

# 基于多尺度特征融合与改进注意力的锈蚀螺栓螺帽检测

孙迪<sup>1,2</sup>, 郭义童<sup>2\*</sup>, 任超<sup>1</sup>, 范海峰<sup>1,2</sup>, 张传雷<sup>2</sup>

(1.天津大学电气自动化与信息工程学院,天津 300072; 2.天津科技大学人工智能学院,天津 300222)

**摘要:**针对输电线路巡检中螺栓螺帽目标尺寸小、数量多、且存在复杂背景遮挡的检测问题,提出具有多尺度特征融合机制和改进注意力机制的螺栓-螺母检测 Transformer (bolt-nut detection Transformer, BN-DETR) 算法。构建跨阶段部分连接网络 (cross stage partial darknet, CSPDarknet) 为主干网络的特征提取模块,通过集成局部感知、全局注意力机制和多层感知机实现多尺度特征的高效聚合,设计基于改进注意力的尺度内特征交互模块,通过动态选择关键点降低计算复杂度,同时保持全局信息交互能力,提出多层次注意力融合机制(全局、局部、像素级)提升特征表达能力。实验结果表明, BN-DETR 算法的平均精度均值@50 较基线算法提高 3%,为电力设施微小缺陷检测提供有效的技术参考。

**关键词:**螺栓螺帽;小目标检测;锈蚀;Transformer

**中图分类号:**TP391

**文献标志码:**A

**引用格式:**孙迪,郭义童,任超,等.基于多尺度特征融合与改进注意力的锈蚀螺栓螺帽检测[J].山东大学学报(理学版),2026,61(1):1-14.

## Based on multi-scale feature fusion and improved attention for rusty bolt and nut detection

SUN Di<sup>1,2</sup>, GUO Yitong<sup>2\*</sup>, REN Chao<sup>1</sup>, FAN Haifeng<sup>1,2</sup>, ZHANG Chuanlei<sup>2</sup>

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; 2. College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300222, China)

**Abstract:** To address the detection challenges associated with bolts and nuts in transmission line inspections—such as small object sizes, large quantities, and complex background occlusions, the method of bolt-nut detection Transformer (BN-DETR) incorporating a multi-scale feature fusion mechanism and enhanced attention mechanisms is proposed. A feature extraction module is constructed utilizing the cross-stage partial darknet (CSPDarknet) as the backbone, which efficiently aggregates multi-scale features through the integration of local perception, global attention mechanisms, and multi-layer perceptron. Scale-wise feature interaction module based on improved attention is designed to dynamically select key sampling points, thereby reducing computational complexity while preserving global information exchange. A multi-level attention fusion mechanism encompassing global, local, and pixel-level attention are introduced to augment feature representation. Experimental results demonstrate that, on a self-constructed dataset of corroded bolts in transmission lines, BN-DETR achieves 3% improvement in the @50 metric relative to the baseline algorithm. The proposed method offers an effective technical reference for the detection of small defects in power infrastructure.

**Key words:** bolt-nut; small object detection; rust; Transformer

## 0 引言

输电线路作为电力系统的重要组成部分,长期暴露在复杂的户外环境中,受到风雨、雷电、温差变化等自然因素的影响,关键部件常出现松动、腐蚀或损坏,若不能及时发现与修复,可能导致设备承载能力下降,甚至引发严重的结构失效。因此,如何高效、准确地检测输电线路部件的运行状态,成为电力系统安全运行中的核心问题之一<sup>[1]</sup>。

收稿日期:2025-07-13; 网络出版时间:2025-12-11

第一作者:孙迪(1986—),女,副教授,硕士研究生导师,博士,研究方向为计算机视觉、深度学习等。E-mail:dsun@tust.edu.cn

\*通信作者:郭义童(2000—),男,硕士研究生,研究方向为计算机视觉。E-mail:gyt23836967@mail.tust.edu.cn

随着图像获取和分析技术的发展,基于计算机视觉的检测手段逐渐成为电力巡检的重要研究方向<sup>[2-9]</sup>。采集图像进行自动化分析,可以显著降低人工依赖与检测成本。然而,输电线路部件种类繁多,其中螺栓螺帽作为数量最大、分布最广的紧固件,对整体结构安全性具有决定性作用。已有研究多聚焦于电塔<sup>[10-11]</sup>、接地线<sup>[12]</sup>、绝缘子<sup>[13]</sup>以及其他大型金具<sup>[14]</sup>的检测,而针对螺栓螺帽锈蚀的自动检测研究仍较少<sup>[15]</sup>。螺栓螺帽一旦发生严重锈蚀,极易导致断裂或失效,因此亟需有效的智能检测算法。

在目标检测研究中,基于卷积神经网络(convolutional neural network, CNN)的快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)<sup>[16]</sup>与 You Only Look Once (YOLO) 系列<sup>[17]</sup>是典型代表。Faster R-CNN 是经典的双阶段检测算法,它分为候选区域提议生成和候选区域分类及边界框回归 2 个阶段。在候选区域提议生成阶段,算法首先在输入图像中生成一系列候选区域,这些候选区域可能包含某些对象。这个步骤通常利用区域建议网络完成。在候选区域分类及边界框回归阶段,第 1 阶段生成的每个候选区域进行更精确的边界框回归和对象类别预测。而 YOLO 算法则属于单阶段检测算法,在单一前向传播中完成目标检测任务,无须生成候选区域。单阶段算法通过密集的网络或锚框直接预测目标的类别和位置。然而,CNN 算法应用于输电线路螺栓螺帽检测时仍面临诸多挑战:(1) CNN 算法倾向于聚焦局部区域的信息提取,易丢失全局细节;(2) 算法结构复杂,设计需依赖大量先验知识;(3) 螺栓螺帽目标尺寸较小,且易受遮挡和复杂背景干扰。

Transformer<sup>[18]</sup>在计算机视觉领域的应用为解决以上问题提供了新的算法。它的自注意力机制可以同时关注输入序列的所有部分,具有天然的全局感知能力。基于 Transformer 的目标检测(detection Transformer, DETR)<sup>[19]</sup>算法是由 Facebook 团队在 2020 年提出的基于 Transformer 的端到端目标检测算法。它既不需要非极大值抑制(non-maximum suppression, NMS)后处理步骤,也不需要锚框等先验知识和约束,完全依靠网络自身实现从输入到目标检测结果的直接转换,使得目标检测的流程更加简洁和直接。然而,DETR 较高的计算量和较慢的收敛速率阻碍了它的大规模应用。为了克服 DETR 自身的缺陷,有学者提出锚点 Transformer 目标检测(anchor-based detection Transformer, anchor DETR)<sup>[20]</sup>、条件 Transformer 目标检测(conditional detection Transformer, conditional DETR)<sup>[21]</sup>、可变形 Transformer 目标检测(deformable detection Transformer, deformable DETR)<sup>[22]</sup>、基于动态锚框的 Transformer 目标检测(dynamic anchor boxes detection Transformer, DAB-DETR)<sup>[23]</sup>的几种变体,这些算法在一定程度上缓解了目标检测中的部分问题,但实用性仍显不足。实时 Transformer 目标检测(real-time detection Transformer, RT-DETR)<sup>[24]</sup>能显著提升实用性,然而它并非针对螺栓螺帽这类小目标检测而设计。

通过对现有研究的综合分析发现,将 CNN 算法与 Transformer 架构进行有机整合并实施针对性优化,能够为输电线路螺栓螺帽检测任务提供较为理想的解决方案。该技术挑战主要源于以下几个关键因素:首先,输电设备所处的复杂背景环境导致螺栓螺帽的视觉特征与周围部件存在高度相似性,显著增加了目标特征的区分难度;其次,金属部件长期暴露在户外环境引发的锈蚀现象,不仅改变了目标表面的纹理特征,更造成目标与背景的灰度值趋同,导致特征混淆问题加剧;此外,不同型号设备所采用的紧固件存在显著尺寸差异,这对检测算法的多尺度适应能力提出了更高要求。基于此,本文通过构建 CNN-Transformer 混合架构的并行特征提取策略,在保持 CNN 算法局部特征捕捉优势的同时,引入 Transformer 的全局建模能力增强复杂场景下的特征表征,并通过改进解码结构的多尺度融合机制提升检测器对尺寸差异的适应能力。

本文针对以上问题进行研究。在复杂背景和图像中占比像素较小的场景下,检测螺栓螺帽的精度易受影响,特别在锈蚀等复杂纹理存在的情况下,单独的 CNN 算法或者 Transformer 算法均难以有效分离目标与背景,为此,本文提出一种结合 CNN 算法局部特征提取与 Transformer 全局感知能力的算法,通过强化特征的提取的和不同维度特征的融合,更好的区分目标和背景。同时,改进现有的注意力机制,使算法能够根据螺栓螺帽图像的内容自适应地调整处理粒度,进一步优化检测性能,因此,本文提出多尺度特征融合机制和改进注意力机制的螺栓-螺母检测 Transformer(bolt-nut detection Transformer, BN-DETR)算法。为了使本文算法在面对锈蚀纹理时能更准确地区分目标与背景,利用跨阶段部分连接网络(cross stage partial darknet, CSPDarknet)为主干网络的特征提取模块,采用上下文特征增强的 CSPDarknet(contextual feature augmentation block cross stage partial darknet, CFAB-CSPDarknet)作为主干网络,通过跨阶段特征提取兼顾局部与全局特征。对于不同尺寸的目标,引入动态注意力机制(dynamic attention, DAttention)<sup>[25]</sup>构建尺度

内特征交互模块,根据螺栓螺帽图像内容自适应调整处理粒度,更精确地捕捉锈蚀螺栓螺帽等微小目标的关键特征。为了增强算法的特征表征能力,设计一种结合局部卷积和全局自注意力机制的特征融合模块,卷积操作提取边缘和纹理等局部细节,自注意力机制建模目标在图像中的全局位置关系与上下文信息,实现局部与全局特征的动态加权融合并对锈蚀螺栓螺帽检测算法优化。

## 1 相关工作

### 1.1 基于 CNN 的目标检测

Girshick 等<sup>[26]</sup>首次将 CNN 算法用于目标检测。基本思路是将图像分成许多候选区域,对每个候选区域进行特征提取和分类。在此基础上,He 等<sup>[27]</sup>构建使用空间金字塔池化的 CNN 算法,接受任意尺寸的输入图像,并能够输出固定大小的特征向量。为进一步平衡检测精度与速度,Girshick 等<sup>[16]</sup>提出在卷积特征图上共享计算,加速检测过程。Ren 等<sup>[28]</sup>则通过引入区域提议网络(region proposal network, RPN),它是一种专用于生成候选区域的卷积网络,克服了 R-CNN 和 Fast R-CNN 算法在区域提取中的效率瓶颈。Kirillov 等<sup>[29]</sup>通过简单的网络连接改变,在基本不增加原有算法计算量的情况下,大幅度提升小物体检测的性能。YOLO 系列算法<sup>[17]</sup>直接作用于整个图像,通过对图像分区域预测边界框及对应类别概率,并经加权与非极大值抑制后处理,最终输出高置信度的检测结果。Lin 等<sup>[30]</sup>使用特征金字塔网络和新的焦点损失函数,以解决对象检测中前景和背景示例之间的不平衡,从而提高准确性。Duan 等<sup>[31]</sup>则使用热图预测每个对象的中心。再借助另一网络预测对象的大小和方向,最终提高检测的精度。尽管基于 CNN 算法的检测算法取得了显著效果,这类算法通常仍依赖复杂的预处理与后处理步骤,在一定程度上限制了应用效率与端到端优化能力。

### 1.2 基于 Transformer 的目标检测

自 Dosovitskiy 等<sup>[32]</sup>将 Transformer 引入计算机视觉领域以后,学者们提出许多基于 Transformer 的目标检测算法。Zhu 等<sup>[22]</sup>通过提高注意机制的效率加速算法的收敛速度并降低算法复杂性,同时通过引入多尺度特征解决目标检测中的性能问题。在此基础上,Wang 等<sup>[33]</sup>提出在抽象的特征空间上进行操作,并自适应地将计算分布在空间域。Roh 等<sup>[34]</sup>则采用稀疏注意力机制,仅让部分关键查询关注最相关的区域,减少计算量,提高推理速度。Meng 等<sup>[21]</sup>通过引入条件解码器,让每个查询点根据当前特征图的位置和上下文动态调整,提高收敛速度、增强目标检测的精度。Wang 等<sup>[20]</sup>借鉴 Faster R-CNN 算法和 YOLO 算法中的先验框设计,引入锚框机制对查询进行初始化,从而提升目标检测的精度。Dai 等<sup>[35]</sup>提出一种动态特征提取机制,使算法能更灵活地关注目标区域,从而提升检测效果。值得注意的是,RT-DETR 算法融合了传统 CNN 算法与 Transformer 的优势,在保持较高检测精度的同时显著提升推理速度,成为更适合边缘设备和工业检测场景的解决方案。尽管上述基于 Transformer 的算法在检测性能上取得进步,但它们仍面临参数规模大、计算冗余的问题。

### 1.3 输电线路金具检测

Luo 等<sup>[36]</sup>设计一个两阶段检测框架,该框架的超小目标感知模块基于混合注意力机制和显著区域提取算法,从高分辨率图像中识别螺栓密集区域,并生成对应的高分辨率图像块,该算法使用 K 均值聚类定位显著区域,无需额外标注。随后,局部螺栓检测模块对这些图像块进行精细缺陷检测,结合自注意力机制和多尺度特征融合策略。Song 等<sup>[37]</sup>提出小目标检测的网络结构,通过无锚点设计、解耦头设计思路,结合多尺度信息特征融合和交互显著区域捕捉策略,构建高效的小目标检测网络。Hua 等<sup>[38]</sup>在 YOLOv8 算法的基础上集成特定的注意力机制和卷积模块,在不增加算法复杂度的情况下提高螺栓检测的准确性和效率。Qi 等<sup>[39]</sup>提出基于优化 YOLOX 算法的级联网络算法,首先定位重要大型金具,然后在这些金具上检测小尺寸螺栓,显著提高了小螺栓的检测准确性。Zhang 等<sup>[40]</sup>则提出基于 Transformer 架构的端到端检测算法,通过引入改进的相对位置编码,考虑不同图像元素之间的全方位相对位置关系,并将位置关系整合到 DETR 算法编码器的自注意力模块中,增强算法对不同目标间相对位置关系的理解能力,该算法还结合了显式的属性知识,如螺栓所需的子组件(例如螺母和销),通过分类后设置属性得分并乘以可学习参数矩阵,将这些属性信息与位置信息融合,更准确地判断螺栓是否存在缺陷及其类别。但是对于输电线路中的螺栓螺帽的复杂背景和目标密集遮挡问题,上述算法仍存在局限性,尚未能实现稳定可靠的检测性能。图 1 是不同环境的螺

栓螺帽样例,图中标注的方形区域为锈蚀目标,圆形区域为正常目标。



图1 无人机拍摄的螺栓螺帽图像示例

Fig.1 Examples image of bolts and nuts captured by a drone

## 2 BN-DETR 算法

### 2.1 BN-DETR 算法的总体框架

BN-DETR 算法总体结构如图2所示。图中 conv 代表卷积操作,  $k$  代表卷积核大小,  $s$  代表步长,  $F_{BN}$  是批归一化处理,使用的激活函数是  $F_{SiLU}$ 。主干网络通过 CFAB 模块与卷积核大小为  $3 \times 3$ , 步长为 2 的卷积逐步提取并融合不同阶段的特征  $S_3, S_4, S_5, F_5$ 。融合块(图3)包含两个  $1 \times 1$  的卷积,调整通道的数量,  $n$  个重参数化卷积特征融合,并通过元素相加输出 2 个分支。使用 DAttention 机制增强小目标区域感知,借助卷积和注意力加权的特征融合(conv attention weighted feature fusion module, CAWFF)模块实现自适应加权特征融合,再将拼接的结果通过不确定性最小查询选择与解码头部输出检测结果。在保持高精度的同时显著提升对螺栓螺母小目标的检测能力与场景适应性。要解决输电线路中螺栓螺帽图像目标检测的问题,在维持较高准确性的同时,保证快速处理能力。BN-DETR 算法的主干网络结合跨阶段局部网络的思想,设计多尺度特征提取的结构,不仅能保留螺栓螺帽的细节特征,还能保证算法面对复杂背景的鲁棒性。随后,引入 DAttention 注意力机制,形成新的尺度内特征交互模块,让算法能够动态聚焦于螺栓螺帽等小目标区域,避免传统全局注意力对背景噪声的过度关注。最后,在混合编码器中引入自适应加权的多元特征融合机制,弥补单一特征在小目标检测中的局限性。轻巧的体系结构和高级功能处理技术的整合不仅优化螺栓螺帽检测中的性能,而且还促进各种操作场景中的可扩展性和适应性。

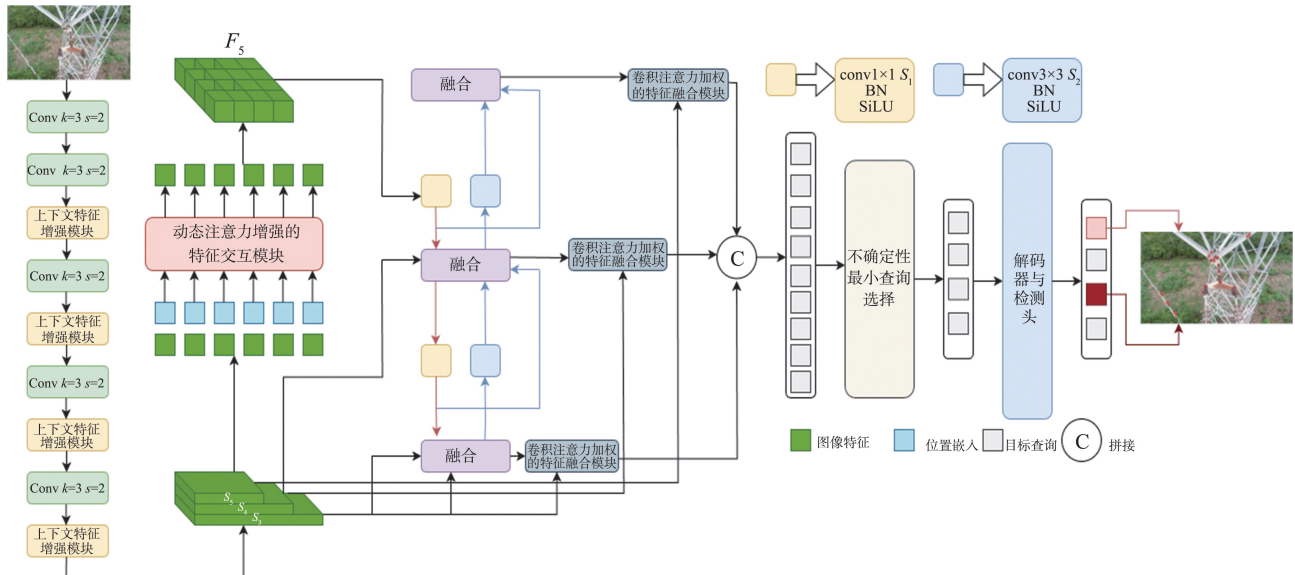


图2 BN-DETR 算法的整体结构图

Fig.2 The overall architecture of BN-DETR algorithm

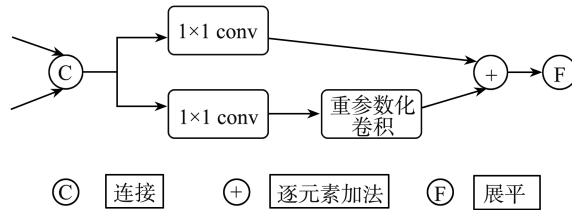


图 3 融合结构图

Fig.3 Fusion structure diagram

### 2.2 具有多尺度特征融合机制的骨干网络

由于输电线路的环境复杂,存在大量干扰背景,如金属反光、树木、电线等。为了解决以上问题,需要增强算法的细节特征提取能力。传统的 CNN 算法从本质上来说还是只在局部进行操作,但是在输电线路的螺栓螺帽检测中,全局信息的处理能力同样重要。本文采用 CSPDarknet 作为主干网络的同时,引入跨阶段特征融合策略与 Transformer 机制,嵌入新设计的 CFAB 模块,克服传统卷积在特征提取中的局限性。CFAB 模块同时保留局部卷积操作与全局 Transformer 处理机制,能够在多个尺度上高效聚合特征,从而在复杂背景中更好地保持目标的结构与细节信息。

CFAB 模块如图 4 所示,输入特征经过卷积层处理,随后输出分为 2 个部分:一部分直接传递至输出,另一个部分则送入多个 CFAB 模块进一步处理。在每个 CFAB 模块中,输入特征图首先通过  $1 \times 1$  卷积层调整通道数,随后经过 split 操作沿通道维度拆分为 2 个分支。第 1 个分支直接保留,作为后续拼接的一部分。第 2 个分支则输入至加性注意力(additiveblock)模块,通过卷积、归一化和激活等一系列操作提取局部特征,增强空间上下文信息。

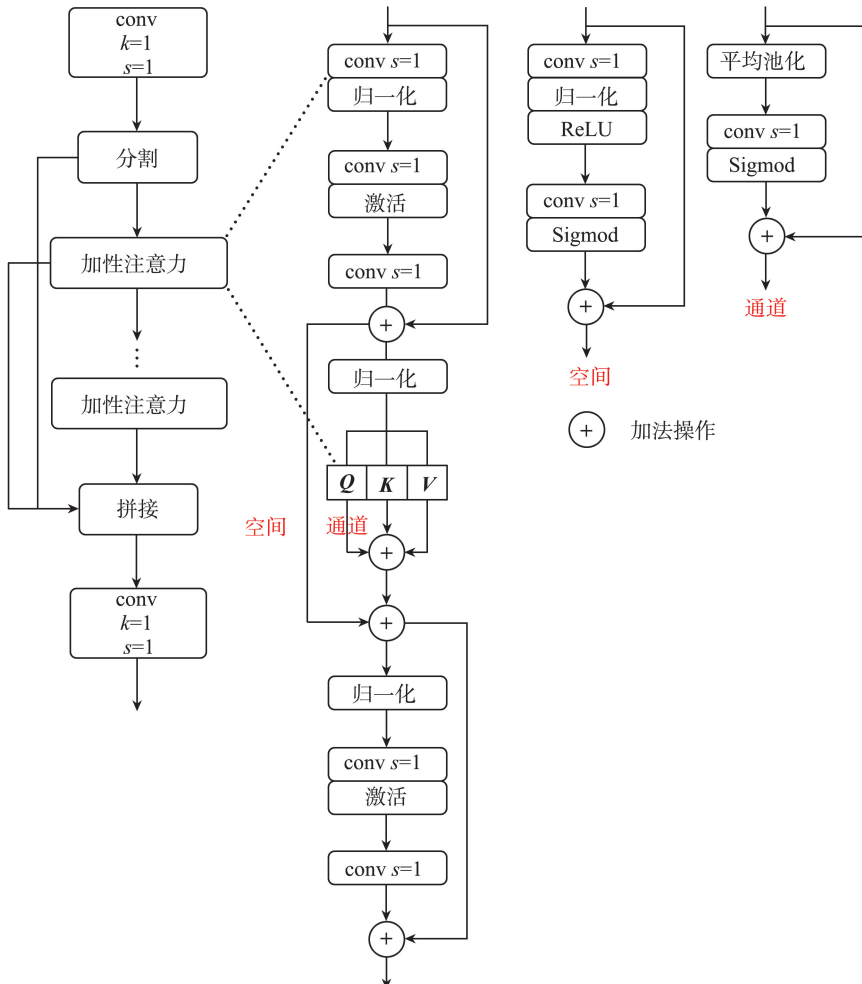


图 4 CFAB 模块结构图

Fig.4 Module of CFAB architecture diagram

第2个分支通过 $1 \times 1$ 卷积生成查询 $Q$ 、键 $K$ 、值 $V$ 。其中, $Q$ 和 $K$ 分别输入2个不同的操作模块:空间操作和通道操作。空间操作通过卷积捕获特征图的空间关系,首先使用 $3 \times 3$ 深度卷积(步长为1,填充为1)独立处理各通道,提取空间局部特征;再通过 $1 \times 1$ 点卷积融合通道信息。该结构增强对局部区域的感知能力,特别适合处理输电线路图像中的复杂背景干扰,同时保持较低的计算复杂度。通道操作用于重新校准通道间的特征重要性。首先,通过全局平均池化压缩空间维度,生成通道描述符;再经由两个全连接层(降维比为16)学习通道间权重关系;最后,通过Sigmoid激活函数生成权重向量,并与原始特征逐元素相乘。该操作自适应地增强与锈蚀螺栓、螺帽检测相关的通道特征(如纹理和边缘信息),抑制无关噪声,从而提高算法在小目标检测中的鲁棒性。

为了降低操作的复杂度和消除冗余特征,本文将 $Q$ 和 $K$ 加法融合以后再与 $V$ 投影操作。将新生成的 $n$ 个分支和第1个分支沿通道维度进行拼接,并通过 $1 \times 1$ 卷积调整通道数,最终得到输出特征。相较于传统的resnet18,引入多级特征聚合CFAB模块的CSPDarknet不仅减少复杂度,更能实现浅层和深层特征有效融合,从而充分保留细粒度信息,即使面对输电线路中复杂背景的干扰,也能精确的保留螺栓螺帽的细节信息。

### 2.3 引入DAttention注意力机制的尺度内特征交互模块

在输电线路的螺栓螺帽图像中,目标尺寸差异显著,螺栓螺帽大小往往取决于固定的器件类型。特别是在目标的尺寸大小不一且分布十分密集的情况下,传统多头自我注意力层在处理此类图像块上都施加统一的全局注意力,忽略特征中不同的潜在频率,不仅限制对复杂特征的建模能力,也在不同尺寸目标的高分辨率图像上具有较大计算量。所以,本文引入DAttention(结构如图5所示)注意力机制。图5中 $w_q$ 、 $w_k$ 、 $w_v$ 、 $w_o$ 为投影矩阵、 $\theta_{\text{offset}}$ 为偏移网络、Reference为参考点、Deformed为变形点、查询向量 $q$ 、键 $k$ 、值 $v$ 、注意力头 $z$ 、位置流 $R$ 。与传统自注意力不同,DAttention的稀疏注意力机制只关注有限的关键点,根据输入特征自适应地调整采样位置,能够更好地处理螺栓螺帽在不同视角或距离下的几何形变,提高检测鲁棒性。同时计算量只与采样点数相关,而非整个特征图。

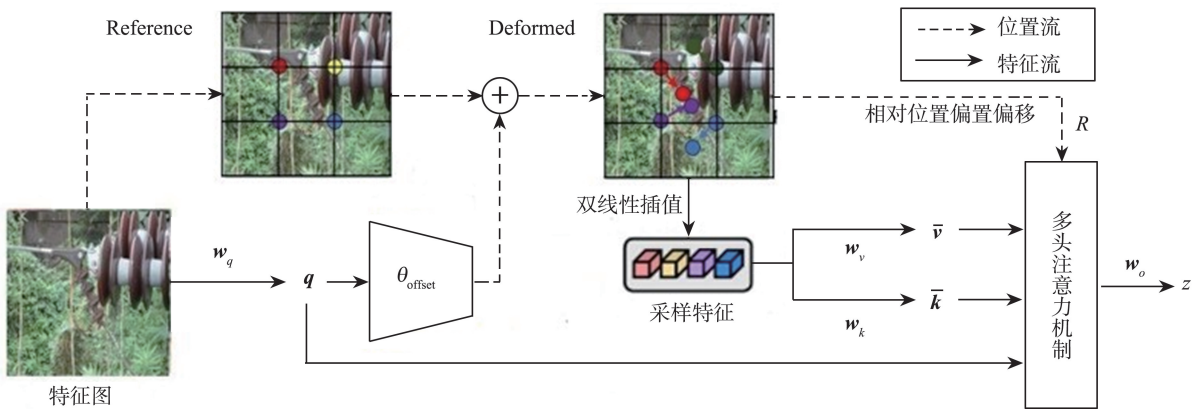


图5 DAttention 结构图

Fig.5 DAttention structural diagram

关键点采样策略。描述采样过程,选取 $E$ 个关键点,例如 $E=64$ ,基于网格采样并引入自适应偏移。即

$$P = \{(x_i, y_i) \mid i = 1, 2, \dots, E\}, \quad (1)$$

式中 $(x_i, y_i)$ 为选取的关键点的坐标。偏移后的网格坐标为

$$(x_h, y_h) = (x_i + \Delta x_i, y_i + \Delta y_i), \quad (2)$$

式中 $\Delta x_i$ 、 $\Delta y_i$ 为可学习偏移量。

复杂度对比。针对传统多头注意力和DAttention计算复杂度的简化对比。分析基于输入特征图尺寸 $H \times W \times C$ ( $H$ 、 $W$ 、 $C$ 分别为高、宽、通道数),其中序列长度 $N = H \times W$ ,DAttention中采样为 $E$ 个关键点( $E \ll N$ )。通过式(3)、(4)可以看出二者差异。传统多头注意力通过对所有位置之间进行全局自注意力计算,捕捉全局上下文关系,计算复杂度来自注意力分数的矩阵运算和加权求和,即

$$O((N)^2 C), \quad (3)$$

式(3)反映注意力计算中  $QK^T$  的矩阵乘法( $N \times N \times C$ ),主导传统多头注意力的二次复杂度,在高分辨率图像中导致极高的计算量。

DAttention 通过稀疏采样仅关注  $E$  个关键点,降低全局交互到局部交互。复杂度为

$$O(ENC), \quad (4)$$

式(4)涵盖采样、注意力权重计算与加权求和(每个查询处理  $E$  个关键点,复杂度与  $N=H \times W$ )呈线性关系。由于  $E$  为小常数,远小于  $N$ ,显著减少计算量。

处理高级特征的计算过程为

$$Q=K=V=F_{\text{flatten}}(S_5), \quad (5)$$

$$F_5=R_{\text{reshape}}(D_{\text{attention}}(Q,K,V)). \quad (6)$$

式(6)中, $F_5$  为经过处理后的  $S_5$  特征, $R_{\text{reshape}}$  改变张量的维度, $D_{\text{attention}}$  为注意力机制。

## 2.4 卷积和注意力加权的特征融合模块

在输电线路的螺栓螺帽图像中,目标的占比通常非常小,极易受到复杂背景的干扰。经过主干网络提取的多尺度特征  $S_3$ 、 $S_4$ 、 $S_5$ 、 $F_5$  涵盖不同维度的多种信息。为充分挖掘特征金字塔的融合潜力,本文提出的 CAWFF 模块有效地传输并融合了目标位置和语义信息,通过关注不同尺度上特征表示的差异和相关性,整合目标位置与语义信息。为了得到最终的权重  $W_{\text{final}}$ ,模块接收 2 组特征:低层次特征  $X$  和高层次特征  $Y$ 。这些特征通常来自网络的不同层,低层次特征包含更多细节信息(如边缘、纹理),高层次特征包含更多语义信息(如目标的类别、位置)。首先将 2 个层次的特征进行初始融合,即

$$I=X+Y, \quad (7)$$

式中: $I$  为初始融合后的特征, $X$  为低层次特征, $Y$  为高层次特征。

随后,将初步融合的特征  $I$  输入跨特征聚合块(cross-feature aggregation block, CAFM)。CAFM 包含 2 个分支:全局分支利用自注意力机制建模长程依赖,捕获高频上下文信息;局部分支则专注于局部结构提取与噪声抑制。在全局分支中,通过卷积生成  $Q$ 、 $K$  和  $V$ ,其形状为  $\hat{H} \times \hat{W} \times \hat{C}$ ,接下来, $Q$  变形为  $\hat{Q} \in \mathbf{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ , $K$  变形为  $\hat{K} \in \mathbf{R}^{\hat{C} \times \hat{H} \times \hat{W}}$ 。通过  $\hat{Q}\hat{K}$  的对应关系计算注意力图  $A \in \mathbf{R}^{\hat{C} \times \hat{C}}$ 。全局注意力机制的分支的输出  $F_{\text{att}}$  为

$$F_{\text{att}}=W_{1 \times 1}A_{\text{attention}}(\hat{Q},\hat{K},\hat{V})+Y, \quad (8)$$

式中: $Y$  是输入特征, $W_{1 \times 1}$  为  $1 \times 1$  卷积, $A_{\text{attention}}$  为注意力机制,即

$$A_{\text{attention}}(\hat{Q},\hat{K},\hat{V})=\frac{\hat{V}S_{\text{softmax}}(\hat{K},\hat{Q})}{\alpha}, \quad (9)$$

式中: $S_{\text{softmax}}$  是归一化函数; $\alpha$  是一个可学习的缩放参数,可以控制  $\hat{Q}$  和  $\hat{K}$  的矩阵乘法得到的结果的大小。

在局部分支中,首先使用  $1 \times 1$  卷积调整通道尺寸,执行通道重组操作,再混合通道信息。随后将特征在通道维度分组,对每组分别施加深度可分离卷积,最后合并各组输出并通过  $3 \times 3$  卷积提取空间特征。即输出为

$$F_{\text{conv}}=W_{3 \times 3 \times 3}(C_s(W_{1 \times 1}(Y))), \quad (10)$$

式中: $F_{\text{conv}}$  是局部分支的输出, $W_{1 \times 1}$  表示  $1 \times 1$  卷积, $W_{3 \times 3 \times 3}$  表示  $3 \times 3 \times 3$  卷积, $C_s$  代表通道重组操作。

CAFM 模块的最终输出由全局与局部分支共同决定,最终输出为

$$F_{\text{out}}=F_{\text{conv}}+F_{\text{att}} \circ \quad (11)$$

为了获得最终精确模拟权重  $W_{\text{final}}$ , $F_{\text{out}}$  的每个通道都根据相应的输入特征进行调整。如图 6 所示,通过输入特征的内容,引导生成最终的、通道特定的模拟结果。特别地, $F_{\text{out}}$  和  $I$  的每个通道以交替的方式重组,通过全局和局部分支的特征计算像素级注意力分数,使算法能够自适应地强调每个空间位置的相关特征,而无需额外的像素级注意力模块,同时保持原有的特征信息,最终精确模拟权重为

$$W_{\text{final}}=(W_{7 \times 7}(C_s(I,F_{\text{out}}))), \quad (12)$$

式中  $W_{7 \times 7}$  是  $7 \times 7$  大核卷积,加权后的最终输出为

$$F_{\text{final}}=I+W_{\text{final}}X+(1-W_{\text{final}})Y. \quad (13)$$

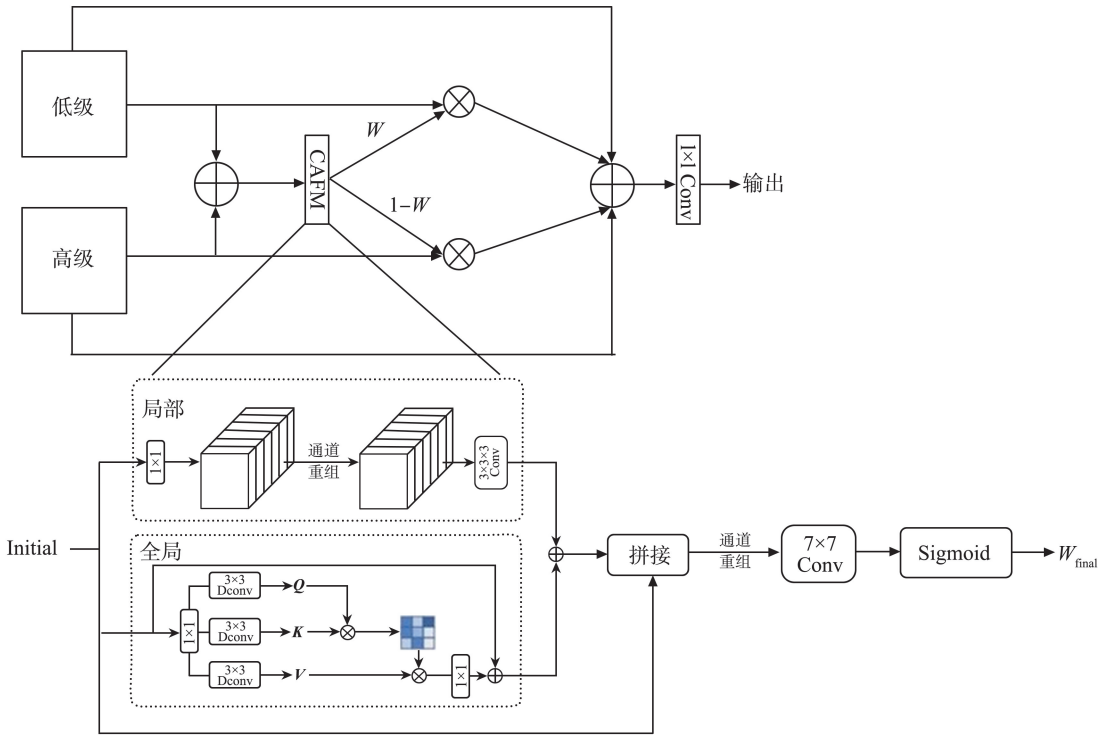


图 6 CAWFF 结构图

Fig.6 CAWFF structural diagram

### 3 实验结果和分析

#### 3.1 数据集和评估指标

本文采用 UAV-BN 数据集消融和对比试验评估 BN-DETR 算法的性能,采用 VisDrone-2019 数据集<sup>[41]</sup>和 a high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection (HIT-UAV) 数据集<sup>[42]</sup>实验验证 BN-DETR 算法的泛化能力。RT-DETR (r18) 是使用残差网络-18 (residual network-18, resnet18) 作为主干网络的 RTDETR 算法,RT-DETR (r34)、RT-DETR (r50) 同理。因 RT-DETR (r18) 算法的轻量级架构适配无人机平台的资源限制,选择 RT-DETR (r18) 算法作为消融实验基线。RT-DETR (r18) 算法的适中参数量和浮点运算次数评估所提模块对性能的贡献。

本文构建专业的数据集 UAV-BN,数据集由输电线路锈蚀和正常的螺栓螺帽图像构成。图 1 为 UAV-BN 数据集图像,数据集中图像的分辨率各不相同,格式有  $960 \times 540$ 、 $2\,000 \times 1\,500$ 、 $2\,736 \times 1\,539$ ,图像 1 160 张。数据集中有两类目标,锈蚀的螺栓螺帽类别名为 lslmxs,正常的螺栓螺帽类别名为 lslmzc。基于 YOLO 算法使用的数据集的标注格式。为了满足实验要求,数据集分为训练集验证集和测试集。其中训练集图像 895 张,验证集图像 131 张,测试集图像 134 张。数据集的各类别标注信息如表 1 所示。

表 1 UAV-BN 数据集标注数量信息  
Table 1 UAV-BN dataset number of data annotation

类型	图像数量	lslmxs	lslmzc
训练集	895	11 995	2 596
验证集	131	1 150	225
测试集	134	940	351
总计	1 160	14 085	3 172

VisDrone-2019 数据集包含从各种城市和郊区收集的图像 10 209 张,包括城市街道、居民区和工业区图像,通过不同高度、相机方向和天气条件下操作的无人机获取图像。数据集中图像的分辨率各不相同,但大多数图像的分辨率为  $2\,000 \times 1\,500$  像素。数据集的标注信息如表 2 所示。

表 2 VisDrone-2019 数据集标注数量信息  
Table 2 VisDrone-2019 dataset number of data annotation

类型	图像数量	Pedestrian	People	Bicycle	Car	Van	Truck	Tricycle	Awning Tricycle	Bus	Motor
训练集	6 471	109 185	38 560	13 069	187 004	32 702	16 284	6 387	4 377	9 117	40 377
验证集	548	8 844	5 125	1 287	14 064	1 975	750	1 045	532	251	4 886
测试集	1 610	21 006	6 376	1 302	28 074	5 771	2 659	530	599	2 940	5 845
总计	8 629	139 035	50 061	15 658	229 142	40 448	19 693	7 962	5 508	12 308	51 108

HIT-UAV 数据集由无人机捕获的高空航空图像组成,包含从学校,高速公路和停车场等位置收集的红外热图像 2 898 张。数据集中的每个图像的分辨率为  $640 \times 512$ 。其中,训练集图像 2 029 张,验证集图像 290 张和测试集图像 579 张。数据集的详细标注信息如表 3 所示。

表 3 HIT-UAV 数据集标注数量信息  
Table 3 HIT-UAV dataset number of data annotation

类型	图像数量	Person	Car	Bicycle	OtherVehicle	DontCare
训练集	2 029	8 473	5 247	3 618	102	110
验证集	290	1 152	719	554	12	7
测试集	579	2 602	1 338	792	34	31
总计	2 898	12 227	7 304	4 964	148	148

### 3.2 实验环境和参数设置

网络实验环境基于 Ubuntu20.04,Python 3.8.10 和 Pytorch 2.4.1,中央处理器是 Intel(R) Xeon(R) Gold 6430,显卡是英伟达 A6000,计算统一设备架构(compute unified device architecture,CUDA)为 12.1 版本。训练参数中,优化器为 AdamW,批处理大小设置为 4,训练轮次为 300,学习率为  $1 \times 10^{-4}$ 。为了避免过度拟合,使用权重衰减策略,重量衰减值设置为 0.000 1。实验选择的输入图像大小为  $640 \times 640$ 。

### 3.3 模块消融实验

模块消融实验在 UAV-BN 数据集上进行,基线算法为 RT-DETR (R18)。实验评价指标为交并比阈值小于 0.5 的平均精度均值(记为平均精度均值@50)和交并比阈值从 0.5 到 0.95 的平均精度均值(记为平均精度均值@50-95),消融实验的结果如表 4 所示,表中的粗体代表最优值,实验是通过逐步整合每个改进模块进行。

表 4 消融实验结果  
Table 4 Ablation experiment results

CFAB-CSPDarknet 模块	Dattention 模块	CAWFF 模块	平均精度均值@50/%	平均精度均值@50-95/%	参数量/ $10^6$	浮点运算次数/ $10^9$
			34.5	13.0	19.8	56.9
✓			36.3	13.4	<b>14.9</b>	<b>49.8</b>
	✓		36.5	13.8	19.8	57.2
		✓	36.7	13.7	21.4	65.1
✓	✓		36.8	13.9	<b>14.9</b>	50.1
✓	✓	✓	<b>37.5</b>	<b>13.9</b>	16.5	58.7

注:“✓”表示当前实验含有这个模块。

实验结果表明,CFAB-CSPDarknet 替换主干网络 Resnet18 时,平均精度均值@50 和平均精度均值@50-95 分别提高 1.8%和 0.4%,说明 CFAB-CSPDarknet 在提取特征方面更有效。参数量和浮点运算次数分别降低  $4.9 \times 10^6$  和  $7.1 \times 10^9$ ,证明 CFAB-CSPDarknet 模块更加轻量。引入 Dattention 注意力机制后,平均精度均值@50 和平均精度均值@50-95 分别提高 2.0% 和 0.8%,同时参数量和浮点运算次数基本不变,证明根据输入特征自适应地调整采样位置以适应不同尺寸目标的有效性。使用卷积和注意力机制加权的特征融合模块增强特征,有效地提高检测的准确性。最后,整合所有提出的改进策略的 BN-DETR 算法,平均精度均值@50 为 3.0%和平均精度均值@50-95 为 1.4%,参数量降低  $3.3 \times 10^6$ ,表明改进策略的 BN-DETR 算法更稳定。

### 3.4 对比实验

本文选取 YOLO 系列算法和 RT-DETR 算法作为对比对象。评价指标为精确率、召回率、平均精度均值@50、平均精度均值@50-95、参数量、浮点运算次数、帧率。以上选取的算法在实时性与准确性之间具有较好的平衡性,在实际电力检测等类似场景中得到广泛应用。

表5为多种算法在螺栓螺帽检测任务中的实验结果。螺栓螺帽检测方面, YOLOv8 算法的平均精度均值@50为34.2%, YOLOv5 算法的平均精度均值@50为32.9%。和 RT-DETR(r34) 算法相比, YOLOv5 算法具有其高效的特征提取能力,且能冗余计算,在保持较低复杂度的同时实现较好的检测效果。相比之下, RT-DETR 算法表现略逊,因为输电线路检测场景中存在大量背景干扰和目标分布不均匀的特点,影响了 RT-DETR 算法的全局特征提取能力。值得注意的是,本文的算法在所有算法中取得最高的平均精度均值@50。与 YOLOv8 算法相比,本文的算法的平均精度均值@50提高3.3%,平均精度均值@50达到37.5%,平均精度均值@50-95也从12.8%提高至14.4%。此外,本文的算法在减少算法复杂度方面表现突出,参数量仅为 $1.65 \times 10^7$ 浮点运算次数为 $5.87 \times 10^{10}$ ,显著小于 RT-DETR(r50) 算法的 $4.19 \times 10^7$ 和 $1.295 \times 10^{11}$ 。这表明该算法在保持较低计算成本的同时,显著提升螺栓螺帽检测能力。

表5 多种算法在螺栓螺帽检测任务中的对比实验对  
Table 5 Comparison experiment of multiple algorithms in bolt nut detection task

算法	精确率/%	召回率/%	平均精度 均值@50/%	平均精度 均值@50-95/%	参数量/ $10^3$	浮点运算 次数/ $10^9$	帧率
基线	54.9	41.0	34.5	13.0	19.8	56.9	79.5
YOLOv3	47.2	38.1	33.0	12.1	98.4	261.8	52.0
YOLOv5	47.0	37.4	32.9	11.4	22.1	<b>52.5</b>	102.0
YOLOv6	23.8	35.7	23.4	7.8	51.2	158.3	98.0
YOLOv8	47.8	38.3	34.2	12.8	23.3	67.4	<b>111.8</b>
YOLOv9	59.0	37.0	36.4	13.5	16.5	60.0	99.0
YOLOv10	47.7	37.4	31.7	12.0	<b>16.4</b>	63.4	96.0
YOLOv11	49.5	35.9	33.5	12.7	20.3	67.7	89.2
RT-DETR (r50)	54.5	39.0	33.4	12.0	41.9	129.5	57.4
RT-DETR (r34)	57.1	42.0	35.5	13.5	31.1	88.8	60.5
本文	<b>63.4</b>	<b>43.5</b>	<b>37.5</b>	<b>13.9</b>	16.5	58.7	81.2

### 3.5 实验结果分析

图7为部分检测结果图分别取自于不同的背景和视角。图7(a)~(f)为基线算法检测结果,图7(g)~(l)为 BN-DETR 算法的检测结果。遮挡是输电线路检测的常见问题,例如螺栓螺帽被其他目标或塔架结构部分覆盖,导致特征不完整。图像中除(c)、(i)外都显示了这种场景。在(a)~(f)图像中,遮挡区域的螺栓螺帽经常被忽略或边界框偏移。例如,左侧塔架上的绝缘子遮挡导致多个小目标漏检,见图7(b)。图7(e)~(f)中,结构部件覆盖的部分螺栓未被检测,召回率较低。这是因为基线依赖传统全局注意力,难以从局部遮挡中恢复关键特征。图7(g)~(l)显示 BN-DETR 算法成功检测更多遮挡目标,边界框紧贴锈蚀区域。绝缘子下的螺栓被准确标注,即使目标被电线部分遮挡,算法仍捕捉到锈蚀纹理。提升了算法在实际巡检中的可靠性,避免了人为干预。背景干扰包括复杂纹理(如植被、建筑物)和光照变化,导致目标与背景灰度趋同或特征混淆。图7(a)、(e)最为显著。图7(a)~(e)中,背景干扰导致识别错误,反映基线对全局噪声的敏感性,易丢失小目标细节。图7(g)~(l)显示 BN-DETR 算法显著减少假检测,目标标注更精确。本文算法过滤了景观干扰,聚焦锈蚀特征,整体平均精度均值@50提高。该算法在干扰下保持高鲁棒性,适合多变户外环境。



图 7 基线算法和 BN-DETR 的检测结果图  
Fig.7 Detection results of baseline and BN-DETR algorithm

### 3.6 泛化实验

为了进一步验证算法的泛化性能,泛化实验采用基线算法和 BN-DETR 算法在 VisDrone-2019 数据集和

HIT-UAV 数据集上进行。实验结果如表 6、7 所示。VisDrone-2019 数据集中有许多图像的目标特别密集, HIT-UAV 数据集的目标中存在较多遮挡, 这都是螺栓螺帽数据集的特点之一。实验的结果表明, 不论是针对密集小目标还是多重遮挡的小目标时 BN-DETR 算法都具有强大的特征提取和融合能力, 说明 BN-DETR 算法的可用性和可扩展性。

表 6 VisDrone-2019 数据集评价指标  
Table 6 VisDrone-2019 dataset evaluation metric

算法	精确率/%	召回率/%	平均精度 均值@ 50/%	平均精度 均值@ 50-95/%	参数量/ $10^6$	浮点运算 次数/ $10^9$
基线	56.0	38.5	37.1	21.2	19.8	<b>56.9</b>
BN-DETR	<b>56.4</b>	<b>39.7</b>	<b>38.4</b>	<b>22.2</b>	<b>16.5</b>	58.7

表 7 HIT-UAV 数据集评价指标  
Table 7 HIT-UAV dataset evaluation metric

算法	精确率/%	召回率/%	平均精度 均值@ 50/%	平均精度 均值@ 50-95/%	参数量/ $10^6$	浮点运算 次数/ $10^9$
基线	<b>88.2</b>	75.9	80.8	51.5	19.8	<b>56.9</b>
BN-DETR	86.9	<b>81.1</b>	<b>82.1</b>	<b>53.0</b>	<b>16.5</b>	58.7

## 4 结语

针对输电线路锈蚀螺栓螺帽检测的问题提出 BN-DETR 算法。利用主干网络、注意力机制、特征融合等多方面的设计减少特征提取过程中的信息丢失, 保留丰富的螺栓螺帽特征信息, 区分复杂背景, 适应目标的多尺度变化。提高螺栓螺帽的检测精度。实验结果表明, 与基线 RT-DETR 算法相比, 本文算法在 UAV-BN 数据集上的平均精度均值@ 50 提高 3.0%, 平均精度均值@ 50-95 提高 0.9%, 同时减少参数量和浮点运算次数。对比实验表明, BN-DETR 算法相较于 YOLO 系列算法和其他主干网络的 RT-DETR 算法在检测精度与算法复杂度的平衡方面较好。此外, 泛化实验证明 BN-DETR 算法在 VisDrone-2019 和 HIT-UAV 数据集上的评价指标较高, 进一步验证本文算法在密集小目标与多重遮挡场景下的泛化能力。BN-DETR 算法在保持较低计算成本的同时, 实现输电线路螺栓螺帽检测能力的显著提升, 特别适用于无人机巡检等对实时性和精度要求较高的场景。未来研究可进一步探索算法轻量化与实时性的优化, 同时结合多模态数据融合技术, 提升复杂环境下的检测性能。

### 参考文献:

- [1] ALHASSAN A B, ZHANG X D, SHEN H M, et al. Power transmission line inspection robots: a review, trends and challenges for future research[J]. International Journal of Electrical Power & Energy Systems, 2020, 118:105862.
- [2] EISENBEISS H. A mini unmanned aerial vehicle (UAV): system overview and image acquisition[J]. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2004, 36(5/W1):1-7.
- [3] 周庆泽, 郭擎, 王海荣, 等. 双判别器深度残差 GAN 高光谱图像融合[J]. 中国图象图形学报, 2024, 29(7):2046-2062. ZHOU Qingze, GUO Qing, WANG Hairong, et al. Dual-discriminator deep residual GAN hyperspectral image fusion[J]. Journal of Image and Graphics, 2024, 29(7):2046-2062.
- [4] 孙彬, 高云翔, 诸葛吴为, 等. 可见光与红外图像融合质量评价指标分析[J]. 中国图象图形学报, 2023, 28(1):144-155. SUN Bin, GAO Yunxiang, ZHUGE Wuwei, et al. Analysis of quality evaluation metrics for visible and infrared image fusion [J]. Journal of Image and Graphics, 2023, 28(1):144-155.
- [5] 韩鲁光, 陈纯毅, 申忠业, 等. 视觉显著性驱动的全景渲染图非局部降噪[J]. 中国图象图形学报, 2024, 29(4):939-952. HAN Luguang, CHEN Chunyi, SHEN Zhongye, et al. Panoramic rendering non-local denoising driven by visual saliency[J]. Journal of Image and Graphics, 2024, 29(4):939-952.
- [6] 张珂, 周睿恒, 石超君, 等. 自监督 E-Swin 的输电线路金具检测[J]. 中国图象图形学报, 2023, 28(10):3064-3076.

- ZHANG Ke, ZHOU Ruiheng, SHI Chaojun, et al. Self-supervised E-Swin transmission line hardware detection[J]. *Journal of Image and Graphics*, 2023, 28(10):3064-3076.
- [7] 赵振兵,蒋志钢,李延旭,等. 输电线路部件视觉缺陷检测综述[J]. *中国图象图形学报*, 2021, 26(11):2545-2560.  
ZHAO Zhenbing, JIANG Zhigang, LI Yanxu, et al. A review of visual defect detection in transmission line components[J]. *Journal of Image and Graphics*, 2021, 26(11):2545-2560.
- [8] AKBARI Y, ALMAADEED N, AL-MAADEED S, et al. Applications, databases and open computer vision research from drone videos and images: a survey[J]. *Artificial Intelligence Review*, 2021, 54:3887-3938.
- [9] DILSHAD N, HWANG J Y, SONG J S, et al. Applications and challenges in video surveillance via drone: a brief survey[C]//2020 International Conference on Information and Communication Technology Convergence (ICTC). Jeju: IEEE, 2020:728-732.
- [10] SAMPEDRO C, MARTINEZ C, CHAUHAN A, et al. A supervised approach to electric tower detection and classification for power line inspection[C]//2014 International Joint Conference on Neural Networks (IJCNN). Beijing: IEEE, 2014:1970-1977.
- [11] MENG Xiaobo, CHEN Wencan, MEI Hongwei, et al. Corrosion mechanism of UHV transmission line tower foot in southern china[J]. *IEEE Transactions on Power Delivery*, 2023, 39(1):210-219.
- [12] ZHANG Heng, YANG Wen, YU Huai, et al. Detecting power lines in UAV images with convolutional features and structured constraints[J]. *Remote Sensing*, 2019, 11(11):1342.
- [13] ZHENG Jianfeng, WU Hang, ZHANG Han. Insulator-defect detection algorithm based on improved YOLOv7[J]. *Sensors*, 2022, 22(22):8801.
- [14] ZHAO Zhenbing, WU Xueliang, QI Yincheng, et al. A knowledge-guided model of fitting detection in aerial transmission line images[J]. *Energy Reports*, 2020, 6:1071-1078.
- [15] LUO Peng, WANG Bo, WANG Hongxia, et al. An ultrasmall bolt defect detection method for transmission line inspection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72:1-12.
- [16] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015:1440-1448.
- [17] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016:779-788.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//In Proceedings of the 31st International Conference on Neural Information Processing Systems. California: Curran Associates Incorporation, 2017:6000-6010.
- [19] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Computer Vision -ECCV 2020. Glasgow: Springer, 2020:213-229.
- [20] WANG Yingming, ZHANG Xiangyu, YANG Tong, et al. Anchor DETR: query design for transformer-based object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence.Vancouver: AAAI, 2022:2567-2575.
- [21] MENG Depu, CHEN Xiaokang, FAN Zejia, et al. Conditional DETR for fast training convergence[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021:3631-3640.
- [22] ZHUXizhou, SU Weijie, LU Lewei, et al. Deformable DETR: deformable transformers for end-to-end object detection [J/OL]. arXiv preprint. (2020-10-15)[2025-10-08]. <https://arxiv.org/abs/2010.04159>.
- [23] LIUShilong, LI Feng, ZHANG Hao, et al. DAB-DETR: dynamic anchor boxes are better queries for detr [J/OL]. arXiv preprint. (2022-03-07)[2025-10-08]. <https://arxiv.org/abs/2203.03605>.
- [24] ZHAOYian, LÜ Wenyu, XU Shangliang, et al. Detsr beat YOLOS on real-time object detection[J/OL]. arXiv preprint. (2023-04-17)[2025-10-08]. <https://arxiv.org/abs/2304.08069>.
- [25] XIA Zhuofan, PAN Xuren, SONG Shiji, et al. Vision transformer with deformable attention [C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022:4784-4793.
- [26] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE, 2014: 580-587.
- [27] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2014, 37(9):1904-1916.

- [28] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 39(6):1137-1149.
- [29] KIRILLOV A, GIRSHICK R, HE K, et al. Panoptic feature pyramid networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach:IEEE, 2019:6399-6408.
- [30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 99:2999-3007.
- [31] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. Centernet++ for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(5):3509-3521.
- [32] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J/OL]. arXiv preprint. (2020-10-22)[2025-10-08]. <https://arxiv.org/abs/2010.11929>.
- [33] WANG Tao, YUAN Li, CHEN Yunpeng, et al. PNP-DETR: towards efficient visual analysis with transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal:IEEE, 2021:4641-4650.
- [34] ROH B, SHIN J W, SHIN W, et al. Sparse DETR: efficient end-to-end object detection with learnable sparsity[J/OL]. arXiv preprint. (2021-11-29)[2025-10-08]. <https://arxiv.org/abs/2111.14330>.
- [35] DAI Xiyang, CHEN Yinpeng, YANG Jianwei, et al. Dynamic DETR: end-to-end object detection with dynamic attention[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021:462-471.
- [36] LUO Peng, WANG Bo, WANG Hongxia, et al. An ultrasmall bolt defect detection method for transmission line inspection [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72:5006512.
- [37] SONG Zhiwei, HUANG Xinbo, JI Chao, et al. Deformable YOLOS: detection and rust warning method of transmission line connection fittings based on image processing technology[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72:2504321.
- [38] HUA Guoxiang, ZHANG Huai, HUANG Chen, et al. An enhanced YOLOv8-based bolt detection algorithm for transmission line[J]. IET Generation, Transmission & Distribution, 2024, 18(24):4065-4077.
- [39] QI Yincheng, HUO Yalin, LIU Shaohang, et al. Optimized YOLOS based transmission line bolt cascade detection[C]//2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Prague: IEEE, 2022:2099-2102.
- [40] ZHANG Ke, LOU Wenshuo, WANG Jiacun, et al. PA-DETR: end-to-end visually indistinguishable bolt defects detection method based on transmission line knowledge reasoning[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72:1-14.
- [41] ZHU Pengfei, WEN Longyin, DU Dawei, et al. Detection and tracking meet drones challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11):7380-7399.
- [42] SUO Jiashun, WANG Tianyi, ZHANG Xingzhou, et al. HIT-UAV: a high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection[J]. Scientific Data, 2023, 10:227.

(编辑:陈丽萍)