

[杆塔缺陷检测]

基于YOLOv5的输电线路杆塔缺陷检测技术研究

文/赵柏然 张妍 马杰 曹成发 贺伟章
国网青海省电力公司西宁供电公司

针对输电线路航检过程中小目标在图像中所占像素比例低, 传统算法难以准确检测的问题, 研究提出一种基于YOLOv5的输电线路杆塔缺陷检测模型。首先, 模型增加一个用于小目标检测的输出层; 其次, 引入全维动态卷积模块和BiLevel路由注意力机制到网络中; 最后, 添加NWD损失函数, 与SiO_U结合作为定位回归损失函数。实验结果表明, 所提出的方法在检测杆塔缺陷方面达到了81.8%的平均准确率, 检测速度为26 FPS。该方法在不牺牲过多检测速度的情况下提高了输电线路杆塔缺陷的检测准确度。

输电线路是电网的重要环节，其安全运行是保障电网安全的基石。随着无人机技术的成熟和计算机图形识别技术的发展，通过深度学习和图像识别方法对输电线路进行巡检已成为重要的研究方向。然而，作为输电线路中尺寸较小的部分，杆塔在巡检图像中的特征表达能力较弱，增加了无人机识别和检测的难度。基于此，针对在复杂环境中难以准确检测输电线路杆塔缺陷的问题，本文介绍一种针对输电线路杆塔缺陷检测的改进算法，该算法基于YOLOv5框架进行了优化和创新。

YOLOv5目标检测算法

1. 本文中算法的总体结构

YOLOv5具有检测速度快、准确度高和鲁棒性强的优点,但在特征提取和特征融合方面仍有较大的改进空间^[2]。因此,本文在此基础上进行了改进。

本文提出的基于YOLOv5的输电线路杆塔缺陷检测算法的结构如图1所示。首先,将数据集中的图片批量缩放到 640×640 ,并发送到集成了ODConv的特征提取网络,然后发送到集成了Biformer注意力机制的颈部网络,将提取的高分辨率低语义和低分辨

率高语义特征图进行融合,得到 20×20 、 40×40 、 80×80 、 160×160 的特征图,再发送到头部进行预测,获得包含位置、分类和置信度参数的候选框。最后,使用NWD-SIoU抑制算法,获得置信度最高的预测框,从而实现输电线路杆塔的准确识别和检测。

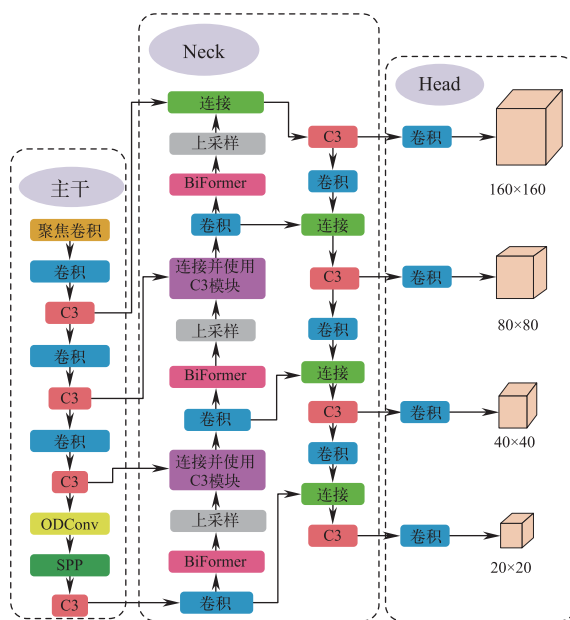


图1 改进的YOLOv5网络结构

2. 插入全维动态卷积

与传统目标检测相比,小尺寸目标检测面临着由于图像中像素占比少以及光线、背景等因素对细节特征影响较大而导致的特征信息提取困难的问题。为在保持参数数量和计算成本相对不变的前提下提升性能,本文创新性地将全维动态卷积(ODConv)技术融入YOLOv5的特征提取网络中^[3-4]。这一举措旨在通过优化卷积操作,增强网络对特征的动态捕捉能力,从而实现性能的显著提升。

ODConv是一种创新的动态卷积技术,它通过在核空间的四个维度上并行学习四种互补的注意力类型,从而有效利用多维注意力机制。这种设计使得ODConv能够在保持参数数量和计算成本相对不变的前提下,显著提升模型的性能。ODConv动态卷积分为

$$y = (\alpha_{ol} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{on} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x_i \quad (1)$$

式中, $x_i \in \mathbb{R}^{H \times W \times C}$ 表示输入特征向量, $\alpha_{oi} \in \mathbb{R}$ 表示卷积核 W_i 的注意力标量; $a_{si} \in \mathbb{R}^{k \times k}$, $a_{ci} \in \mathbb{R}^{c_{in}}$, $a_{fi} \in \mathbb{R}^{c_{out}}$ 表示新引入的三个维度上的注意力,这四种注意力是通过使用多注意力模块 $\pi i(x)$ 计算得到的;而 $\odot$ 表示沿着核空间不同维度的乘法操作。

多头注意力模块 $\pi i(x)$ 的架构如图2所示。在这个结构中,输入 x 首先经过全局平均池化(GAP)操作,被压缩成一个长度为 c_{in} 的特征向量。接下来,这个特征向量通过全连接层(FC)和修正线性单元(ReLU)进行进一步处理。在这个过程中,全连接层将压缩后的特征向量映射到一个较低维度的空间,这个空间的维数是原始空间的一定缩减率。

在四个分支头中,每个分支都有一个输出尺寸分别为 $k \times k$ 、 $c_{in} \times 1$ 、 $c_{out} \times 1$ 、 $n \times 1$ 的全连接层。这些全连接层用于生成归一化的注意力权重,分别表示为 a_{si} 、 a_{ci} 、 a_{fi} 和 a_{wi} 。这些注意力权重通过应用 Sigmoid 或 Softmax 函数来生成,以确保它们的值在 0~1 之间,并且所有注意力权重的总和为 1。

这种多头注意力机制允许模型在不同的子空间中并行地学习特征,从而提高模型对不同类型特征的捕捉能力。通过这种方式,模型可以更好地理解和处理复杂的输入数据,提高其在各种任务上的

性能。

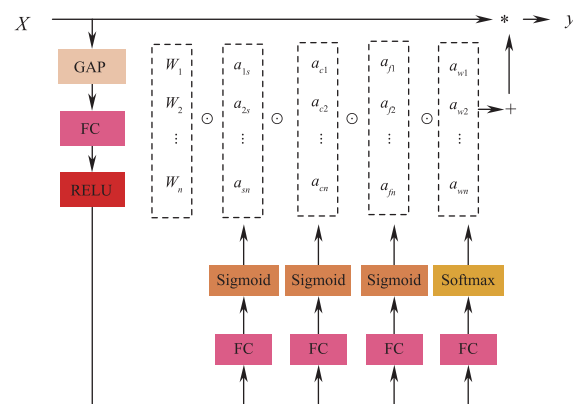


图2 多头注意力模块 $\pi i(x)$ 的示意图

在ODConv中,有四种注意力类型 a_{si} 、 a_{ci} 、 a_{fi} 、 a_{wi} 。这四种注意力类型彼此互补,共同增强模型的特征提取能力。具体来说,卷积核 W_i 在位置、通道、过滤器和核的维度上分别与不同的注意力机制相乘。这种设计使得每个输入的卷积操作在各个维度上都有所变化,从而能够捕捉到更加丰富和多样的上下文信息。这种改进特别有助于提升小目标检测的性能,因为小目标往往需要更多的上下文信息来准确识别。通过这种方式,模型在处理复杂场景时能够表现得更加出色,提高了整体的检测准确度和鲁棒性。

引入BiLevel路由注意力机制

在实际的YOLOv5特征融合网络中,融合后的特征会丢失一些重要细节,特别是对于小目标检测任务,这将导致整个网络的检测准确度下降。本文在改进后的特征融合层中引入了BiLevel路由注意力机制(Biformer)。该机制依据输入图像的特性,动态调整注意力权重,从而在不同位置或特征上赋予不同程度的关注,以引导网络更多地聚焦于与不同尺度传输线相关的小目标特征信息。此外,Biformer采用了参数共享策略,即多个位置或通道共享相同的权重。这种设计不仅减少了模型的参数量,使其更加轻量化,而且不会显著增加计算负担。Biformer的结构如图3所示。该结构由两部分组成;在第一部分中,图像数据通过重叠补丁嵌入转换为一维向量序列。随后在第二部分中,结构采用了一系列连续的补丁合并模块和Biformer块,以逐步降低输入的空间分辨率并增加通道数量。在每个

Biformer块中，模型首先通过 3×3 深度卷积隐式编码相对位置信息。接下来，依次应用扩展比率为 e 的BRA块和两层MLP块，以实现交叉位置关系建模和逐位置嵌入。

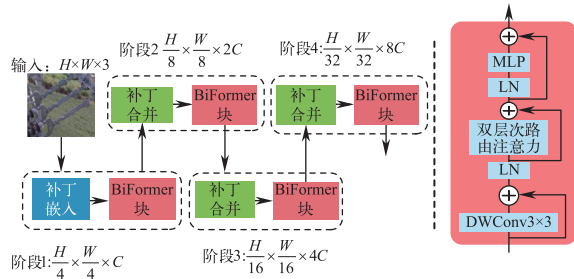


图3 Biformer结构图

理论上，区域间有向图路由使得特征图能够通过部分选择相关内容进行优化，这是Biformer注意力机制的核心，因此它不仅能够在总体上选择更精确的目标，还能在局部区域获得更准确的感受野内容。也就是说，基于双层路由的Biformer注意力机制具有更准确的选择感受野的能力。因此，本文认为将Biformer引入YOLOv5中是可行的，以优化小目标检测的效果。

改进损失函数

本文使用SIoU替代原始YOLOv5网络中的CIoU，其参数和示意图如图4所示。SIoU损失函数涵盖角度成本、距离成本以及形状成本。

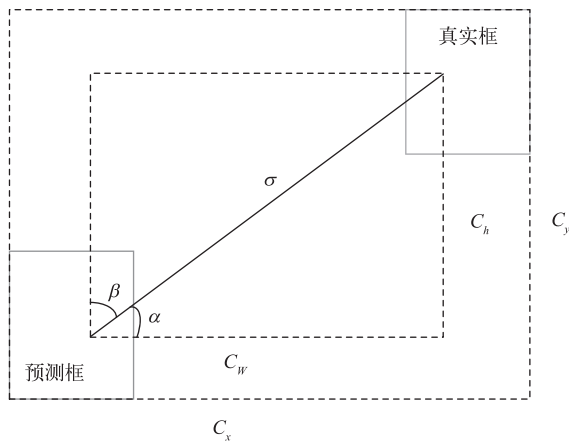


图4 SIoU示意图

(1) 角度成本

角度成本的计算公式为

$$\Delta = 1 - 2 \times \sin^2 \left(\arcsin(x) - \frac{\pi}{4} \right) \quad (2)$$

式中， x 为 a 的正弦值； a 可以通过两帧之间高度差与两帧中点之间的距离的比值获得。

(2) 距离成本

距离成本反映了预测框中心点与真实框中心点之间的距离。具体定义为

$$\Delta = \sum_{l=x,y} (1 - e^{-\gamma l}) \quad (3)$$

式中， $\gamma=2-\sqrt{e}$ 。

当 α 趋向于0时，距离成本的贡献显著减少。相反， α 越接近 $4\pi/4$ ，距离成本的贡献就越大。

(3) 形状成本

形状成本 ω 的定义为

$$\Omega = (1 - e^{-\omega_w})^\theta + (1 - e^{-\omega_h})^\theta \quad (4)$$

式中， θ 的值定义了形状成本的权重，本文中 θ 的值设定为1； ω_w 为两帧宽度差的绝对值与最大宽度值的比例； ω_h 为两帧高度差的绝对值与最大高度值的比例。

综上所述，SIoU的定义为

$$L_{\text{SIoU}} = 1 - I_{\text{IoU}} + \frac{\Delta + \Omega}{2} \quad (5)$$

与其他损失函数相比，SIoU引入了角度因素的考量，这使得预测框能够更迅速地逼近真实框，从而显著提升了损失函数的收敛速度。在计算真实框与预测框之间的宽高关系时，SIoU采取了独立计算的方式，有效避免了在预测框与真实框宽高比相同情况下惩罚项归零的问题。这使得损失函数的收敛更加稳定。

为了进一步提高模型对输电线路针孔缺陷的检测准确度，本文引入基于归一化瓦瑟斯坦距离的NWD损失函数，并提出结合SIoU和NWD的方法，如式(6)所示，其中 $L_{\text{NWD-SIoU}}$ 作为定位损失函数， α 表示SIoU的权重比例。如式(7)和式(8)所示，NWD采用二维高斯分布来评估预测框与标记框之间的相似程度， $W_2^2(N_a, N_b)$ 代表它们之间的归一化瓦瑟斯坦距离。

这种技术能够统一地表征模型所识别物体的空间分布间距。此外，NWD对于跨尺度的目标表现出非敏感性，因此在衡量小目标图像中预测框与标记框的相似度方面尤为适用。

$$L_{\text{NWD-SIoU}} = \alpha \text{SIoU} + (1 - \alpha) \text{NWD} \quad (6)$$

$$\text{NWD}(N_a, N_b) = \exp \left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C} \right) \quad (7)$$

$$W_2^2(N_a, N_b) = \left\| \begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right\|_2^2 \quad (8)$$

实验与结果分析

1. 数据集的建立

数据集是根据标准检查流程通过无人机(UAV)从输电线收集的。研究对象的螺栓销作为硬件的一个固定部分,最常出现在各部件的连接处。在早期阶段,手动选择了560个有效且清晰的可用数据,并按照8:2的比例分为训练集和测试集。电塔主体上的螺栓没有销钉,因此未标记,所以只标记了其他连接配件上的螺栓。实际上,由于螺母缺失和螺母松动的数据集非常少,本研究仅考虑两类:正常螺栓和缺销钉。

对上述连接配件进行标记,并使用labelme软件对所有图像进行标记,包括正常螺栓和缺销钉的螺栓。标记的正常螺栓和缺销钉螺栓的数量分别为1 076和288。由于数据集较小,通过对数据集中图像进行翻转、缩放、裁剪和颜色变换来扩展数据集。扩展后的数据集共有2 240个样本。扩展后的数据集见表1。

表1 数据集中故障的名称和数量

数据集	图像数量	正常螺栓/件	缺销钉的螺栓/件
训练集	1 792	3 010	862
测试集	448	1 294	290
总数据集	2 240	4 304	1 152

2. 实验环境和网络模型训练

实验中使用的操作系统是Windows 10, GPU是NVIDIA GeForce GTX 2080,深度学习框架是PyTorch 1.3.1。在模型训练过程中,由于数据量小且形式多样,使用Imagenet预训练的主干网络来预训练模型,并根据自建数据集对本文中的模型参数进行微调。初始学习率设置为0.01,优化策略采用随机梯度下降法,动量因子设置为0.937,训练周期为24个epoch,当epoch达到16时,学习率降至0.001。

3. 评估指标

本文使用所有类别的平均精度(AP)和平均平均准确度(mAP)作为准确性的评估指标。每秒帧数(FPS)被用作速度评估指标,代表每秒可以处理的图像数量。mAP代表所有类别的平均准确度均

值。其值越高,表明模型的性能越优越,目标识别的准确性也更高。AP作为PR曲线下面积的量化表示,其数值越大,意味着识别的精确度越高。计算公式为

$$mAP = \frac{1}{C} \sum_{i=1}^C AP(i) \quad (9)$$

$$AP = \int_0^1 P(r) dr \quad (10)$$

4. 消融实验

本文提出的基于YOLOv5的输电线销钉缺陷检测方法包括多尺度特征金字塔、ODConv、Biformer注意力机制和NWD-SIoU损失函数。为评估这些模块的有效性,本节进行了消融实验,以比较和分析它们的效果。实验结果见表2。

在表2中,逐步引入改进策略,从上至下进行实验。通过分析实验结果,可以观察到不同改进策略对检测性能的具体影响。在实验1中,向基线模型添加了一个小目标检测层。检测层的大部分特征来自主干网络的浅层特征。改进后的模型在测试集上的mAP%增加了1.2%,表明将包含大量小目标信息的浅层特征引入特征融合网络可以有效地向模型添加小目标的细粒度信息,从而大大提高了模型的检测性能。

表2 消融实验的结果

实验	小目标检测层	ODConv	Biformer	NWD-SIoU	标准AP (%)	损失	MA (%)	FPS
1	-	-	-	-	79.7	78.1	78.9	31
2	✓	-	-	-	80.7	79.5	80.1	30
3	✓	✓	-	-	81.4	79.8	80.6	29
4	✓	✓	✓	-	82.1	81.0	81.0	27
5	✓	✓	✓	✓	82.5	81.1	81.8	26

在实验3中,基于实验2,使用ODConv改进模型的主干部分,以更有效地细化销钉缺陷的细粒度特征。实验结果显示,模型的mAP%增加了0.5%,检测速度减少了1帧,表明ODConv可以提高小目标的检测准确度,并对检测速度影响不大。

在实验4中,基于实验3,将Biformer整合到YOLOv5的颈部层。带有Biformer的模型的mAP%值增加了0.7%,检测准确度略有提高,对模型的检测速度影响不大。

与实验4相比,实验5在实验4的基础上,将边界框定位损失的计算方法从CIoU改进为SIoU,并

融合了基于归一化Wasserstein距离的NWD位置回归损失函数。实验结果显示，NWD-SIoU损失函数可以提高模型的效果，mAP%有小幅增加，且对模型推理速度的影响较小。因此，这种策略几乎无损地提高了小目标的检测精度。

5. 对比实验

为验证所提出方法在输电线销钉缺陷检测中的优越性，使用输电线螺栓数据集进行了实验，并将该方法与现有的代表性目标检测算法进行了对比，包括单阶段检测网络（SSD，YOLOv4，Retinanet）和两阶段检测网络（如FasterRCNN，CascadeRCNN）。实验结果见表3。结果表明，与其他代表性目标检测算法相比，所提出的方法提高了输电线销钉缺陷的检测准确度，同时没有牺牲太多的检测速度，并且还能满足实时检测的需求。

表3 实验结果

算法	标准AP (%)	损失	mAP (%)	FPS	参数量
Fastercnn	80.1	78.7	79.4	24	41.1×106
Cascadecnn	79.8	78.5	79.15	22	67.1×106
SSD	76.4	74.2	75.3	31	24.8×106
YOLOv4	77.3	76.1	76.7	28	60.1×106
RetinaNet	77.6	76.4	77	30	36.2×106
本文的方法	82.5	81.1	81.8	26	53.7×106

6. 视觉分析

使用基线算法和优化算法检测销钉缺陷的视觉结果分别如图5和图6所示。与图5a和图5b相比，可以发现在光照条件较差的销钉情况下，基线算法存在漏检问题，而优化算法能够检测到销钉位置，并将其识别和分类为缺失销钉。

通过比较图6a和图6b，可以发现基线算法将正常销钉误识别为缺失销钉，存在明显的识别错误。优化算法成功检测到销钉位置，并消除了干扰目标。以上结果表明，优化算法提高了销钉缺陷的检测准确度。



(a) 基线算法 (b) 优化算法

图5 样本1的测试结果比较



(a) 基线算法 (b) 优化算法

图6 样本2的测试结果比较

结束语

本文提出一种基于YOLOv5的输电线销钉缺陷检测算法YOLOv5-S。首先，在YOLOv5网络的基础上增加了一个针对小目标的专用特征图。该特征图由较浅层的特征图融合而来，具有更准确的位置信息，使模型更容易检测到螺栓位置。之后，使用ODConv替换YOLOv5主干部分的局部卷积，以便网络更好地提取小目标的详细特征信息，而不会大幅增加参数。然后，将Biformer注意力机制引入颈部结构，以增强不同层次的特征融合，减少目标信息的丢失，提高网络检测准确度。最后，提出了一种新的损失函数形式，结合了NWD位置回归损失和SIoU损失。这种损失函数不仅避免了惩罚项为0的情况，还降低对小物体位置偏差的敏感性。EA

(收稿日期：2024-10-30)

