受关注的基于 Transformer 的大型检测模型 DETR<sup>[25]</sup> 的检测结果,其检测精度均高于 YOLO 系列的算法模型,但其模型过大,不利于部署边缘端,因此本文未使用该类检测方法。实验表明,文中改进算法的检测精度有一定的提高,出现漏检的情况低,更加满足对实际检测的需要。

为了更直观地展示改进模型的检测效果,选取包含多尺度缺陷目标的典型场景图像进行推理实验,具体推理结果如图 13 所示,算法可以排除雨雾等噪声干扰,准确地识别出不同尺寸的破损、闪络缺陷,具有较强的泛化性和鲁棒性。

5  结    论

本文提出了一种基于 BDDM-YOLOv8 的绝缘子缺陷检测方法,该方法对数据集中闪络绝缘子及破损绝缘子的检测效果更优,可以解决绝缘子缺陷检测任务中背景噪声复杂、缺陷尺度跨度较大等问题。创新点包含以下 4 点:在原 YOLOv8 的主干网络部分,本文引入了可分离注意力模块以及 DAF-CA 注意力模块来增强骨干部分的特征提

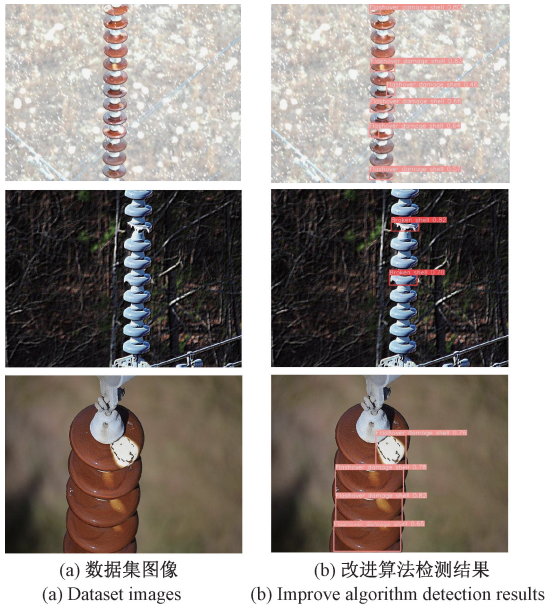


图 13 绝缘子缺陷图像推理实验结果

Fig. 13 Experimental results of insulator defect image inference

取能力,在颈部网络中融合了特征融合模块 BiFPN,它能够充分融合多尺度特征,除此之外,本文替换原始损失函数为 MPDIoU 损失函数,以进一步提升模型的检测精度。本文实验证明,本文所提出的模型有更高的检测精度和更低的参数量及浮点数。未来的研究任务将围绕进一步完善数据集,增强模型的泛化能力和鲁棒性,在保证检测精度的同时使模型进一步轻量化。

### 参考文献

- [1] REN SH 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, 2016, 39(6): 1137-1149.
- [2] TERVEN J, CORDOVA-ESPARZA D M, ROMERO G J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [3] 贺元帅, 纪超, 王博雅, 等. 基于改进 Faster R-CNN 的绝缘子缺陷检测识别与定位[J]. 电工技术, 2024(1): 176-179.  
HE Y SH, JI CH, WANG B Y, et al. Identification and localization of insulator defects based on improved Faster R-CNN[J]. Electrical Technology, 2024(1): 176-179.
- [4] THECKEDATH D, SEDAMKAR R R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks[J]. SN Computer Science, 2020, 1(2): 79.
- [5] SENGUPTA A, YE Y, WANG R, et al. Going deeper in spiking neural networks: VGG and residual architectures[J]. Frontiers in Neuroscience, 2019, 13: 95.
- [6] 赵博, 马宏忠, 张潇, 等. 定向识别航拍绝缘子及其缺陷检测方法研究[J]. 电子测量与仪器学报, 2023, 37(5): 240-251.  
ZHAO B, MA H ZH, ZHANG X, et al. Research on directional identification aerial photography of insulators and their defect detection methods[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(5): 240-251.
- [7] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision(ECCV), 2018: 3-19.
- [8] TANG H Y, LIANG SH, YAO D, et al. A visual defect detection for optics lens based on the YOLOv5-C3CA-SPPF network model[J]. Optics Express, 2023, 31(2): 2628-2643.
- [9] ZHENG ZH H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [10] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 10781-10790.
- [11] ZHU X ZH, SU W J, LU L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection[J]. ArXiv preprint arXiv:2010.04159, 2020.
- [12] SU P, HAN H Z, LIU M, et al. MOD-YOLO: Rethinking the YOLO architecture at the level of feature information and applying it to crack detection[J]. Expert Systems with Applications, 2024, 237: 121346.
- [13] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv:2307.07662, 2023.
- [14] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [15] 曾勇杰, 范必双, 杨涯文, 等. 改进 YOLOv8 算法在风机叶片缺陷检测上的应用[J]. 电子测量与仪器学报, 2024, 38(8): 26-35.  
ZENG Y J, FAN B SH, YANG Y W, et al. Application of improved YOLOv8 algorithm in fan blade defect detection[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 26-35.

- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [17] LIU T L, LUO R H, XU L Q, et al. Spatial channel attention for deep convolutional neural networks[J]. Mathematics, 2022, 10(10): 1750.
- [18] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [19] 廖丽瑛,刘洪. 基于改进 YOLOv8 的绝缘子自爆缺陷检测[J]. 电子测量技术, 2024, 47(18): 138-144.
- LIAO L Y, LIU H. Self explosion defect detection of insulators based on improved YOLOv8[J]. Electronic Measurement Technology, 2024, 47(18): 138-144.
- [20] JOCHER G, CHAURASIA A, STOKEN A, et al. Ultralytics/YOLOv5: v7.0-YOLOv5 sota realtime instance segmentation [J]. Zenodo, 2022, DOI: 10.5281/zenodo.3908559.
- [21] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. ArXiv preprint arXiv: 2209.02976, 2022.
- [22] 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[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [23] WANG C Y, YE H I, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Springer, Cham, 2025: 1-21.
- [24] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.
- [25] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.

## 作者简介

杨茜, 硕士研究生, 主要研究方向为目标检测。

E-mail: 940630258@qq.com

熊炜(通信作者), 副教授, 硕士生导师, 主要研究方向为计算机视觉与模式识别。

E-mail: xw@mail.hbut.edu.cn