

面向电力设备的多维交互注意力增强 YOLOv8n 多模态检测算法

童超¹, 饶鑫威², 徐十三悦², 王海波³, 王皓天⁴

(1. 国网江西省电力有限公司电力科学研究院, 江西 南昌 330009; 2. 江西师范大学数字产业学院, 江西 上饶 334000; 3. 南昌科晨电力试验研究有限公司, 江西 南昌 330096; 4. 国网江西省电力有限公司石城县供电分公司, 江西 赣州 342700)

摘要:为了解决可见光与红外图像融合导致颜色不连续、目标分布碎片化的问题,文中提出了一种面向电力设备的多维交互注意力增强YOLOv8n多模态检测算法。针对融合图像颜色不连续、缺乏视觉一致性现象,通过增强通道、高度、宽度三个维度之间特征相互作用,设计并引入了多维交互注意力模块;为了避免过拟合,使用平滑技术对Slide Loss 损失函数进行改进,替换了YOLOv8n的完全交并比(complete intersection over union, CIoU) Loss 损失函数。实验结果表明,改进YOLOv8n算法在可见光和红外融合图像目标检测任务中表现优异,在精确率、召回率和交并比(intersection over union, IoU)高阈值范围内的性能上均优于多种改进的YOLO系列算法,能够有效提升目标检测的精度。

关键词: 可见光与红外图像融合; YOLOv8n; 目标检测; 注意力机制; 双池化融合

中图分类号:TM 727 TP 391 文献标志码:A 文章编号:1006-348X(2025)03-0022-06

0 引言

在电力设备的运行监控与维护中,如何准确、高效地检测电力设备中的关键部件及其异常状态,对于保障电力系统的稳定运行,及时发现潜在故障具有重大意义。传统目标检测方法往往依赖单一种类的图像,如可见光图像或红外图像,这在一定程度上限制了检测的准确性和鲁棒性。可见光图像虽然能够提供丰富的纹理和细节信息,但在光照条件不佳或存在遮挡的情况下,目标的可辨识性会大幅下降;红外图像则能够通过热辐射信息突出目标物体,但其图像分辨率通常较低,且难以提供精细的纹理特征。因此,如何充分利用可见光与红外图像的互补信息,提升电力红外图像目标检测的性能,成为当前研究的一个重要方向^[1]。

近年来,随着深度学习技术的飞速发展,基于卷积神经网络(convolutional neural network, CNN)

的目标检测算法取得了显著进展,其中YOLO系列算法因其速度快、精度高而被广泛应用于各种实际场景。然而,面对复杂多变的电力红外图像环境,尤其是可见光与红外图像融合后的特殊场景,现有的目标检测算法仍存在一些不足之处^[2];文献[3]通过引入空间到深度卷积层(space-to-depth convolutional block, SPD-Conv)实现轻量化设计,结合卷积块注意力模块(convolutional block attention module, CBAM),增强关键特征提取能力,但对复杂背景下的小目标检测能力有待提升;文献[4]通过引入空洞卷积与均值池化扩大感受野,嵌入轻量型三维注意力机制(simple parameter-free attention module, SimAM),增强对感兴趣区域的关注,尽管检测精度显著提升,但在处理极端复杂背景或极端小目标时的性能有待提升;文献[5]通过K-means聚类优化锚点框尺寸,加入浅层特征图,增强小目标特征信息,并引入高效通道注意力机制

收稿日期:2025-04-22

基金项目:国家自然科学基金项目(面向电力设备监测的非对称双摄红外图像超分辨率重建与定位研究,62362040)。

作者简介:童超(1992),男,硕士,高级工程师,从事多光谱目标检测技术研究。

(efficient channel attention, ECA),提升特征提取能力,尽管检测精度显著提升,但模型复杂度增加导致检测速度下降;文献[6]通过多尺度特征融合和特征细化,显著提升了电力设备目标检测的精度,尤其是对小目标和多角度目标的检测能力,但模型复杂度较高,计算成本增加。

为了克服这些困难,文中提出了一种面向电力设备的多维交互注意力增强 YOLOv8n 多模态检测算法。通过采用轮廓角方向-要素匹配方法(contour angle orientation-coarse to fine, CAO-C2F)配准与加权融合技术,实现可见光和红外图像的精准对齐和有效融合,并在此基础上作出以下改进:1)在检测头前添加多维交互注意力模块;2)使用平滑技术对 Slide Loss 损失函数进行改进,替换 YOLOv8n 的完全交并比(complete intersection over union, CIoU)Loss 损失函数。

1 可见光与红外图像融合技术

在电力设备检测中,准确识别和定位设备的关键部件及其异常状态,对于保障电力系统的稳定运行和及时发现潜在故障具有重要意义。电力设备在运行过程中,可能会受到多种因素的影响,如光照变化、物体遮挡、背景复杂等。虽然可见光图像能够提供丰富的纹理和细节信息,但在这些复杂条件下,目标的可辨识性会显著下降。红外图像能够通过热辐射信息突出目标物体,尤其是在夜间或低光照条件下,能够提供可见光图像所无法提供的关键信息^[7]。然而,红外图像的分辨率通常较低,难以提供精细的纹理特征。因此,将可见光图像与红外图像进行融合,可以有效整合两类图像的互补信息,提升目标检测的准确性和鲁棒性。这种融合技术不仅能够增强目标的可辨识性,还能够提高检测算法的适应性和可靠性,尤其是复杂环境和低光照条件下的检测性能^[8]。

为了充分利用可见光图像和红外图像的互补信息,文中使用 CAO-C2F 配准^[9]与加权融合的方法对两类图像进行处理。CAO-C2F 配准是一种基于特征点匹配的图像配准算法,通过提取图像中的特征点并计算其相似性,能够有效确定可见光图

像与红外图像之间的几何变换关系,实现精准的空间对齐。在配准完成后,进一步采用加权融合策略,对图像进行融合。加权融合通过分别为可见光图像和红外图像分配不同的权重,将图像信息进行加权求和,生成融合后的图像。在这一过程中,根据可见光图像和红外图像的特点,合理分配权重,即保留可见光图像丰富的纹理和细节信息,同时又突出红外图像中目标的热辐射特征。通过 CAO-C2F 配准与加权融合的结合,文中成功将两类图像信息进行了有效整合,为后续的目标检测任务提供了更高质量的输入图像,融合的图片效果如图 1 所示。

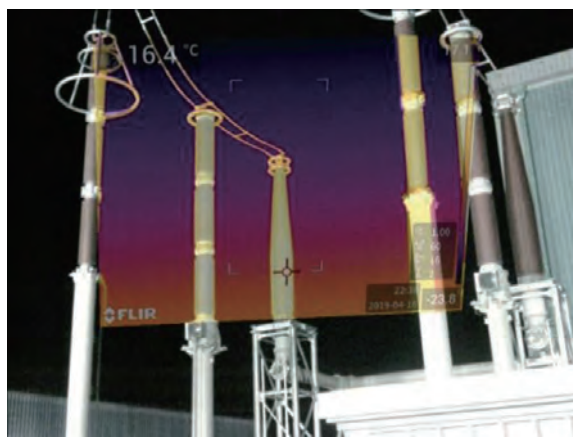


图1 可见光与红外图像融合效果

2 可见光与红外融合图像检测模型

2.1 网络整体结构

结合不同 YOLO 模型的特点,文中选定 YOLOv8n 作为基线模型。YOLOv8n 由特征提取网络(Backbone)、特征融合网络(Neck)和检测头(Head)三部分组成。特征提取网络采用跨阶段部分连接(cross stage partial connection, CSP) Darknet53 作为骨干网络,相较于之前的 Darknet53,其通过 CSP 优化了梯度流,提升了特征提取效率,同时减少了计算量。YOLOv8n 的主要创新点在于:1)使用 C2f 模块替代 YOLOv5 的 C3 模块,通过交叉卷积+双分支结构增强梯度流动,显著提升特征提取效率;2)移除 Focus 模块,采用更高效卷积层,在保持精度的同时,降低 30% 计算量;3) Neck 部分优化特征金字塔网络(feature pyramid networks,

FPN)与路径聚合网络(path aggregation networks, PAN)结构,通过双向特征聚合,提升多尺度检测能力,尤其增强小目标检测精度;4)摒弃锚框机制,直接预测目标中心点与宽高,结合解耦头设计(分类与回归分支分离),降低计算复杂度的同时,提升收敛速度。

为了解决融合图像中的颜色不连续性和目标检测难题,如图2所示,文中提出了一种改进YOLOv8n算法:1) 针对融合图像颜色不连续现象,以及缺乏视觉一致性,以通过增强通道、高度、宽度三个维度之间特征相互作用为中心,设计并引入了多维交互注意力模块;2) 为了避免过拟合,使用平滑技术对Slide Loss 损失函数进行改进,并替换YOLOv8n的CIoU Loss 损失函数。

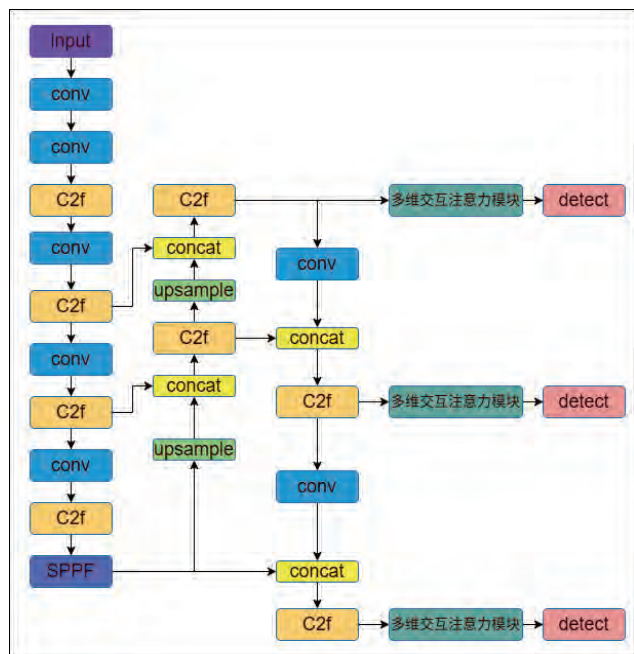


图2 改进的YOLOv8模型结构

2.2 多维交互注意力模块

用于融合图像的可见光图像与红外图像由于大小不同,且图像视觉信息密度不同,不可避免地造成了以下问题:1)可见光图像与红外图像融合交界处存在显著的颜色不连续现象,缺乏视觉一致性;2)虽然可见光图像与红外图像具有相同的检测目标,但由于大小的不一致性,导致被检测目标可能存在于融合图像处、可见光图像处、融合图像部分与可见光图像交界处,使图像中目标特征更加难以检测。

因此,文中改进了复合协同注意力机制^[10],通过对通道、高度、宽度三个维度同时增加权重,提高模型准确率;通过加入通道的特征信息、组归一化(group normalization, GN)层以及正则残差结构对模型进行修改创新;通过增强通道、高度、宽度三个维度之间特征相互作用,以提高模型对电力设备检测的准确率,如图3所示。

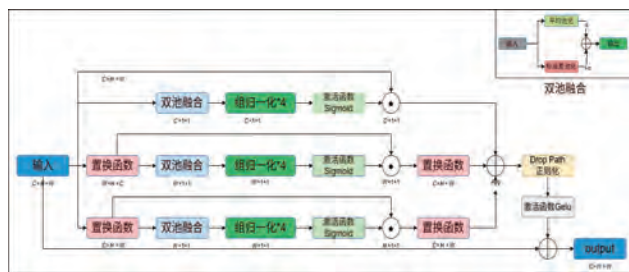


图3 多维交互注意力模块

为了更好地获得三个维度的权重特征,通过将输入的特征图 $F^{\text{input}}=C \times H \times W$ 复制为相同的四份,前三部分通过置换函数重新排列为通道特征图 $F^{\text{ch}}=C \times H \times W$ 、宽度特征图 $F^{\text{wi}}=W \times H \times C$ 、高度特征图 $F^{\text{he}}=H \times C \times W$ 三个特征,第四部分用于之后的残差结构。之后需要将 F^{ch} 、 F^{wi} 、 F^{he} 通过池化转为一维向量 F^{ch} 、 F^{wi} 、 F^{he} ,其中池化层为混合池化,通过平均池化与标准差池化的结合来更好地聚合跨维度特征响应。为更好地适应模型,文中使用双池化融合,即在同时使用两种池化层时,没有简单地将池化结果相加,而是采用了类残差结构的方式,将两种池化结果融合在一起,使其获得自适应机制,从而显著提高特征描述符的表示能力,其中输入特征图 $F^t=[f_1^t, f_2^t, \dots, f_n^t] t \in \{\text{ch}, \text{wi}, \text{he}\}$ 。池化后获得平均特征向量 $F^{t-\text{avg}}=[f_1^{t-\text{avg}}, f_2^{t-\text{avg}}, \dots, f_n^{t-\text{avg}}]$, 标准差池化向量 $F^{t-\text{std}}=[f_1^{t-\text{std}}, f_2^{t-\text{std}}, \dots, f_n^{t-\text{std}}]$ 。双池融合特征向量 $F^{t-\text{mean}}$ 平均池化如公式(1)所示,标准差池化如公式(2)-(3)所示,池化残差如公式(4)所示。

$$f_n^{t-\text{avg}} = \sum_{i=1}^H \sum_{j=1}^W f_n^t(i,j)/(H \times W) \quad (1)$$

$$S_n^t = \sum_{i=1}^H \sum_{j=1}^W (f_n^t(i,j) - f_n^{t-\text{avg}})^2 \quad (2)$$

$$f_n^{t-\text{std}} = \sqrt{S_n^t} / (H \times W) \quad (3)$$

$$F^{t-\text{mean}} = \alpha \times F^{t-\text{avg}} + (1 - \alpha) \times F^{t-\text{std}} \quad (4)$$

式(4)中的 $0 < \alpha < 1$ 是一个可训练的参数,通过随机梯度下降算法进行优化,这种自适应机制,在图

像特征提取的不同阶段,为平均池化与标准差池化分配不同的权重,以提高优化后续权重选择的能力。

使用4组GN层对 $F^{i \cdot \text{mean}}$ 进行处理,GN层可以对每个子特征进行独立归一化,而不会引入批次统计噪声,有效减轻子特征之间的语义干扰,同时稳定深层网络训练,通过一次Sigmoid激活函数后,分别与 F^{ch} 、 F^{wi} 、 F^{he} 点乘,将三个通道通过置换函数重新排列为 $C \times H \times W$ 的形式后,逐元素相加并除以3,使得输出与输入的特征图大小一致。

进行DropPath正则化处理,在训练过程中随机丢弃部分路径来防止过拟合,提高模型泛化能力;再通过GELU激活函数,在负值处仍然保留微小激活,缓解了神经元死亡问题,并且增强模型的非线性表达能力。为了缓解梯度消失问题,同时加速训练过程中的信息流通,再与四份特征中余下的 F^{input} 进行Shortcut连接,允许网络直接传递信息到输出,有助于缓解深度网络中梯度消失的问题,同时可以加速训练过程中的信息流动。

2.3 损失函数

在目标检测任务中,尤其是小目标检测领域,样本不平衡问题一直是研究的热点和难点。具体而言,易样本数量通常远多于难样本,这种不平衡现象在训练过程中可能导致模型对易样本过度拟合,而对难样本的识别能力不足。为了解决这一问题,已有研究提出了多种损失函数的改进方案,例如,Slide Loss^[11]通过计算预测框与真实框之间的交并比(intersection over union, IoU)来区分样本的难易程度,并以全部边界框IoU数值的平均值作为阈值 μ ,将样本分为难样本和易样本,并给予难样本更高的权重。然而,这种方法在训练初期可能会导致模型欠拟合,进而影响模型的整体精度和泛化能力。

Slide Loss是通过动态调整难样本的权重来优化模型的训练过程,但在实际应用中,该方法在训练初期可能会导致模型对难样本的过度关注,从而引发欠拟合问题,这主要是因为,在训练初期,模型的识别能力相对较弱,此时对难样本赋予过高权重,可能会使模型陷入局部最优,进而影响其对整体数据的拟合能力。因此,如何在训练初期平衡易样本和难样本的权重,同时避免模型过拟合,是解决样本不平衡问题的关键。

为了解决上述问题,文中使用了一种基于指数加权移动平均法(exponential moving average, EMA)思想的改进方案。EMA是一种常用的平滑技术,能够有效地减少数据的波动,从而提高模型的稳定性和泛化能力。文中将EMA应用于阈值 μ 的调整过程中,使其在训练初期维持在较高值,给予大多数样本相同的权重,这样让模型在训练初期能够充分学习到易样本的特征,避免因过早关注难样本而导致的欠拟合问题。随着模型的迭代和准确度的逐渐提高,阈值 μ 将逐渐降低,使模型能够逐渐关注到那些IoU小于阈值 μ 的难样本。这种动态调整策略不仅能够有效解决样本不平衡问题,还能提高模型的泛化能力和整体性能。损失函数如公式(5)–(7)所示。

$$f(x) = \begin{cases} 1 & x \leq \mu_i - 0.1 \\ e^{1-\mu_i} & \mu_i - 0.1 < x < \mu_i \\ e^{1-x} & x \geq \mu_i \end{cases} \quad (5)$$

$$\mu_i = d_i \times \mu_{\text{iou}} + (1 - d_i) \times \hat{\mu} \quad (6)$$

$$d_i = 1 - e^{-i} \quad (7)$$

式中: i 表示当前的迭代次数; $\hat{\mu}$ 是超参数,表示经验所得的平均IoU值; μ_{iou} 是所有边界框的IoU值的平均值。

3 改进的yolov8算法实验及结果分析

3.1 不同算法对比实验

文中使用的数据集为自建数据集,红外图像与可见光图像均为403张,训练集、验证集、测试集分别为281张、61张、61张,共8种标签,分别为电阻器(resistor)、断路器(interrupter)、隔离开关(disconnect_switch)、电抗器(reactor)、穿墙套管(wall_tube);均压环(equalizing_ring)、接地扁铁(Grounding_flat_iron)、接地刀闸(Grounding_switch)。

文中模型训练总批次为400,输入图片大小为640×640,工作进程为4。

为了验证文中所改进的YOLOv8n算法在可见光和红外融合图像目标检测任务中的性能,文中与多种改进的YOLO系列算法进行了对比实验。选取的对比算法包括YOLOv8n^[12]、YOLO-inner^[13]、YOLO-FCA^[14]、YOLO-rdnet^[15]和YOLO-BCN^[16]。为了更

准确地评估文中算法的性能,实验在特定数据集上进行,并采用精确率、召回率、mAP50 和 mAP50-95 四个评价指标,实验结果见表 1。

文中提出的改进算法在可见光和红外融合图像的目标检测任务中表现优异。与 YOLOv8n 对比,文中算法的精确率提升 0.7%,召回率提升 1.6%,mAP50 提升 0.9%,mAP50-95 提升 1.0%。综合比较各种算法,文中算法精确率与 YOLO-rdnet 持平,同为最高;召回率仅低于最高的 YOLO-inner1.3%;mAP50 和 mAP50-95 均为最高。总体而言,改进的算法在 mAP50 和 mAP50-95 的性能上均优于对比算法,综合性能表现优异,能够有效提升可见光和红外融合图像目标检测的精度。

表 1 不同模型试验对比

算法	精确率	召回率	mAP50	mAP50-95
YOLOv8n	88.6	75.5	83.4	59.1
YOLO-inner	85.9	78.4	83.5	57.9
YOLO-FCA	87.3	75.6	82.5	56.8
YOLO-rdnet	89.3	65.9	76.2	47.6
YOLO-BCN	86.3	72.4	81.4	57.8
文中算法	89.3	77.1	84.3	60.1

3.2 不同算法的可视化分析

为了更直观地分析文中所提改进 YOLOv8n 算法的检测效果,选取 2 张同时包含目标物体及其关键特征的融合图像,这 2 张图像受到了光照不均、目标尺寸较小、图像噪声等因素的干扰。不同算法检测结果的可视化效果如图 4、图 5 所示。

由图 4、图 5 可知,YOLO-rdnet 和 YOLO-FCA 复杂背景下对部分目标物体检测效果差,其中 YOLO-FCA 在图 4 中错误地检测出了 equalizing ring 设备,图 4、图 5 中 YOLO-rdnet 分别重复检测了 reactor 和 wall_tube 设备,两者检测效果均不理想。YOLO-inner、YOLO-BCN 和 YOLOv8n 虽然检测出了大部分目标物体及其关键特征,但对于小尺寸目标存在漏检,检测效果有待提高。而文中算法检测出了所有的目标物体及其关键特征,且置信度整体较高,检测效果最佳。

由表 1 结果与图 4、图 5 检测结果可知,文中提出的算法模型在检测精度上较其他模型有显著提升,尤其在高 IoU 阈值范围内的性能表现更为突出。这表

明,文中提出的模型在复杂场景下对目标的识别与定位更为精准,能够有效减少误检与漏检情况。

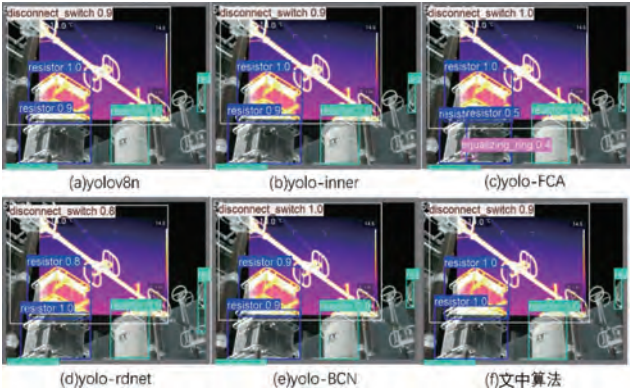


图 4 融合图像中不同算法目标检测结果对比图 1

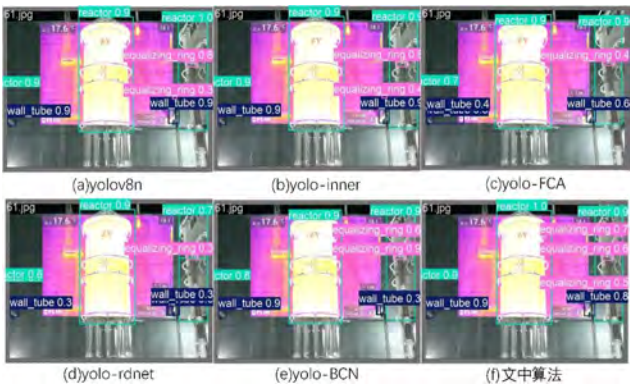


图 5 融合图像中不同算法目标检测结果对比图 2

4 结语

文中针对电力红外图像目标检测任务,提出了一种面向电力设备的多维交互注意力增强 YOLOv8n 多模态检测算法。通过 CAO-C2F 配准与加权融合技术,实现了可见光图像与红外图像的精准对齐和有效融合。在此基础上,设计的改进注意力机制模块进一步增强了模型对融合图像特征的提取能力,特别是在处理融合图像交界处颜色不连续和目标检测困难等问题时表现出色。实验结果表明,文中提出的改进 YOLOv8n 算法在可见光和红外融合图像目标检测任务中表现优异,与多种改进的 YOLO 系列算法相比,在精确率、召回率和高 IoU 阈值范围内的性能上均优于对比算法,有效提升了目标检测的精度。

尽管文中方法在电力红外图像目标检测任务中取得了显著的性能提升,但仍存在一些可以进一步改

进的地方。例如,当前的融合方法主要依赖于加权求和策略,未来可以探索更先进的融合策略,以更好地保留两类图像的互补信息。此外,虽然改进的注意力机制模块在一定程度上解决了融合图像的特征提取问题,但在面对更复杂的场景时,仍可能面临挑战。因此,进一步优化注意力机制模块的设计,提高模型对复杂场景的适应能力,也是未来工作的方向之一。

参考文献:

- [1] 李淑慧,蔡伟,王鑫,等.深度学习框架下的红外与可见光图像融合方法综述[J/OL].计算机工程与应用,2025: 1-19[2025-04-22]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20250122.1601.007.html>.
- [2] 朱自文,宋晓鸥,崔巍,等.可见光-红外图像融合的目标检测综述[J/OL].计算机工程与应用,2025: 1-26[2025-04-22]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20250321.1453.006.html>.
- [3] 贺强,刘洋,游鑫,等.基于改进YOLOv9的电力设备红外目标检测模型[J/OL].计算机技术与发展,2025: 1-9[2025-04-22]. <https://doi.org/10.20165/j.cnki.ISSN1673-629X.2025.0112>.
- [4] 杨达伟,杨明圣,付博.改进YOLOv7的电力设备红外图像多目标检测[J].红外技术,2025,47(3):326-334.
- [5] 王秋忆,周浩,郑婷婷.改进RetinaNet的电力设备目标检测方法[J].计算机与现代化,2024(1):47-52,58.
- [6] 屈璐瑶.基于多尺度特征融合与特征细化的电力设备目标检测研究[D].天津:天津大学,2023.
- [7] 赵杰,陈嘉晋.变电站低照度场景红外可见光图像融合[J].电气技术,2025,26(3):22-29.
- [8] 陈海秀,卢海涛,房威志,等.基于多重特征的红外与可见光图像融合算法[J/OL].电光与控制,2025: 1-8[2025-04-22]. <http://kns.cnki.net/kcms/detail/41.1227.TN.20250320.0952.002.html>.
- [9] Jiang Q, Liu Y, Yan Y, et al. A contour angle orientation for power equipment infrared and visible image registration[J]. IEEE Transactions on Power Delivery, 2020, 36(4): 2559-2569.
- [10] Yang Yu, Yi Zhang, Zeyu Cheng, et al. MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition[J]. Engineering Applications of Artificial Intelligence, 2023, 126(Part C): 107079.
- [11] Ziping Yu, Hongbo Huang, Weijun Chen, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. Pattern Recognition, 2024, (155): 110714.
- [12] Yaseen M. What is YOLOv8: An in-depth exploration of the internal features of the next-generation object detector[DB/OL]. (2024-08-28) [2025-04-22]. <https://arxiv.org/abs/2408.15857>.
- [13] Zhang H, Xu C, Zhang S. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[DB/OL]. (2023-11-14) [2025-04-22]. <https://arxiv.org/abs/2311.02877>.
- [14] Wang J, Xu C, Yang W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[DB/OL]. (2022-06-14) [2025-04-22]. <https://arxiv.org/abs/2110.13389>.
- [15] Kim D, Heo B, Han D. DenseNets reloaded: paradigm shift beyond ResNets and ViTs[C]//European Conference on Computer Vision. Cham: Springer, 2024: 395-415.
- [16] Khaled A. BCN: batch channel normalization for image classification[C]//International Conference on Pattern Recognition. Cham: Springer, 2025: 295-308.