



一种多模态协同学习的变电设备缺陷识别方法

王纯一, 王雪

引用本文:

王纯一, 王雪. 一种多模态协同学习的变电设备缺陷识别方法[J]. *智能系统学报*, 2026, 21(4): 1004–1012.

WANG Chunyi, WANG Xue. Multimodal collaborative learning method for defect recognition in power substation equipment[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 1004–1012.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202509009>

您可能感兴趣的其他文章

多视图主动学习的多样性样本选择方法研究

Diversity sample selection method of multiview active learning classification

智能系统学报. 2021, 16(6): 1007–1014 <https://dx.doi.org/10.11992/tis.202007037>

利用残差密集网络的运动模糊复原方法

Image restoration with residual dense network

智能系统学报. 2021, 16(3): 442–448 <https://dx.doi.org/10.11992/tis.201912002>

自步稀疏最优均值主成分分析

Sparse optimal mean principal component analysis based on self-paced learning

智能系统学报. 2021, 16(3): 416–424 <https://dx.doi.org/10.11992/tis.201911028>

弹性网络核极限学习机的多标记学习算法

Multi-label learning algorithm of an elastic net kernel extreme learning machine

智能系统学报. 2019, 14(4): 831–842 <https://dx.doi.org/10.11992/tis.201806005>

基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

基于加权边缘弱化引导滤波的人脸光照补偿

Face illumination compensation based on weighted edge-weakening guided image filter

智能系统学报. 2018, 13(3): 373–379 <https://dx.doi.org/10.11992/tis.201612011>

DOI: 10.11992/tis.202509009

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20260407.1005.002>

一种多模态协同学习的变电设备缺陷识别方法

王纯一¹, 王雪²

(1. 国网上海浦东供电公司电力调度控制中心, 上海 200122; 2. 西北工业大学 计算机学院, 陕西 西安 710072)

摘要:为解决变电设备缺陷识别任务中多谱段图像融合特征学习不鲁棒难题, 提出了一种融合语义增强与模态特征解耦的多模态协同学习方法。该方法引入具备语义理解能力的大模型 DeepSeek, 对红外与紫外图像的语义标注进行分析与补充, 丰富标注中蕴含的语义信息表达; 设计了模态共享与模态特有特征提取模块, 通过共享查询向量和特有查询向量分别建模两种模态间的共性特征和差异特征, 以提升多模态特征表示能力; 引入特征对齐损失和模态解耦损失, 分别用于增强模态间共享特征的一致性和抑制冗余信息干扰。实验结果表明, 该方法在变电设备缺陷识别任务中优于现有主流方法, 消融实验及可视化结果进一步验证了其在多模态协同学习方面的有效性。结论可为多模态融合表征学习方法及应用提供参考。

关键词: 红外图像; 紫外图像; 大语言模型; 多模态建模; 协同学习; 注意力机制; 变电设备; 缺陷识别

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2026)04-1004-09

中文引用格式: 王纯一, 王雪. 一种多模态协同学习的变电设备缺陷识别方法 [J]. 智能系统学报, 2026, 21(4): 1004-1012.

英文引用格式: WANG Chunyi, WANG Xue. Multimodal collaborative learning method for defect recognition in power substation equipment[J]. CAAI transactions on intelligent systems, 2026, 21(4): 1004-1012.

Multimodal collaborative learning method for defect recognition in power substation equipment

WANG Chunyi¹, WANG Xue²

(1. STATE GRID Shanghai Pudong Electric Power Supply Company, Shanghai 200122, China; 2. School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: To address the poor robustness of fused feature learning for multispectral images in substation equipment defect recognition, this study proposes a multimodal feature learning method that integrates semantic enhancement and modal feature decoupling. This method first introduces DeepSeek, a large-scale model with semantic understanding capabilities, to analyze and supplement the semantic annotations of infrared and ultraviolet images, thereby enriching the semantic information in the annotations. Modality-shared and modality-specific feature extraction modules are then designed. These modules employ shared and modality-specific query vectors to model common and distinct features across the two modalities, enhancing multimodal feature representation. Furthermore, feature alignment and decoupling losses are introduced to strengthen trans-modal shared feature consistency and suppress interference from redundant information. Extensive experimental results demonstrate that the proposed method outperforms existing mainstream approaches in electrical equipment defect recognition. Ablation studies and visualization results further validate its effectiveness in multimodal collaborative learning. This conclusion provides a reference for research and application of multimodal fusion representation learning methods.

Keywords: infrared image; ultraviolet image; large language model; multimodal modeling; collaborative learning; attention mechanism; power substation equipment; defect recognition

随着电力系统规模的不断扩大和设备运行环境的日益复杂, 电网包含输电、变电及配电设备

数量急剧增加, 其运行环境也呈现高电压、大负荷和多工况的特点^[1-3]。在这种背景下, 电力设备的运行状态直接关系到整个电力系统的安全与稳定。一旦发生异常或故障, 可能导致供电中断、设备损坏, 甚至引发大范围的电力事故, 带来严

收稿日期: 2025-09-05. 网络出版日期: 2026-04-07.

基金项目: 国家电网有限公司总部科技项目 (5700-202019186A-0-0-00).

通信作者: 王雪, E-mail: xwang@nwpu.edu.cn.

©《智能系统学报》编辑部版权所有

重的经济损失和社会影响^[4]。因此,开展电力设备异常检测与故障识别,不仅是保障电网安全稳定运行的重要前提,更是提升电力系统智能化管理水平的核心环节。

电力设备缺陷的发展具有动态演化性,不同类型缺陷在演化的不同阶段,在不同光谱段下会呈现出不同程度的响应。基于可见光(visible, RGB)图像的检测通过摄像头捕捉设备外观、结构及状态的视觉特征差异,适用于可直接或间接通过肉眼(或光学放大)观察到物理形态、颜色、位置变化的缺陷类型,难以识别设备内部发热、局部放电等隐性故障,存在一定的检测盲区。为克服该问题,近年来,红外(infrared, IR)与紫外(ultraviolet, UV)成像技术被广泛引入电力巡检^[5-7],为设备状态评估提供了更全面、多维的信息支持。红外图像反映设备表面的热分布信息,对过热、接触不良等热故障具有良好的响应能力;紫外成像感知电晕、局部放电等放电类故障,是识别绝缘缺陷的重要手段^[8-10]。两者在光谱层面互补,分别揭示了设备在热异常和电异常方面的不同表现。然而,如何充分融合红外与紫外图像,使其协同服务于电力设备异常识别任务,仍面临诸多挑战。一方面,不同模态间存在显著的特征差异和语义鸿沟;另一方面,异常区域通常具有非结构化、不规则等特征,难以通过简单拼接或浅层融合手段实现有效多模态建模。

近年来,国内外研究人员围绕多模态信息融合在电力巡检中的应用开展了大量探索^[11-17]。文献^[18]提出一种基于YOLOv8n(you only look once version 8 nano)的电力设备红外图像实例分割算法,通过充分利用上下文和全局信息,增强了模型对复杂场景的理解能力。文献^[19]设计了一个面向输变电设备缺陷的语料库,涵盖文本语料、实体语料和事件语料,并且采用LoRA(low-rank adaptation)方法对LLaMA(large language model architecture)大语言模型^[20]进行了微调,并结合电网领域知识图谱对语料数据进行了质量优化。文献^[21]针对配电室中可见光图像易受环境干扰、红外图像纹理特征不明显的问题,提出了一种改进的红外与可见光图像融合网络。该方法在特征提取与重建阶段引入坐标注意力模块,有效融合不同模态信息,提升图像信息量。文献^[22]提出了一种基于提示学习的多谱段融合识别方法,通过引入红外与紫外图像的可训练提示词(prompts),将其作为唯一可更新的参数参与模型训练。该策略在提升识别性能的同时,大幅减少了训练所需

参数量,有效降低了大模型对下游数据规模的依赖。然而,现有多模态融合方式多基于浅层拼接或注意力机制,难以有效建模红外与紫外图像间深层语义关联,尤其在处理复杂场景和弱特征异常时表现不稳。此外,现有方法对语义信息的利用较为有限,缺乏对异常区域的高层语义理解与解释能力,导致模型泛化性不足。

随着大规模预训练文本-视觉模型的发展,跨模态理解与表达能力不断增强,在图像识别^[23-26]、目标检测^[27]和异常分析^[28]等多个领域展现出强大的泛化能力。根据可学习软提示词的应用方式,现有方法大致分为基于文本的^[29-30]和基于视觉的方法^[31]。相比传统方法,大规模预训练模型能够更好地融合视觉信息与语义信息,为复杂场景下的异常判别提供了新的思路。在电力设备巡检场景中,异常往往呈现多样化、模态差异显著且语义表达模糊等特点,传统基于单一模态或浅层融合的检测方法难以应对这类复杂场景。

为解决红外与紫外模态融合不充分、异常区域语义表达不足、模态差异难以有效建模等问题,本文提出了一种结合语义增强与模态特征解耦的大模型驱动多模态协同学习方法,通过充分挖掘多模态数据的语义潜力与互补特性,提升变电设备缺陷识别的准确性与鲁棒性。本文主要贡献如下:

1) 提出一种基于模态特征解耦的多模态协同学习框架,通过设计模态共享特征提取模块和模态特有特征提取模块,分别建模红外与紫外模态的共性与差异特征,显著提升多模态特征表示能力。

2) 在多模态协同学习训练过程中,引入跨模态特征对齐损失与模态解耦损失,在强化模态间共享特征一致性的同时,有效抑制冗余信息干扰。

3) 提出一种利用大语言模型对数据集基础标注进行语义扩充的增强方法,可生成细粒度结构化文本标注,有利于在设备状态视觉表征与状态语义间建立有效关联。

1 数据集构建

1.1 数据集标注的语义扩充

现有的多模态电力设备异常识别数据集通常仅标注设备与缺陷类型,标注方式多采用简短关键词组合,如“infrared and ultraviolet abnormal ammeter and wire”,缺乏结构化的语义描述。这类描述仅涵盖设备类型与模态类别,既无法体现红外图像的热异常特征或紫外图像的放电现象,也缺

乏对异常表现的语义细化与上下文信息补充,导致模型无法捕捉到与缺陷相关的视觉细节,难以建立异常表现与设备状态之间的语义联系。

为克服上述不足,本文引入大语言模型^[20,32-33]对原始标注进行语义扩展,如图 1 所示,采用 DeepSeek 大模型将原始标注“Infrared and ultraviolet abnormal ammeter and wire”扩充为“Infrared and ultraviolet abnormal ammeter and wire. Infrared images show significant hotspots. Ultraviolet images reveal corona discharge or arc flashes, indicating serious issues such as insulation degradation or connector faults”。用于语义扩充的中文结构化提示词模板为“请基于电力设备巡检知识,将以下标签扩充为‘模态特征 + 异常表现 + 故障原因’的结构化文本: {原始标签}。要求: 红外需描述热分布(如热点), 紫外需描述放电特征(如电晕), 内容需匹配设备类型与状态。”在模型选择方面, 本工作使用 DeepSeek 官方开放的 DeepSeek-V3 应用程序编程接口, 通过兼容接口发起调用, 无需本地部署即可实现批量语义扩充, 确保流程的便捷性与可复现性。

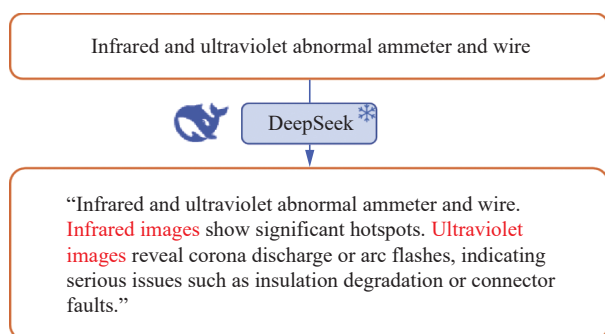


图 1 利用预训练模型对数据集标注进行扩充
Fig. 1 Annotation augmentation with pre-trained models

相比传统标签, 该增强标注方式具有以下优势: 1) 显式语义增强: 引入了对红外图像“热点异常”和紫外图像“电晕放电/电弧闪络”的具体描述, 使模型更容易学习到异常类型与模态特征之间的因果关系; 2) 支持复杂推理: 通过描述可能的故障原因(如绝缘老化、连接器故障), 赋予模型一定的语义推理能力, 有助于识别难以从图像直接观察到的深层故障模式; 3) 增强模态对齐效果: 更细致的模态描述有助于优化模态共享与特有特征的提取过程, 提升多模态融合的有效性。

1.2 扩充后的变电设备数据集

本文采用的数据集来源于文献[22], 该数据集包含紫外、红外和可见光 3 种谱段下的变电设备图像, 具备多模态、多场景和多设备类型的特点。数据集中共覆盖 6 类典型变电设备, 分别为

干式变压器、隔离开关、电流表与导线、油浸式变压器、冲击器以及开关柜, 涵盖了变电站巡检中常见的设备类型与故障场景。在原始标注策略中, 文献[22]采用了“状态+类型”的组合方式对图像样本进行标注, 其中“状态”指设备运行是否正常, “类型”指设备类别。为了提升模型对异常类型及其多模态表现形式的理解能力, 本文利用图 1 所示方法对原始标签进行语义增强, 生成更丰富且结构化的文本标注, 为后续多模态特征学习与异常判别提供 stronger 的语义支撑。图 2 给出了部分语义增强后的标注。

缺陷标注:

“Norm impulse voltage generator. Infrared images show no abnormal thermal patterns. Ultraviolet images reveal no corona discharge or electrical sparking. The equipment is operating under normal conditions.”

“Infrared abnormal ammeter and wire. Infrared images reveal localized overheating at terminals or conductors, indicating possible poor contact or overload. No abnormality is observed in ultraviolet images.”

“Infrared and ultraviolet abnormal ammeter and wire. Infrared images show significant hotspots. Ultraviolet images reveal corona discharge or arc flashes, indicating serious issues such as insulation degradation or connector faults.”

“Norm dry-type transformer. Temperature distribution is uniform in infrared images, with no hotspots. Ultraviolet images show no signs of corona or partial discharge. The transformer operates normally.”

图 2 语义增强后的数据集标注示例
Fig. 2 Semantically enhanced annotation examples

2 多模态视觉特征建模

本文所提出的多模态视觉特征建模方法流程如图 3 所示。该方法主要包含了单模态特征提取、多模态特征建模和模型训练损失函数 3 个部分。具体地, 以配准后的三谱段图像以及经过语义增强的紫外/红外谱段提示词为输入, 单模态特征提取模块首先提取到紫外可见融合特征以及红外可见融合特征, 并传入多模态特征建模模块。该模块基于模态特征解耦策略和交叉注意力机制, 分别提取跨模态共享特征和单独模态私有特征, 最终得到多模态融合图像特征, 用于识别电力设备故障。除交叉熵损失外, 本文引入跨模态特征对齐损失、模态解耦损失以及双向文本图像对比损失, 共同用于模型训练。这种联合监督学习和无监督学习的训练方式有利于引导模型学习更具判别性的特征表示, 从而提高模型的收敛速度和性能。本节将分别对这 3 部分内容进行详细介绍。为方便理解, 表 1 列出了本文用到的主要数学符号, 并给出解释。

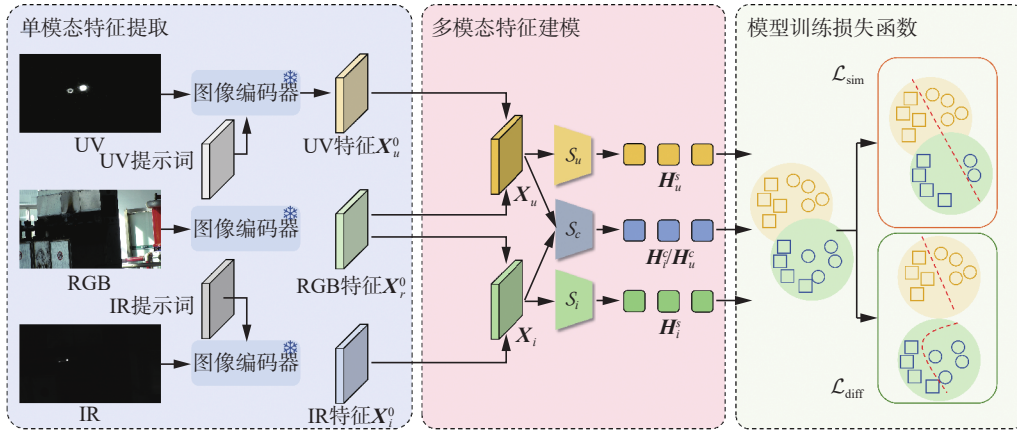


图3 本文提出的多模态特征建模流程

Fig. 3 The pipeline of our proposed multi-modal feature modeling

表1 数学符号

Table 1 Mathematical symbols

符号	说明	符号	说明
X_r^0	红可光图像特征	X_u^0	紫外图像特征
X_i^0	红外图像特征	D	单模态特征维度
L	单模态特征长度	X_u	紫外可见融合特征
X_i	红外可见融合特征	N	训练过程中批量大小
X_q^c	共享特征查询向量	N_q	查询向量的数量
H_u^c	由 X_u 提取共享特征	H_i^c	由 X_i 提取共享特征
$X_{q,u}^s$	紫外模态查询向量	$X_{q,i}^s$	红外模态查询向量
H_u^s	由 X_u 提取私有特征	H_i^s	由 X_i 提取私有特征
H_u	最终紫外特征	H_i	最终红外特征
V	多模态融合图像特征	T	语义增强后文本特征
$\mathcal{L}_{sim}$	共享特征对齐损失	$\mathcal{L}_{diff}$	私有特征差异损失
$\mathcal{L}_{v2t}$	图像到文本对比损失	$\mathcal{L}_{t2v}$	文本到图像对比损失
$\mathcal{L}_{clip}$	对比损失函数	$\mathcal{L}_{ce}$	交叉熵损失函数
$\mathcal{L}_{total}$	总损失函数	λ_1	$\mathcal{L}_{clip}$ 损失项权重
λ_2	$\mathcal{L}_{sim}$ 损失项权重	λ_3	$\mathcal{L}_{diff}$ 损失项权重

2.1 单模态特征提取

本文利用经典多模态模型 CLIP (contrastive language-image pre-training) 的图像分支编码器提取图像特征。冻结预训练模型参数可大幅减少训练参数量。不同于传统词元 (token) 形式的提示词, 对于紫外和红外图像分支, 本文借鉴文献 [22] 的方法, 通过在这两个分支分别添加与输入图像尺寸一致 (例如 224 像素 $\times$ 224 像素) 的二维可训练提示词, 来学习紫外和红外图像中的故障信息。这种设计能够适配不同模态中故障信息的建模与提取需求, 从而增强多模态图像中异常特征的表达能力。最终得到的可见光、紫外和红外特

征分别表示为

$$X_r^0, X_u^0, X_i^0 \in \mathbb{R}^{L \times D}$$

式中: D 为特征维度, L 为展平后的图像特征长度。

由于紫外和红外分支仅包含放电、发热等异常信息, 缺乏对设备外观及类型等信息的完整描述, 需要将可见光特征分别与紫外、红外特征融合, 以弥补二者在背景理解上的不足。融合时采用不同模态特征相加的方式, 得到的特征向量为 X_u 和 X_i 。

2.2 红外和紫外图像共享特征提取

红外和紫外图像均属于非常规可见光谱段, 分别对应设备的热辐射与局部放电现象, 具有高度互补的异常信息表达能力。尽管它们的物理成像机制不同, 但二者在设备故障检测任务中表现出一定的特征共性, 例如均能反映出高温、放电、老化等异常区域。因此, 本文设计了一个共享特征提取模块 S_c , 以增强模型对跨模态异常特征的统一感知能力。具体地, 定义一组可学习的查询向量 $X_q^c \in \mathbb{R}^{N_q \times D}$, N_q 为查询向量的数量。对于不同模态的特征向量, 在 X_q^c 与 $\{X_u, X_i\}$ 之间建立交叉注意力机制。

$$H_m^c = \text{Cross-Attn}(X_q^c, X_m, X_m), m \in \{u, i\}$$

式中 $H_m^c \in \mathbb{R}^{N_q \times D}$ 为对齐的特征表达, 不同模态特征使用共享的查询向量来提取一致性的表达。对于紫外模态, X_q^c 为查询变量 (query), 键向量 (key) 和值向量 (value) 均为 X_u 。对于红外模态同理。

在训练过程中, 为了约束红外特征和紫外特征在共享异常区域上的表达一致性, 本文设计了模态一致性约束机制。具体地, 与文献 [31] 类似, 在共享特征提取模块输出的红外和紫外特征之间 $\{H_u^c, H_i^c\}$, 引入如下特征对齐损失:

$$\mathcal{L}_{sim} = \|E(H_u^c) - E(H_i^c)\|_2 + \|C(H_u^c) - C(H_i^c)\|_2$$

式中: $E(\cdot)$ 为均值, $C(\cdot)$ 为协方差, $\|\cdot\|_2$ 为欧氏距离。上述损失通过最小化不同模态特征分布间距离进行约束, 使得模型在学习模态共享特征时增强对异常区域的跨模态一致性表达能力, 从而有效减小红外与紫外图像在特征空间中的分布差异, 同时有利于消除特征维度间的相关性。

2.3 红外和紫外图像独享特征提取

尽管红外与紫外图像在异常检测中具有一定共性信息, 但二者的成像机制和关注焦点存在差异: 红外图像擅长捕捉设备表面的温度分布特征, 而紫外图像则能揭示局部放电、击穿等电气异常现象。因此, 二者还包含不同模态的差异性信息。为有效提取和保留各模态的独享特征, 在共享特征提取模块之后, 为紫外和红外图像分别设计了模态特有分支 $\mathcal{S}_u$ 和 $\mathcal{S}_i$, 分别用于提取各自模态独享特征。具体地, 在紫外和红外分支分别定义可学习的查询向量 $\mathbf{X}_{q,u}^s$ 和 $\mathbf{X}_{q,i}^s \in \mathbb{R}^{N_u \times D}$ 。对于每个分支, 令键向量和值向量相同, 使用上述查询向量提取独享特征。

$$\mathbf{H}_m^s = \text{Cross-Attn}(\mathbf{X}_{q,m}^s, \mathbf{X}_m, \mathbf{X}_m), m \in \{u, i\}$$

注意, 这里不同模态的注意力权重不共享。

同样地, 在训练过程中, 为了约束不同分支提取到特有的表达, 本文设计了模态私有特征差异损失, 又称作模态解耦损失。该损失旨在增强单独模态分支学到的特有表示, 使其尽可能捕捉本模态下更具判别性或显著性的特征信息, 同时避免其退化为模态共享特征的冗余表达。具体地, 对于两个模态的私有特征 $\{\mathbf{H}_u^s, \mathbf{H}_i^s\}$, 其差异损失定义为

$$\mathcal{L}_{\text{diff}} = \sqrt{\text{Tr}\left(\left((\mathbf{H}_u^s)^T \mathbf{H}_i^s\right)^T \left((\mathbf{H}_u^s)^T \mathbf{H}_i^s\right)\right)}$$

式中 $\text{Tr}(\cdot)$ 为矩阵的迹。迹可量化矩阵特征值之和, 通过最小化乘积矩阵的迹, 不仅能够增强各模态的表达多样性, 也可避免共享与特有表征之间的冗余或冲突^[34], 进一步提升多模态特征在变电设备缺陷检测任务中的判别能力。

2.4 损失函数

在提取紫外和红外图像的共享及独享特征后, 先对两类特征分别相加后融合, 得到最终的紫外特征向量 $\mathbf{H}_u$ 和红外特征向量 $\mathbf{H}_i$; 再将这两个向量相加融合, 最终得到完整的多模态融合图像特征 $\mathbf{V}$ 。另一方面, 对于文本特征, 将经过语义增强的图像标注输入到 CLIP 的文本编码器中进行特征提取, 得到的特征表示为 $\mathbf{T}$ 。训练过程中采用与原始 CLIP 一致的对比学习损失 (contrastive loss) 策略: 通过最大化图像-文本正样本对之间的相似度, 并最小化其与其他负样本对的相似

度, 从而对齐图文语义空间, 提升模型对缺陷语义的理解能力。该策略具体可表示为图像到文本方向的对比损失:

$$\mathcal{L}_{v2t} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\text{sim}(\mathbf{V}_i, \mathbf{T}_i))}{\sum_{j=1}^N \exp(\text{sim}(\mathbf{V}_i, \mathbf{T}_j))}$$

以及文本到图像方向的对比损失:

$$\mathcal{L}_{t2v} = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\text{sim}(\mathbf{T}_i, \mathbf{V}_i))}{\sum_{j=1}^N \exp(\text{sim}(\mathbf{T}_i, \mathbf{V}_j))}$$

式中 N 为批量大小。总对比损失函数可表示为

$$\mathcal{L}_{\text{clip}} = \mathcal{L}_{v2t} + \mathcal{L}_{t2v}$$

最终, 训练过程中的总损失函数表示为

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{ce}} + \lambda_1 \mathcal{L}_{\text{clip}} + \lambda_2 \mathcal{L}_{\text{sim}} + \lambda_3 \mathcal{L}_{\text{diff}} \quad (1)$$

式中: $\mathcal{L}_{\text{ce}}$ 为交叉熵分类损失, 超参数 λ_1 、 λ_2 、 λ_3 用于平衡不同损失项的权重。

3 变电设备故障识别实验结果与分析

3.1 数据集

本文在文献 [22] 提出的数据集上验证所提方法性能。该数据集面向变电设备故障识别任务, 涵盖了电力系统中常见的 6 类设备, 共计 12 种不同的“电力设备+状态”类别。其中, 干式变压器、开关柜、电流表与导线这 3 类设备均包含正常和故障状态, 而油浸式变压器、隔离开关、冲击电压发生仅有正常状态数据。电力设备发生故障时, 通常会伴随有红外热异常 (如发热) 或紫外光异常 (如局部放电) 等现象, 这些多谱段传感数据为变电设备故障自动识别提供了重要依据。该数据集共收集了 1697 组已配准的三谱段图像, 图像分辨率为 1920 像素 × 1080 像素, 其中 1352 组用于训练, 345 组用于测试。图 4 给出了该数据集中不同类别数据的分布情况, 反映该数据集存在一定的类别分布不均衡现象。

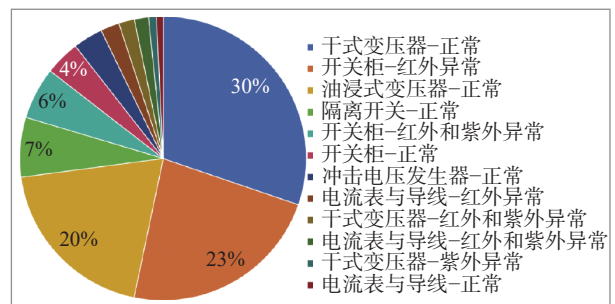


图 4 变电设备故障识别数据集中不同类别数据分布情况
Fig. 4 Class distribution of substation equipment fault identification dataset

3.2 实验细节

本实验采用 CLIP 模型作为基础架构, 图像编码器选用 CLIP-ViT-B/32^[23] (vision transformer base model, 32 表示图像块大小)。为充分利用预训练模型在图文对齐上的表达能力, 实验中冻结了 CLIP 模型的全部参数, 仅对新增模块 (如 prompts、共享/独享特征提取模块) 进行训练。在输入阶段, 所有图像被统一缩放至 224 像素×224 像素分辨率, 并进行归一化处理以适配 CLIP 图像编码器。训练过程中, 批量大小设为 64, 使用带有预热机制的学习率调度策略以稳定训练初期的参数更新, 初始学习率设置为 40, 动量参数设为 0.9。式(1)中超参数 λ_1 设置为 1.0, λ_2 和 λ_3 均设置为 0.5。所有实验在 NVIDIA GeForce RTX 3080 单卡图形处理器 (graphics processing unit, GPU) 上进行。

3.3 与现有先进方法的对比

本文提出方法与现有先进方法^[22-23,29-31] 在变电设备故障识别数据集上的性能对比结果如表 2 所示。

表 2 在变电设备故障识别数据集上的对比实验结果
Table 2 Comparison results on substation equipment fault Identification dataset

模型	准确率/%	F_1
CLIP ^[23]	25.22	0.151
linear CLIP ^[23]	73.33	0.604
padding prompt ^[31]	82.35	0.754
random patch ^[31]	47.25	0.394
fixed patch ^[31]	78.55	0.631
IR+UV prompt ^[22]	90.14	0.891
ArGue ^[29]	92.75	0.916
ATPrompt ^[30]	93.33	0.926
本文方法	96.52	0.947

注: 加粗代表最优。

CLIP 模型^[23] 为没有增加任何训练参数的原始预训练模型。linear CLIP^[23] 对原始 CLIP 骨干网络进行冻结, 仅训练通过全连接层 (full connection, FC) 构成的分类头, 且使用与本文方法完全一致的训练参数。填充提示词 (padding prompt) 模型^[31] 在图像边缘添加可学习的填充区域。相关填充参数在训练过程中更新, 从而引导模型关注边缘或背景信息。随机图块 (random patch) 模型^[31] 在原始图像中随机选取一个位置, 将该区域替换为一个可学习的图块参数。该图块在训练过程中被优化, 从而注入额外的引导信息。固定图块 (fixed patch) 模型^[31] 与 random patch 模型类似, 但使用固定位置的图像块进行替换, 以避免训练过程的随机性。实验中图像块的尺寸设置为 30 像

素×30 像素。红外紫外提示词 (IR+UV prompt) 模型^[22] 利用红外和紫外图像信息作为引导, 设计提示词融入模块, 以增强对多模态信息的理解能力。ArGue 模型^[29] 利用视觉属性增强模型的语义理解能力。ATPrompt (attribute-anchored textual prompt) 模型^[30] 通过软提示学习与属性相关的额外表征。

实验结果表明, 原始预训练 CLIP 模型在未进行任何领域微调情况下的 zero-shot 推理结果较差, 准确率仅有 25.22%。这源于自然图像与电力设备图像之间存在较大的特征分布差异。重新训练分类头的 linear CLIP 模型准确率提升到 73.33%。引入不同形式的 prompt 策略进一步提升了 CLIP 模型在变电设备故障识别任务中的表现, 其中 padding prompt、fixed patch prompt 和 IR+UV prompt 分别取得了 82.35%、78.55% 和 90.14% 的准确率, 验证了可学习提示信息在提升模型感知能力方面的重要作用。相比之下, random patch 模型的效果较差, 仅为 47.25%, 表明随机位置的提示难以稳定提供有效引导。本文提出方法在所有对比方法中表现最佳, 其准确率达到 96.52%, 不仅显著优于现有的多种 prompt 设计, 也进一步证明了所提方法在引导多模态特征学习、提升任务性能方面的有效性。 F_1 分数表现出相似的模型性能变化趋势。

3.4 不同网络模块对实验结果影响

为验证本文提出方法中各模块的有效性, 本节在 IR+UV prompt 模型的基础上, 逐步引入共享特征提取模块与独享特征提取模块, 构建多组对照实验以量化分析不同模块对模型性能的贡献。实验结果如表 3 所示。IR+UV prompt 模型单独使用时的准确率为 90.14%, 引入共享特征提取模块后, 准确率提升至 93.91%, 说明该模块能够增强不同模态间的语义对齐能力, 促进了多模态特征融合。单独引入独享特征提取模块后, 准确率提升至 91.59%, 验证了该模块在捕捉模态特有信息、补充差异性特征方面的积极作用。

表 3 不同网络模块对实验结果的影响
Table 3 Effects of different network modules on experimental results %

模型	准确率
IR+UV prompt	90.14
共享特征提取模块	93.91
独享特征提取模块	91.59
本文方法	96.52

注: 加粗代表最优。

当两者共同使用时,模型性能进一步提升至 96.52%。这一结果表明,共享特征提取模块与独有特征提取模块在跨模态信息互补以及模态特异性保留方面发挥着关键作用,二者联合使用能够显著增强模型对有效特征的提取与表征能力。

3.5 不同超参数对实验结果影响

为进一步探究超参数对模型效果的作用,本文通过多组对照实验分析了查询向量数量 N_q 对模型性能的影响,相关实验结果汇总于表 4。实验结果表明,该参数在测试范围内的数值变动,对模型最终预测准确率的抑制或提升作用均十分有限。无论 N_q 取何实验值,模型都能稳定输出较高的识别精度,抗参数扰动能力较强。横向对比所有参数组合的实验数据后发现, $N_q = 8$ 为该超参数的最优取值,在此设置下模型的识别准确度最佳。综合实验结论与模型实用型考量,本文最终选用 $N_q = 8$ 作为模型的固定参数。

表 4 不同超参数对实验结果的影响

Table 4 Effects of different hyper-parameter values on experimental results %

参数	准确率
$N_q = 4$	95.94
$N_q = 8$	96.52
$N_q = 12$	94.78

注:加粗代表最优。

3.6 可视化结果

图 5 给出了本文提出方法对部分典型变电设备缺陷进行识别的可视化结果,其中涵盖了正常的开关柜、红外与紫外异常的开关柜、红外与紫外异常的干式变压器以及正常的油浸式变压器。通过对比可以观察到,当红外图像中存在明显的发热区域或紫外图像中出现放电特征时,本文所提方法能够有效实现多模态图像的联合建模,准确捕捉异常信号并识别缺陷类别。这进一步验证了所提方法在融合多谱段信息、识别复杂异常模式方面的有效性,为其在相关任务中的应用可行性提供了有力支撑。



(a) 正确识别示例1: 正常的开关柜



(b) 正确识别示例2: 红外和紫外异常的开关柜



(c) 正确识别示例3: 红外和紫外异常的干式变压器



(d) 正确识别示例4: 正常的油浸式变压器

图 5 本文方法的可视化结果

Fig. 5 Visualization results of our method

4 结论

1) 本文针对变电设备红外与紫外图像中的缺陷识别任务,提出了一种融合语义增强与模态特征解耦的多模态特征学习方法。通过引入具备强语义理解能力的大模型 DeepSeek 对红外与紫外图像的原始标签进行补充与优化,显著提升了数据的语义表达能力。

2) 在此基础上,本文设计了模态共享与模态特有的特征提取模块,并且结合特征对齐损失与模态解耦损失,分别建模了红外与紫外模态之间的共性特征与差异特征,充分挖掘多模态数据的互补优势。

3) 对比实验验证了所提方法在变电设备缺陷识别任务中的优越性能,消融实验和可视化分析也充分证明了各关键模块的有效性。

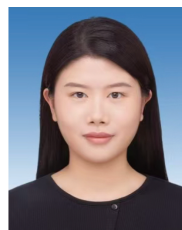
参考文献:

- [1] 王成山, 李鹏. 分布式发电、微网与智能配电网的发展与挑战[J]. 电力系统自动化, 2010, 34(2): 10-14, 23.
WANG Chengshan, LI Peng. Development and challenges of distributed generation the micro-grid and smart distribution system[J]. Automation of electric power systems., 2010, 34(2): 10-14, 23.
- [2] 张智刚, 康重庆. 碳中和目标下构建新型电力系统的挑战与展望[J]. 中国电机工程学报, 2022, 42(8): 2806-2819.
ZHANG Zhigang, KANG Chongqing. Challenges and prospects for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2819.
- [3] 程其云, 李峰, 罗澍忻, 等. 新型电力系统规划方法框架及关键支撑技术[J]. 电网技术, 2025, 49(6): 2219-2231.
CHENG Qiyun, LI Feng, LUO Shuxin, et al. Research on the planning methodology framework and key supporting technologies for new power systems[J]. Power system technology, 2025, 49(6): 2219-2231.

- [4] 秦潘昊,陈威宇,胡秦然,等.新型电力系统设备状态监测与故障诊断传感芯片关键技术与展望[J].*电力系统自动化*,2024,48(6):83–95.
QIN Panhao, CHEN Weiyu, HU Qinran, et al. Key technologies and prospect of equipment condition monitoring and diagnosis sensor chips for new power systems[J]. *Automation of electric power systems*, 2024, 48(6): 83–95.
- [5] 范澜珊,刘云鹏,刘一瑾,等.面向电力设备红外-可见光图像配准的自适应监督重训算法[J].*高电压技术*,2025,51(4):1785–1800.
FAN Lanshan, LIU Yunpeng, LIU Yijin. Adaptive algorithm with supervision and retraining for power equipment infrared and visible images registration[J]. *High voltage engineering*, 2025, 51(4): 1785–1800.
- [6] 赵振兵,冯烁,席悦,等.大模型时代:电力视觉技术新起点[J].*高电压技术*,2024,50(5):1813–1825.
ZHAO Zhenbing, FENG Shuo, XI Yue, et al. The era of large models: a new starting point for electric power vision technology[J]. *High voltage engineering*, 2024, 50(5): 1813–1825.
- [7] 刘传洋,吴一全.基于红外图像的电力设备识别及发热故障诊断方法研究进展[J].*中国电机工程学报*,2025,45(6):2171–2196.
LIU Chuanyang, WU Yiquan. Research progress of power equipment identification and thermal fault diagnosis based on infrared images[J]. *Proceedings of the CSEE*, 2025, 45(6): 2171–2196.
- [8] 刘云鹏,李泳霖,裴少通,等.基于紫外光辐射照度特征的污秽瓷绝缘子绝缘状态评估方法[J].*高电压技术*,2023,49(4):1622–1631.
LIU Yunpeng, LI Yonglin, PEI Shaotong, et al. Evaluation of the insulation state of contaminated porcelain insulators based on characteristics of ultraviolet irradiance [J]. *High voltage engineering*, 2023, 49(4): 1622–1631.
- [9] 宋鹏,白晓晨,林晓焕,等.电力传输线异常放电紫外光检测定位仿真与模拟验证[J].*光学精密工程*,2022,30(10):1160–1169.
SONG Peng, BAI Xiaochen, LIN Xiaohuan, et al. Simulation and verification of ultraviolet detection and location of abnormal discharge in power transmission line[J]. *Optics and precision engineering*, 2022, 30(10): 1160–1169.
- [10] 律方成,牛雷雷,王胜辉,等.基于紫外成像和改进YOLOv3的瓷悬式绝缘子放电严重程度评估[J].*高电压技术*,2021,47(2):377–386.
LÜ Fangcheng, NIU Leilei, WANG Shenghui, et al. Discharge severity assessment of porcelain suspension insulators based on UV images and improved YOLOv3[J]. *High voltage engineering*, 2021, 47(2): 377–386.
- [11] 黄志鸿,杜瑞,张辉.面向复杂电力环境场景理解的可见光和红外图像特征级融合方法[J].*智能系统学报*,2025,20(3):631–640.
HUANG Zhihong, DU Rui, ZHANG Hui. Feature-level fusion method of visible and infrared images for scene understanding in complex power environments[J]. *CAAI transactions on intelligent systems*, 2025, 20(3): 631–640.
- [12] 陶岩,张辉,黄志鸿,等.面向配电网典型部件的热故障精准判别方法[J].*智能系统学报*,2025,20(2):506–515.
TAO Yan, ZHANG Hui, HUANG Zhihong, et al. Accurate identification of thermal faults for typical components of distribution networks[J]. *CAAI transactions on intelligent systems*, 2025, 20(2): 506–515.
- [13] 张建良,张晓洁,麻坚,等.面向新型电力系统的故障诊断技术研究进展(二):机器学习和大语言模型技术[J].*实验技术与管理*,2025,42(12):54–70.
ZHANG Jianlian, ZHANG Xiaojie, MA Jian, et al. A review of fault diagnosis technology for the new type power systems(part II): machine learning and large language modeling techniques[J]. *Experimental technology and management*, 2025, 42(12): 54–70.
- [14] 赵俊华,文福拴,黄建伟,等.基于大语言模型的电力系统通用人工智能展望:理论与应用[J].*电力系统自动化*,2024,48(6):13–28.
ZHAO Junhua, WEN Fushuan, HUANG Jianwei, et al. Prospect of artificial general intelligence for power systems based on large language model: theory and applications[J]. *Automation of electric power systems*, 2024, 48(6): 13–28.
- [15] 丁剑涛.面向电力设备缺陷数据的多模态知识图谱研究[D].西安:长安大学,2022.
DING Jiantao. Research on multimodal knowledge graph for power equipment defect data[D]. Xi'an: Chang'an University, 2022.
- [16] 赵欣洋,刘志远,赵宇嘉,等.基于异源特征融合的电力设备分割与异常发热检测[J].*国外电子测量技术*,2024,43(11):177–186.
ZHAO Xinyang, LIU Zhiyuan, ZHAO Yujia, et al. Segmentation and anomalous heat detection of power equipment based on heterogeneous feature fusion[J]. *Foreign electronic measurement technology*, 2024, 43(11): 177–186.
- [17] 单其尚.从情感分析到变压器故障诊断的多模态融合方法迁移应用[D].南京:南京邮电大学,2023.
SHAN Qishang. Multimodal fusion method migration application from sentiment analysis to transformer fault diagnosis[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.

- [18] 李冰, 杜喜英, 王玉莹, 等. 基于改进 YOLOv8n 的变电设备红外图像实例分割算法[J]. 电子测量技术, 2024, 47(10): 151–159.
LI Bing, DU Xiying, WANG Yuying, et al. Instance segmentation algorithm for infrared images of substation equipment based on improved YOLOv8n[J]. *Electronic measurement technology*, 2024, 47(10): 151–159.
- [19] 陈梦实. 基于语言模型的电网输变电设备缺陷语料库系统的设计与实现[D]. 北京: 北京邮电大学, 2024.
CHEN Mengshi. Design and implementation of defect corpus system for power grid equipment[D]. Beijing: Beijing University of Posts and Telecommunications, 2024.
- [20] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMa: open and efficient foundation language models[EB/OL]. (2023–02–27)[2025–05–20]. <https://arxiv.org/abs/2302.13971>.
- [21] 杨文昊. 基于多模态信息融合的故障源定位方法研究[D]. 绵阳: 西南科技大学, 2024.
YANG Wenhao. Research on fault source localization method based on multimodal information fusion[D]. Mianyang: Southwest University of Science and Technology, 2024.
- [22] 姚一杨, 杜泽星, 周果清, 等. 基于提示学习的电力设备故障多谱段融合识别方法[J]. 西北工业大学学报, 2025, 43(2): 410–417.
YAO Yiyang, DU Zexing, ZHOU Guoqing, et al. Multi-spectral fusion power equipment fault recognition based on prompt learning[J]. *Journal of Northwestern Polytechnical University*, 2025, 43(2): 410–417.
- [23] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the 38th International Conference on Machine Learning. Online: IMLS, 2021.
- [24] LIU Haotian, LI Chunyuan, LI Yuheng, et al. Improved baselines with visual instruction tuning[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024.
- [25] HAN Zeyu, GAO Chao, LIU Jinyang, et al. Parameter-efficient fine-tuning for large models: a comprehensive survey[EB/OL]. (2024–03–21)[2025–05–20]. <https://arxiv.org/abs/2403.14608>.
- [26] ZHU Lianghui, LIAO Bencheng, ZHANG Qian, et al. Vision mamba: efficient visual representation learning with bidirectional state space model[C]//Proceedings of the 41st International Conference on Machine Learning. Vienna: IMLS, 2024.
- [27] REN Tianhe, CHEN Yihao, JIANG Qing, et al. DINO-X: a unified vision model for open-world object detection and understanding[EB/OL]. (2024–11–21)[2025–05–22]. <https://arxiv.org/abs/2411.14347>.
- [28] CHEN Qifeng, CHEN Yingcong, FANG Cheng, et al. HAWK: learning to understand open-world video anomalies[C]//Advances in Neural Information Processing Systems 37. Vancouver: NeurIPS, 2024.
- [29] TIAN Xinyu, ZOU Shu, YANG Zhaoyuan, et al. ArGue: attribute-guided prompt tuning for vision-language models[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024.
- [30] LI Zheng, SONG Yibing, CHENG Mingming, et al. Advancing textual prompt learning with anchored attributes [C]//2025 IEEE/CVF International Conference on Computer Vision. Honolulu: IEEE, 2025.
- [31] BAHNG H, JAHANIAN A, SANKARANARAYANAN S, et al. Exploring visual prompts for adapting large-scale models[EB/OL]. (2022–03–31)[2025–04–20]. <https://arxiv.org/abs/2203.17274>.
- [32] ACHIAM J, ADLER S, AGARWAL S, et al. GPT-4 technical report[EB/OL]. (2023–03–15)[2025–04–20]. <https://arxiv.org/abs/2303.08774>.
- [33] LIU Aixin, FENG Bei, XUE Bing, et al. Deepseek-V3 technical report[EB/OL]. (2024–12–27)[2025–04–20]. <https://arxiv.org/abs/2412.19437>.
- [34] CHEN Chen, QI Jiahao, LIU Xingyue, et al. Weakly misalignment-free adaptive feature alignment for UAVs-based multimodal object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024.

作者简介:



王纯一, 助理工程师, 主要研究方向为多模态大模型、目标检测。
E-mail: millaywang@hotmail.com。



王雪, 副研究员, 主要研究方向为多模态学习与三维视觉。主持国家自然科学基金等科研项目 5 项, 发表学术论文 30 余篇。E-mail: xwang@nwpu.edu.cn。