

基于深度学习的电力缺陷检测研究综述

张朔华

北京印刷学院信息工程学院, 北京

收稿日期: 2026年7月18日; 录用日期: 2026年8月19日; 发布日期: 2026年8月27日

摘 要

电力设备缺陷检测是输电线路、变电站和新能源设备智能运维中的重要任务。随着无人机、巡检机器人、可见光相机、红外热像仪等设备的应用, 电力巡检图像数据快速增长, 推动了基于深度学习的自动检测方法发展。本文围绕电力缺陷检测对象与任务特点, 梳理两阶段目标检测、一阶段目标检测、Transformer视觉模型和多模态融合技术在电力缺陷检测中的研究进展。现有研究表明, 两阶段方法在定位精度和区域分析方面具有优势, 但推理速度和部署复杂度较高; YOLO等一阶段方法具有实时性和工程部署优势, 但在小目标、复杂背景和全局语义建模方面仍存在不足; Transformer能够增强长距离依赖和高层语义表达, 多模态融合则为图像、文本、红外和点云信息协同利用提供了新的方向。最后总结当前研究存在的问题及未来的发展趋势。

关键词

电力缺陷检测, YOLO, 目标检测, Vision Transformer, 多模态融合

A Review of Deep Learning-Based Power Defect Detection

Shuohua Zhang

School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: July 18, 2026; accepted: August 19, 2026; published: August 27, 2026

Abstract

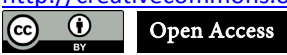
Power equipment defect detection is an important task in intelligent operation and maintenance of transmission lines, substations and photovoltaic facilities. With the increasing use of UAVs, inspection robots, visible cameras and infrared sensors, massive inspection images have promoted the development of deep learning-based automatic detection methods. This review summarizes the detection objects and task characteristics of power defects, and discusses two-stage detectors, one-

stage detectors represented by the YOLO series, visual Transformer models and multimodal fusion techniques. Existing studies show that two-stage methods are advantageous in localization and region-level analysis but are relatively complex for real-time deployment. One-stage YOLO-based methods are efficient and deployment-friendly, while small defects, cluttered backgrounds and limited global semantic modeling remain challenging. Transformer models provide long-range dependency modeling, and multimodal fusion offers a possible way to combine visual, textual, infrared and point-cloud information. Finally, summarize the existing problems of current research and future development trends.

Keywords

Power Defect Detection, YOLO, Object Detection, Vision Transformer, Multimodal Fusion

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



1. 引言

电力设备长期运行于户外复杂环境中，容易受到风沙、覆冰、雷击、污秽、机械振动和温度变化等因素影响[1]。绝缘子破损、金具锈蚀、导线断股、异物悬挂和变电设备表面异常如果不能及时发现，可能导致闪络、短路或停电事故。因此，缺陷检测不仅关系到电网安全稳定运行，也关系到巡检效率、运维成本和作业人员安全。本文从电力缺陷检测对象与任务特点出发，重点介绍了基于深度学习算法的电力缺陷检测研究现状。最后总结了该领域存在的主要问题和挑战，以及未来的发展趋势。

从研究脉络看，电力缺陷检测经历了由人工判读、传统图像处理、深度学习目标检测到多模态智能感知的转变。早期方法多依赖阈值、边缘和纹理等人工特征；深度学习阶段则以卷积神经网络、两阶段检测器和单阶段检测器为核心；近年来，视觉 Transformer 和视觉语言模型又进一步推动了全局语义建模和多模态融合研究。

2. 电力缺陷检测对象与任务特点

随着无人机、巡检机器人和固定摄像头的应用，电力系统积累了大量可见光、红外和视频数据，为基于计算机视觉的自动检测提供了基础。电力缺陷检测对象具有明显的设备类型差异。输电线路巡检主要关注绝缘子、导线、金具、杆塔及通道环境等对象，常见的电力线路巡检对象与任务特点如表 1 所示。

Table 1. Characteristics of common equipment defects and inspection tasks in power line inspection

表 1. 电力巡检中常见设备缺陷及检测任务特点

设备	典型缺陷	任务特点	主要难点
绝缘子	破损、闪络、掉串、污秽、裂纹	缺陷类型多，既包含明显结构性缺陷，也包含表面纹理类缺陷	破损、裂纹等缺陷尺度较小；背景复杂时容易出现漏检和误检
导线	断股、磨损、异物缠绕、弧垂异常	多为细长目标，缺陷沿导线分布	导线在图像中占比小，断股、磨损等缺陷细节不明显
金具	锈蚀、变形、缺失	金具结构复杂，缺陷通常表现为局部颜色、形状或结构变化	锈蚀与阴影、污渍相似；小部件缺失不易发现；不同拍摄角度下形态变化较大
防震锤	移位、脱落、锈蚀	目标尺寸较小，通常依附于导线附件，需要判断其位置状态是否正常	在远距离巡检图像中分辨率低，移位和脱落需要结合导线位置进行空间判断

续表

异物	塑料薄膜、鸟巢、 风筝、广告布	目标形态不规则，颜色、大小、位置 差异明显	类别变化大，形态无固定模式；小尺寸 异物容易漏检
均压环	倾斜、脱落、锈蚀	具有较明显的环状结构，缺陷多与 位置、姿态和表面状态有关	倾斜、脱落需要准确识别结构位置关系； 锈蚀区域可能较小
杆塔本体	变形、裂纹、锈蚀、 螺栓松动或脱落	目标结构复杂，既包含大尺度结构 异常，也包含螺栓等小尺度部件异常	杆塔背景通常复杂，构件交错严重； 结构遮挡和拍摄角度变化影响较大

电力巡检图像具有尺度变化大、背景复杂和缺陷样本稀缺等特点。无人机拍摄时距离、角度和光照变化明显，同一设备可能在图像中呈现不同尺度；绝缘子掉串、闪络、裂纹和导线断股等缺陷往往面积较小，与设备纹理或背景结构相近，容易被天空、树木、杆塔和导线交叠区域干扰。电力缺陷检测不仅要求“看见目标”，还要求“判断状态”。例如绝缘子检测可以只定位绝缘子本体，也可以进一步识别破损、烧蚀或掉串等状态。因此，检测模型需要兼顾部件定位、缺陷定位和缺陷语义解释。

3. 基于深度学习的电力缺陷检测方法

深度学习目标检测方法通过端到端的特征学习机制，减少了对人工特征设计和规则阈值设定的依赖，使电力缺陷检测逐步由传统图像处理方法转向数据驱动的自动识别方法。根据检测流程的不同，现有方法通常可分为两阶段目标检测算法和一阶段目标检测算法。其中，两阶段方法通常具有较高的检测精度，但模型结构相对复杂、推理速度较慢；一阶段方法以 SSD、YOLO 系列算法为代表，具有检测速度快、端到端训练和工程部署便利等优势，因而在电力巡检场景中得到广泛关注。随着视觉模型的发展，Transformer [2] 和 Vision Transformer (ViT) [3] 等方法被引入缺陷检测任务，用于增强模型的全局上下文建模能力；同时多模态融合方法也逐渐应用于电力缺陷检测，通过融合可见光图像、红外图像、文本描述或巡检信息等多源数据，提高模型对复杂缺陷场景的理解能力。本文将从两阶段目标检测算法、一阶段目标检测算法、Transformer 视觉模型以及多模态融合方法四个方面对相关研究进行综述。

3.1. 基于两阶段目标检测算法

两阶段目标检测算法通常将检测过程划分为候选区域生成和目标分类回归两个阶段，R-CNN [4] [5] 系列算法就是典型的两阶段检测算法。其基本思想是先通过区域建议网络或目标定位模块生成可能包含目标的候选区域，再对候选区域进行精细分类、边界框回归或进一步的缺陷分析。两阶段方法的优势在于候选区域分析较精细，适合缺陷区域边界不清、需要进一步分类判断的任务，但其推理流程相对复杂，在无人机巡检实时处理和边缘端部署中存在一定限制。

Faster R-CNN [6] 是典型的两阶段目标检测算法，其核心结构包括特征提取网络、区域建议网络和检测分类网络。如图 1 所示，输入图像经过卷积神经网络提取特征信息后生成共享特征图；随后，RPN 基于 Anchors 生成候选区域，并通过分类分支和回归分支对候选框进行筛选与优化；候选区域经过 RoI Pooling 将不同尺寸的目标区域转换为固定尺寸特征后输入到 RoI Head，通过分类分支和边界框回归分支完成目标类别预测和位置精修，最终输出目标类别和位置。对于电力缺陷检测而言，该方法能够先从复杂场景中筛选出疑似目标区域，再进行精细识别，有助于降低背景冗余信息对检测结果的干扰。

Ling 等[7]针对航拍图像中的自爆玻璃绝缘子定位问题，提出了 Faster R-CNN 与 U-Net 相结合的检测方法。该方法首先利用 Faster R-CNN 定位绝缘子串，并裁剪出目标区域，以减少复杂背景干扰；随后利用 U-Net 对裁剪后的绝缘子图像进行像素级分割，从而识别自爆破损区域。该方法将自爆绝缘子检测转化为“绝缘子串定位”和“破损区域分割”两个子任务，能够有效缓解航拍图像中目标占比小、信噪比较低的问题。

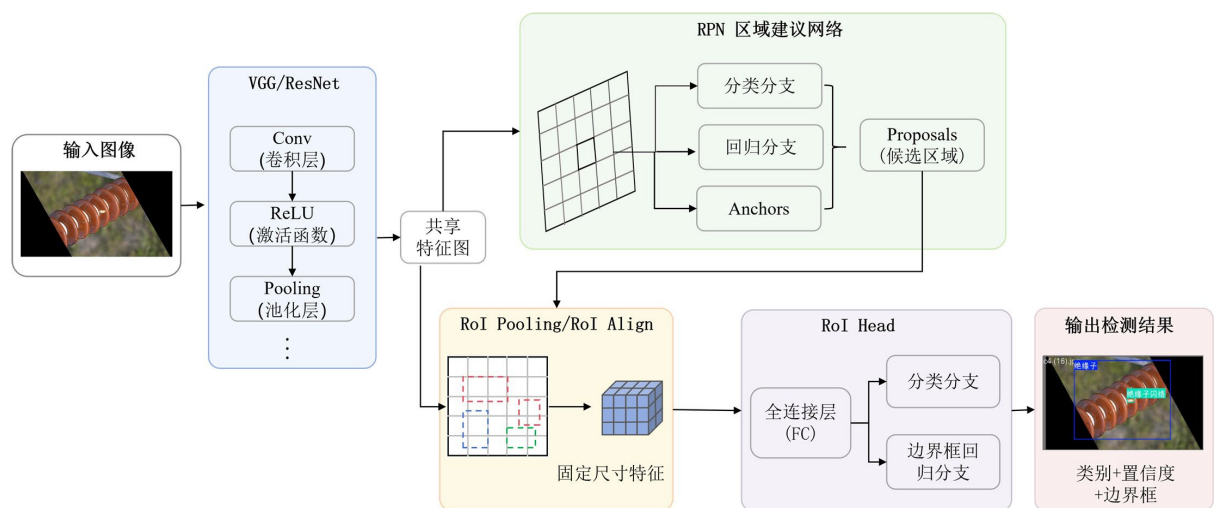


Figure 1. Architecture of Faster R-CNN target detection
图 1. Faster R-CNN 目标检测结构图

3.2. 基于一阶段目标检测方法

与两阶段目标检测算法需要先生成候选区域再进行分类回归不同，一阶段目标检测算法直接在输入图像上进行目标类别预测和位置回归，通过单个网络完成从特征提取到检测结果输出的全过程。该类方法省去了候选区域生成步骤，在保证较高检测精度的同时显著提升了推理速度，具有结构简单、计算效率高和易于工程部署等特点，因此更加适用于无人机巡检、在线监测等对实时性要求较高的电力缺陷检测场景。一阶段目标检测算法主要包括 SSD [8]、YOLO (You Only Look Once) [9]-[11] 系列等。

在 SSD 系列算法的应用方面，相关研究主要围绕多尺度特征提取、默认框优化、样本扩充和主干网络改进等方向展开。徐海洋等[12]针对变电站巡检图像中不同设备尺寸差异显著的问题，提出了一种改进 SSD 的变电站电力设备识别方法。该方法主要对两个方面进行了改进，一是加强多尺度特征提取，提高大尺度特征图的有效性；二是优化默认框生成策略，使默认框更加符合变电站设备尺度差异较大的特点。同时，该研究还结合迁移学习和数据增强技术提升模型鲁棒性，实验结果表明，相较于传统 SSD 算法，其召回率提升了 42.24%，能够更好地解决变电站场景中大、小尺寸设备同时识别的问题。尤振飞等[13]针对电力巡检图像中绝缘子定位准确性受训练图像数量和质量限制的问题，提出了基于风格迁移和改进 SSD 的绝缘子定位方法。该方法首先利用风格迁移进行样本扩充，并通过改进 LeNet-5 对图像质量进行评定和清洗，以提高训练数据质量；随后通过融合原 SSD 算法前三个尺度的检测特征图，并利用特征金字塔结构进行特征重构，从而增强模型对局部细节特征和全局语义特征的提取能力，提高绝缘子定位精度。

YOLO 系列算法采用端到端回归思想，将目标分类与边界框定位任务统一到单个网络中完成，在保证较高检测精度的同时具有更快的推理速度，因此更加适用于无人机巡检和边缘设备部署等实时检测场景。表 2 列举了近几年 YOLO 系列算法在电力缺陷检测中的具体应用。

Chen 等[14]针对无人机巡检过程中图像分辨率低、目标特征不明显的问题，提出了一种结合超分辨率重建与 YOLOv3 的电力组件检测方法。该方法利用 SRCNN 对低质量巡检图像进行增强，提高输入图像的细节信息，再通过 YOLOv3 完成电力组件定位和识别，有效提升了复杂巡检环境下目标检测的准确性和鲁棒性。

针对输电线路关键部件检测过程中存在的小目标漏检和复杂背景干扰问题，Tu Renwei 等[15]提出了

一种改进 YOLOv3 的输电线路关键部件检测方法。该方法通过调整 YOLOv3 多尺度检测结构,将原有检测层进行优化,同时利用 K-means++对目标框尺寸重新聚类获得更加适合电力设备的数据先验信息,并结合旋转、缩放以及光照变化等数据增强策略提升模型泛化能力。在轻量化 YOLO 模型方面, Liu 等[16]针对复杂航拍图像中的绝缘子检测任务提出了 MTI-YOLO 模型。该方法以 YOLO-tiny 为基础,面向绝缘子检测中复杂背景干扰、不同尺度目标识别困难以及嵌入式部署资源受限等问题进行改进。首先,研究者构建了 CCIN_detection 复合绝缘子数据集;其次,在 MTI-YOLO 网络中引入多尺度检测头、多尺度特征融合结构和空间金字塔池化模块,以增强模型对不同尺寸绝缘子的特征表达能力。实验结果表明,MTI-YOLO 的平均精度分别比 YOLO-tiny 和 YOLOv2 提高 17%和 9%,同时其内存占用分别比 YOLOv2 和 YOLOv3 降低 25.6%和 38.9%,在复杂背景和强光照条件下均表现出较好的检测效果。该研究说明,针对电力巡检场景中的绝缘子目标,轻量化 YOLO 模型可以通过多尺度特征融合和 SPP 结构,在检测精度、运行速度和存储开销之间取得较好平衡。

Table 2. Research on applications of YOLO series algorithms in power defect detection

表 2. YOLO 系列算法在电力缺陷检测中的应用研究

文献	模型	检测对象	改进目的	主要改进	实验结果	算法优势
[14]	SRCNN-YOLOv3	电力组件	提升低分辨率图像下的目标识别能力	引入 SRCNN 超分辨率重建方法,对低分辨率巡检图像进行增强	准确度比 Faster R-CNN 和 SDD 高 1%~3%	增强低分辨率图像细节表达,适合图像质量较差的巡检场景
[15]	改进 YOLOv3	输电线路关键部件	改善小目标漏检和复杂背景干扰问题	采用 K-means++重新聚类 Anchor,并结合旋转、缩放、光照变化等数据增强策略	Precision 比 YOLOv3 提高 6%,计算量减少 8.49%,检测速度提升约 6%	提高模型对输电线路关键部件尺度变化和复杂背景的适应能力
[16]	MTI-YOLO	绝缘子检测	再精度、速度和开销之间取得平衡	引入多尺度检测头、多尺度特征融合结构和 SPP 模块	AP 分别比 YOLO-tiny 和 YOLOv2 提高 17%和 9%	保持轻量化优势,提高不同尺度绝缘子的检测效果
[17]	Insu-YOLO	无人机绝缘子缺陷	提升微小绝缘子缺陷检测能力	引入 GSConv 降低参数量,引入 CARAFE 特征重组模块和增加小目标检测层	相比 YOLOv8n, mAP 提升 3.95%,内存使用量为 9.2MB	降低模型参数量并增强多尺度特征融合能力,
[18]	PowerStrand-YOLO	导线断股、松股	提高细长型导线缺陷的检测精度和定位能力	引入 C2f_DCNv4、EMA 注意力机制和 Shape-IoU 损失函数	Precision 为 95.4%, Recall 为 96.2%,速度达到 250 帧/s	适合细长目标和复杂航拍背景下的实时监测
[19]	RD-YOLO	电力部件锈蚀缺陷	提升不规则锈蚀区域和难样本检测能力	引入 CCARM、RFB、ECBAM、对比样本策略和 Focal Loss	平均精度达到 95%,比原始 YOLOX 提高 9%	强化锈蚀缺陷的多尺度特征提取和难样本学习能力
[20]	YOLO-IASE	变电站火焰、安全帽	提高复杂变电环境下不规则目标检测鲁棒性	基于 YOLO 改进框架,引入智能自适应场景增强、CCBS 模块和 δ -EIoU 损失函数	Recall 相较于 YOLOv5s 提升 12.1%	对火焰形态变化、光照变化和遮挡目标具有较强适应性
[21]	SDA-YOLO	绝缘子自爆缺陷	提升小尺寸自爆缺陷识别能力	基于 YOLO11s,增加小目标检测层,引入 DBB 多分支结构,并采用 AIFI 模块增强特征表达	Precision 提升 6.6%,速度保持 93.6 帧/s	兼顾小目标检测精度和实时检测需求
[22]	DB-YOLO	绝缘子多尺度缺陷	解决大、小尺寸缺陷同时存在的问题	设计大小缺陷双分支结构,引入 Mamba 机制、ShuffleNetV2 轻量化结构	mAP50 达到 94.4%,比 YOLO11n 提升了 2.3%, Recall 提升了 0.7%	增强大尺度和小尺度缺陷特征表达,并提高推理效率

针对无人机绝缘子缺陷检测中存在的小目标缺陷难以识别的问题,Chen 等[17]提出了基于 YOLOv8 的 Insu-YOLO 模型。该方法在 YOLOv8 基础上引入 GSConv 轻量化卷积降低模型参数量,同时采用 CARAFE 特征重组模块优化特征上采样过程,提高不同尺度特征之间的信息交互能力。此外,通过增加小目标检测层,加强浅层细节特征与深层语义特征融合,提高了微小缺陷目标的检测能力。实验结果显示,Insu-YOLO 相比 YOLOv8n 基准模型 mAP 提升 3.95%,同时保持较高检测速度,实现了无人机巡检条件下绝缘子缺陷的实时检测。

针对输电导线断股、松股等细长型缺陷目标难以识别的问题,Deng 等[18]提出了 PowerStrand-YOLO。该方法以 YOLOv8 为基础,面向无人机航拍场景下高压输电导线缺陷检测任务进行改进。由于输电导线在图像中像素占比较小,且容易受到复杂背景干扰,传统检测模型在准确率和推理效率方面仍存在不足。PowerStrand-YOLO 通过引入 C2f_DCNv4 模块增强多尺度特征提取能力,利用 EMA 注意力机制抑制背景干扰并突出缺陷相关特征,同时采用 Shape-IoU 损失函数增强模型对导线异常形态和空间结构的定位能力。实验结果显示,该方法在导线缺陷检测中取得了 95.4%的 precision、96.2%的 recall 和 250 FPS 的检测速度,相比 YOLOv8 基线模型在精度和召回率方面均有所提升,体现出较好的实时检测与工程应用价值。

针对输电线路电力部件锈蚀缺陷形态不规则、尺度变化大以及样本数量有限等问题,Liu 等[19]提出了 RD-YOLO 模型。该方法以 YOLOX 为基础,首先设计 Coordinate Channel Attention Residual Module (CCARM)增强主干网络的多尺度特征提取能力;其次,在 PANet 中引入 Receptive Field Block (RFB)和 Efficient Convolutional Block Attention Module (ECBAM),加强深层语义特征与浅层空间特征之间的融合;最后,采用对比样本策略和 Focal Loss 对模型进行训练优化,以缓解锈蚀缺陷样本少、缺陷特征不明显和难样本学习不足等问题。实验结果表明,RD-YOLO 在自建数据集上的锈蚀缺陷检测平均精度达到 95%,比原始 YOLOX 提高 9%,在电力部件识别和锈蚀缺陷检测中表现出较好的应用前景。

在变电站安全监测场景中,Jiang 等[20]提出了 YOLO-IASE 模型,用于复杂变电站环境下的火焰和安全帽检测。该研究指出,现有 YOLOv5、YOLOv8 和 CBAM-YOLOv5 等方法虽然在实时检测方面取得了一定进展,但在复杂变电站环境中仍存在自适应数据增强不足、对不规则目标鲁棒性较弱等问题,难以稳定应对火焰形态变化、光照变化以及小目标遮挡等情况。为此,YOLO-IASE 引入智能自适应场景增强方法,对实时训练数据进行动态增强,以提升模型对复杂环境的适应能力;同时设计结合 CoordConv 的 CCBS 模块,以增强空间感知能力和目标定位精度,并提出 δ -EIoU 损失函数,降低模型对极端长宽比目标的敏感性。实验结果表明,该模型在变电站火焰和安全帽检测中具有较好的精度、鲁棒性和实时处理能力,其中火焰检测召回率提升 12.1%。

近年来,随着 YOLO11 模型的提出,研究进一步关注复杂场景下的小目标缺陷检测能力以及模型轻量化部署问题。Yang 等[21]针对绝缘子自爆缺陷尺寸小、背景干扰严重的问题,提出了一种基于 YOLO11s 的 SDA-YOLO 模型。该方法增加小目标检测层,利用浅层和深层特征融合提升模型对微小缺陷的感知能力;同时引入 DBB 多分支卷积结构增强不同尺度目标特征提取能力,并采用 AIFI 模块替换原有注意力结构,提高关键区域特征表达能力。实验结果表明,SDA-YOLO 相比 YOLO11s 基准模型准确率提升 6.6%,同时保持 93.6 帧/s 的实时检测速度。

针对绝缘子缺陷检测中不同尺度目标同时存在的问题,Fan 等[22]进一步提出基于 YOLO11 的 DB-YOLO 模型。该方法设计双分支检测结构,分别针对大尺度缺陷和小尺度缺陷进行特征学习,从而缓解传统检测网络对于不同尺度目标关注不足的问题。同时,该方法引入 Mamba 机制增强复杂背景下的全局特征建模能力,并利用轻量化 ShuffleNetV2 替换部分网络结构,提高模型推理效率。该研究表明,通过多分支结构与全局特征建模方法结合,可以进一步提升 YOLO 网络在复杂电力巡检环境中的

缺陷识别能力。

针对输电线路无人机巡检图像中小目标、复杂背景和遮挡等问题, Lu 等[23]提出了 AIR-YOLO 模型。该模型面向复杂背景下的输电线路巡检任务, 以 YOLO11 为基础进行系统级优化: 一是设计 AIFI 模块替代传统 SPPF, 通过多头自注意力增强跨尺度语义交互; 二是构建轻量化 ASF-Small Neck, 引入高分辨率 P2 特征层以保留小缺陷的细粒度信息; 三是在检测头中嵌入空间与通道混合 SEAM 注意力机制, 提高遮挡情况下的特征重标定能力; 四是采用面向细长结构的 PowerIoU 损失函数, 以提升绝缘子、防震锤等高长宽比目标的定位精度。该方法在检测精度、计算效率和模型复杂度之间取得较好平衡, 适合部署于电力巡检边缘设备。

3.3. Transformer 电力缺陷检测中的应用

近年来, 随着 Transformer 在计算机视觉领域的发展, 其基于自注意力机制的全局特征建模能力逐渐被应用于电力缺陷检测任务。相比传统卷积神经网络主要依赖局部感受野提取特征, Transformer 能够建立图像不同区域之间的长距离依赖关系, 更好地捕获设备结构信息和缺陷区域的上下文语义特征, 因此在复杂背景、小目标缺陷以及细粒度识别任务中具有较大的应用潜力。

针对输电线路巡检过程中背景干扰强、缺陷区域难以定位的问题, 李帷韬等[24]提出了一种结合强化学习与 Transformer 的输电线路缺陷检测方法。该方法首先利用 DetNet 提取图像特征, 并通过深度 Q 网络筛选包含有效缺陷信息的区域, 再利用 Transformer 编码器进行全局特征建模, 从而增强模型对复杂场景下缺陷目标的识别能力。实验结果表明, 该方法在输电线路缺陷航拍图像检测任务中取得了 89.7% 的平均检测精度, 具有较好的检测效果和泛化能力。

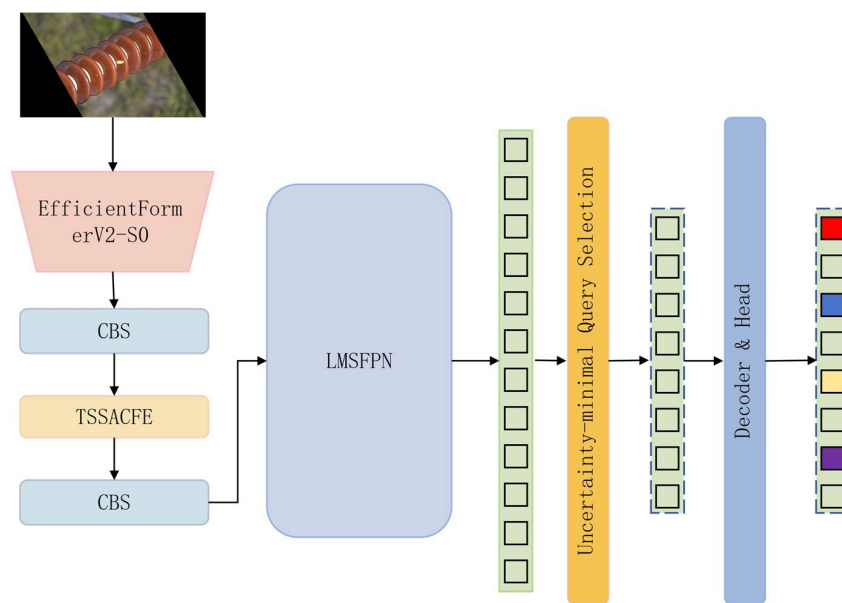


Figure 2. ELT-RTDETR network structure

图 2. ELT-RTDETR 网络结构

刘梓良等[25]提出了一种轻量化 ELT-RTDETR 模型, 如图 2 所示。通过引入 EfficientFormerV2 轻量化主干网络、LMSFPN 多尺度特征融合结构以及基于 TSSA 注意力机制的特征增强模块, 在降低计算复杂度的同时提升了模型对微小缺陷的检测能力。实验结果显示, 该方法在变电站缺陷数据集上的检测精度达到 82.1%, 相比传统 RT-DETR 具有更好的精度和部署效率。

3.4. 多模态电力缺陷检测研究

随着深度学习目标检测技术的发展，基于单一视觉信息的电力缺陷检测方法已经取得了较好的应用效果，但在实际巡检环境中仍存在一定局限性。一方面，无人机拍摄图像容易受到光照变化、天气条件、背景遮挡以及拍摄角度等因素影响，使部分缺陷区域表现出特征弱、尺度小和纹理不明显等问题；另一方面，电力设备缺陷通常具有较强的领域语义属性，仅依靠图像中的颜色、纹理和结构信息难以充分表达设备状态及缺陷含义。因此，近年来研究人员逐渐将多模态学习方法引入电力缺陷检测任务，通过融合不同来源的信息，实现视觉特征与语义信息的互补，提高模型对复杂场景和细粒度缺陷的识别能力。

Table 3. Applications of multimodal information in power defect detection
表 3. 多模态信息在电力缺陷检测中的应用

文献	模态类型	融合层次	方法/模型	主要方法设计	解决的问题	算法优势
[26]	可见光图像 + 红外图像	特征级融合/中期融合	CWSP-YOLO	1. 基于 YOLOv11 构建可见光与红外双分支网络； 2. 引入 Cross-Modality Fusion Transformer 模块； 3. 结合注意力机制和改进损失函数优化性能	解决单一视觉模态受光照、背景干扰影响，导致绝缘子缺陷误检和漏检的问题	利用可见光纹理信息与红外热信息互补，增强复杂环境下绝缘子缺陷区域的特征表达能力
[27]	图像 + 文本	语义级融合/视觉语言对齐	电力缺陷视觉语言大模型	1. 利用 YOLOv8 自动构建图像、提示信息 and 标签组成的多模态数据集； 2. 采用 LoRA 进行参数高效微调； 3. 引入 Q-Former 模块，实现视觉特征与文本语义信息对齐	解决传统检测模型语义理解能力不足、人工标注成本高的问题	通过图像 - 文本交互辅助完成缺陷识别、状态分析和问答，提高模型的电力领域语义理解能力
[28]	图像 + 文本 + 领域知识	知识级融合/空间推理融合	视觉语言多模态大模型	1. 利用 Swin Transformer 提取多尺度视觉特征； 2. 通过大语言模型学习电力领域专业知识； 3. 采用对比学习实现图像特征与文本描述对齐； 4. 引入区域级指令微调，增强空间关系推理能力	解决无人机巡检中缺陷样本不足、小样本弱纹理以及复杂背景干扰问题	不仅能够识别缺陷目标，还能够结合设备空间关系和领域知识进行推理，提升复杂巡检场景下的智能理解能力

目前，多模态电力缺陷检测主要包括两类研究方向：一类是融合可见光、红外等不同传感器数据，通过多源视觉信息增强缺陷特征表达能力；另一类是融合图像与文本信息，利用视觉语言模型实现缺陷场景理解和语义推理。相关代表性研究如表 3 所示。

多模态方法能够利用不同信息源之间的互补性，提高电力缺陷检测模型的环境适应能力。在多源视觉信息融合方面，Du 等[26]针对无人机绝缘子缺陷检测中单一可见光图像容易受到光照和背景干扰的问题，提出了一种基于 YOLOv11 的多模态检测方法 CWSP-YOLO。该方法构建可见光与红外图像双分支网络，通过 Cross-Modality Fusion Transformer 模块实现不同模态特征之间的交互融合，利用 Transformer 的全局建模能力增强缺陷区域特征表达，同时结合注意力机制和改进损失函数优化检测性能。在视觉与语言信息融合方面，Wang 等[27]提出了一种面向电力缺陷场景的图像 - 文本多模态大模型，通过 YOLOv8 自动构建包含图像、提示信息和标签的多模态数据集，并结合 LoRA 参数高效微调方法和 Q-Former 模块，实现视觉特征与语言语义信息的有效对齐。该方法能够通过图像文本交互辅助电力人员完成场景识别、缺陷分析以及知识问答，提高了模型在专业电力场景中的理解能力。

为解决无人机巡检中缺陷样本稀缺、复杂背景干扰以及现场实时识别需求等问题，Dong 等[28]提出

了一种基于视觉语言多模态大模型和空间推理学习的设备缺陷识别方法。该方法利用 Swin Transformer 提取图像中的多尺度视觉特征，同时通过大语言模型学习电力领域专业知识，并利用对比学习实现图像特征与文本描述之间的语义对齐。此外，该方法进一步引入区域级指令微调，使模型能够结合设备空间关系进行多步推理，从而增强复杂巡检场景下的缺陷识别能力。实验结果表明，该方法能够有效提升复杂背景和小目标缺陷情况下的检测性能。

4. 总结

本文围绕基于深度学习的电力缺陷检测方法展开综述，分析了电力缺陷检测对象、任务特点以及两阶段目标检测、一阶段目标检测、Transformer 视觉模型和多模态融合方法的研究进展。总体来看，两阶段方法具有较好的精细定位能力，但推理流程较复杂；以 YOLO 为代表的一阶段方法具有速度快、端到端训练和易部署等优势，适用于无人机巡检和在线监测等实时场景；Transformer 能够增强模型的全局上下文建模能力，多模态融合方法则通过结合可见光、红外、文本和领域知识等信息，提高模型对复杂缺陷场景的理解能力。尽管现有深度学习方法显著提升了电力缺陷检测的准确性和自动化水平，但距离实际智能运维应用仍存在一定差距。未来研究可以从以下几个方面进一步发展：

(1) 融合电力设备机理知识，提高模型可解释性与可靠性。当前深度学习检测模型主要依赖大量巡检数据进行特征学习，其预测过程通常缺乏明确的物理解释，而电力设备缺陷往往与长期运行过程中的材料老化、机械疲劳、电气放电以及环境因素密切相关。因此，未来研究可以探索数据驱动模型与电力设备运行机理相结合的方法，将设备结构信息、故障演化规律以及专家知识融入深度学习框架，构建具有物理约束和可解释性的智能检测模型。

(2) 面向小样本的缺陷学习方法。目前电力缺陷检测仍面临罕见缺陷样本稀缺的问题，导致模型难以获得充分的训练样本。同时，不同输电线路、设备型号以及运行环境之间存在较大的场景差异，已训练模型容易出现泛化能力不足的问题。因此，未来需要进一步研究小样本学习和零样本学习方法，降低模型对大规模人工标注数据的依赖。

(3) 从二维图像检测向三维空间感知与缺陷定位发展。当前电力缺陷检测大多基于二维 RGB 图像，虽然能够实现目标识别，但容易受到拍摄角度变化、遮挡以及尺度变化影响，难以准确描述缺陷在设备空间结构中的位置关系。随着无人机多视角采集、三维重建和神经辐射场等技术的发展，未来可进一步融合图像、点云以及空间几何信息，实现电力设备三维建模和缺陷空间定位。通过构建数字孪生场景，模型不仅能够检测缺陷，还能够实现缺陷区域精准定位、变化趋势跟踪以及设备状态动态评估，为智能巡检提供更加丰富的信息支持。

(4) 多模态大模型驱动的智能运维系统。未来电力缺陷检测的发展方向将不再局限于单一目标识别，通过融合可见光图像、红外热成像、巡检视频、气象信息、地理信息以及设备运行参数等多源数据，并结合视觉语言大模型和领域知识库，可以实现设备状态综合分析、缺陷原因推理以及维修策略辅助生成。同时，结合轻量化模型压缩、边缘计算和云边协同技术，有望实现从云端智能分析到现场实时决策的完整闭环，提高电力系统智能运维水平。

参考文献

- [1] 杨育熙, 夏启辉, 谭佳欣, 等. 基于改进 YOLOv9 的高压电缆缺陷检测算法研究[J]. 工程科学学报, 2025, 47(11): 2269-2280.
- [2] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. arXiv:1706.03762.
- [3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2020) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929.

- [4] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/cvpr.2014.81>
- [5] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/iccv.2015.169>
- [6] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
- [7] Ling, Z., Qiu, R.C., Jin, Z., *et al.* (2018) An Accurate and Real-Time Self-Blast Glass Insulator Location Method Based on Faster R-CNN and U-Net with Aerial Images. arXiv:1801.05143.
- [8] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al.* (2016) SSD: Single Shot Multibox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [9] Redmon, J. (2018) YOLOv3: An Incremental Improvement. arXiv:1804.02767.
- [10] Varghese, R. and Sambath, M. (2024) YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, 18-19 April 2024, 1-6. <https://doi.org/10.1109/adics58448.2024.10533619>
- [11] Khanam, R. and Hussain, M. (2024) YOLOv11: An Overview of the Key Architectural Enhancements. arXiv:2410.17725.
- [12] 徐海洋, 卢泉. 改进 SSD 的变电站电力设备识别方法[J]. 计算机与数字工程, 2024, 52(1): 240-246.
- [13] 尤振飞, 罗麟, 赵健, 等. 基于风格迁移及改进 SSD 算法的绝缘子定位方法[J]. 电瓷避雷器, 2022(6): 180-189.
- [14] Chen, H., He, Z., Shi, B. and Zhong, T. (2019) Research on Recognition Method of Electrical Components Based on YOLOv3. *IEEE Access*, **7**, 157818-157829. <https://doi.org/10.1109/access.2019.2950053>
- [15] Tu, R., Zhu, Z., Bai, Y., Gao, M. and Ge, Z. (2021) Key Parts of Transmission Line Detection Using Improved YOLOv3. *The International Arab Journal of Information Technology*, **18**, 747-754. <https://doi.org/10.34028/iajit/18/6/1>
- [16] Liu, C., Wu, Y., Liu, J. and Han, J. (2021) MTI-YOLO: A Light-Weight and Real-Time Deep Neural Network for Insulator Detection in Complex Aerial Images. *Energies*, **14**, Article 1426. <https://doi.org/10.3390/en14051426>
- [17] Chen, Y., Liu, H., Chen, J., Hu, J. and Zheng, E. (2023) Insu-YOLO: An Insulator Defect Detection Algorithm Based on Multiscale Feature Fusion. *Electronics*, **12**, Article 3210. <https://doi.org/10.3390/electronics12153210>
- [18] Deng, Z., Li, J., Huang, J., Jiang, S., Wu, Q. and Yang, R. (2025) PowerStrand-YOLO: A High-Voltage Transmission Conductor Defect Detection Method for UAV Aerial Imagery. *Mathematics*, **13**, Article 2859. <https://doi.org/10.3390/math13172859>
- [19] Liu, C., Liu, J., Wu, Y. and Sun, Z. (2025) RD-YOLO: Towards Rust Defect Detection for Future Unmanned Transmission Lines Maintenance. *IEICE Transactions on Information and Systems*, **108**, 1348-1358. <https://doi.org/10.1587/transinf.2024edp7265>
- [20] Jiang, J., Zhang, Y., Zhou, X., AL-Bukhaiti, K., Anping, W., Cheng, X., *et al.* (2025) Enhanced YOLO-IASE for Robust Safety Inspection in Complex Substation Environments. *Signal, Image and Video Processing*, **19**, Article No. 686. <https://doi.org/10.1007/s11760-025-04281-7>
- [21] Yang, Z., Xu, W., Chen, N., Chen, Y., Wu, K., Xie, M., *et al.* (2025) SDA-YOLO: Multi-Scale Dynamic Branching and Attention Fusion for Self-Explosion Defect Detection in Insulators. *Electronics*, **14**, Article 3070. <https://doi.org/10.3390/electronics14153070>
- [22] Fan, Z., Zhao, Y., Liu, C. and Qiu, J. (2025) DB-YOLO: A Dual-Branch Parallel Industrial Defect Detection Network. *Sensors*, **25**, Article 6614. <https://doi.org/10.3390/s25216614>
- [23] Lu, L., Wu, X., Zhang, S., Liu, J., Jiang, J. and Gu, Y. (2026) AIR-YOLO: Enhanced Small-Object Detection for Transmission Line Inspection in Complex Backgrounds. *Multimedia Systems*, **32**, Article No. 170. <https://doi.org/10.1007/s00530-026-02234-9>
- [24] 李帷韬, 侯建平, 张倩, 等. 基于强化学习和 Transformer 的输电线路缺陷智能检测方法研究[J]. 高电压技术, 2023, 49(8): 3373-3384.
- [25] 刘梓良, 尼鹿帕尔·艾克木, 伊力哈木·亚尔买买提. ELT-RTDETR: 轻量化变电站缺陷检测的 Transformer 增强模型[J]. 哈尔滨工业大学学报, 2026, 58(4): 141-155.
- [26] Du, Z., Geng, Y., Wang, H., Zhang, H. and Hu, Y. (2026) Aerial Insulator Defect Detection Method Based on CWSP-YOLO. *Scientific Reports*, **16**, Article No. 18652. <https://doi.org/10.1038/s41598-026-40782-2>
- [27] Wang, Q., Zhang, J., Du, J., Zhang, K., Li, R., Zhao, F., *et al.* (2024) A Fine-Tuned Multimodal Large Model for Power

-
- Defect Image-Text Question-Answering. *Signal, Image and Video Processing*, **18**, 9191-9203.
<https://doi.org/10.1007/s11760-024-03539-w>
- [28] Dong, F., Xu, X. and Dou, Z. (2026) An On-Site Equipment Defect Recognition Method for UAV Inspection Scenarios via Spatial Reasoning Learning with Vision-Language Multimodal Large Models. *IEEE Access*, **14**, 84787-84796.
<https://doi.org/10.1109/access.2026.3689947>