第 52 卷 第 7 期：2973-2985 高电压技术 Vol.52, No.7: 2973-2985
2026 年 7 月 31 日 High Voltage Engineering July 31, 2026
DOI: 10.13336/j.1003-6520.hve.20260604

领域知识引导的视觉语言模型配电线路金具锈蚀检测方法

赵振兵[1,2,3,4]，田本喜[1]，高凯鹏[1]，唐辰康[1]，李浩鹏[1,2,3]
（1. 华北电力大学电气与电子工程学院，保定 071003；2. 河北省电力物联网技术重点实验室(华北电力大学)，保定 071003；3. 华北电力大学电力物联智慧化技术河北省工程研究中心，保定 071003；4. 华北电力大学复杂能源系统智能计算教育部工程研究中心，保定 071003）

摘　要：针对配电线路金具锈蚀检测中缺陷表征困难、语义信息不足及类别边界模糊等问题，提出一种领域知识引导的视觉语言检测方法。首先，设计低对比特征增强模块，通过多尺度方向特征提取、边界引导调制和通道自适应重标定，增强低对比锈蚀区域的纹理、边界及关键通道响应。其次，构建知识引导表征模块，将配电领域知识转化为文本提示，并利用视觉条件提示适配器动态优化文本嵌入，提高模型对锈蚀形态和金具状态的语义理解能力。最后，引入跨模态语义对齐策略，通过区域-文本对比分类损失和语义对齐正则项约束视觉特征与文本语义的一致性，增强相似类别的判别能力。实验结果表明，所提方法在自建配电线路金具缺陷数据集上，平均精度均值达到 87.1%；在开放语义检测和零样本泛化实验中，平均精度均值分别达到 81.2%和 44.2%，推理速度达到 34.9 帧/s。所提方法可兼顾检测精度、开放类别泛化能力和推理效率，为配电线路金具缺陷检测提供技术支撑。
关键词：配电线路；缺陷检测；视觉语言模型；领域知识；低对比特征增强；跨模态对齐；开放词汇检测

Domain Knowledge-guided Vision-language Model for Corrosion Detection of Distribution Line Fittings

ZHAO Zhenbing[1,2,3,4], TIAN Benxi[1], GAO Kaipeng[1], TANG Chenkang[1], LI Haopeng[1,2,3]
(1. Department of Electronic and Communication Engineering, North China Electric Power University, Baoding 071003, China; 2. Hebei Key Laboratory of Power Internet of Things Technology (North China Electric Power University), Baoding 071003, China; 3. Hebei Engineering Research Center of Intelligent Technology for Power Internet of Things, North China Electric Power University, Baoding 071003, China; 4. Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, North China Electric Power University, Baoding 071003, China)

Abstract: To address the issues such as difficulty in defect representation, insufficient semantic information, and ambiguous class boundaries in corrosion detection of distribution-line fittings, a domain knowledge-guided vision-language detection method is proposed. First, a low-contrast feature enhancement module is designed to enhance textures, boundaries, and key channel responses of low-contrast corrosion regions through multi-scale directional extraction, boundary-guided modulation, and channel adaptive recalibration. Second, a knowledge-guided representation module is constructed to transform distribution-line domain knowledge into textual prompts and to dynamically optimize text embeddings with a vision-conditioned prompt adapter, improving the model's semantic understanding of corrosion patterns and fitting states. Finally, a cross-modal semantic alignment strategy is introduced to constrain visual-semantic consistency through region-text contrastive classification loss and semantic alignment regularization, enhancing discrimination of similar categories. The experimental results show that the proposed method achieves 87.1% mean average precision (mAP50) on a self-built defect dataset of distribution-line fittings. In open-semantic detection and zero-shot generalization experiments, the mAP50 values reach 81.2% and 44.2%, respectively, with an inference speed of 34.9 frames per second. The proposed method can achieve a balance among detection accuracy, open-category generalization and inference effi-

基金资助项目：国家自然科学基金(62571189；62373151；62371188；62303184)；中央高校基本科研业务费专项资金(2023JC006；2025MS118)。
Project supported by National Natural Science Foundation of China (62571189, 62373151, 62371188, 62303184), Fundamental Research Funds for the Central Universities (2023JC006, 2025MS118).

ciency, providing a technical support for defect detection of distribution-line fittings.

Key words: distribution lines; defect detection; vision-language model; domain knowledge; low-contrast feature enhancement; cross-modal alignment; open-vocabulary detection

0 引言

在新型电力系统建设背景下，电力视觉感知与无人机智能巡检已成为配电设备运行状态精准感知和智能运维的重要技术支撑[1-2]。配电线路设备数量庞大、分布广泛，长期服役于风吹、雨淋、污秽和腐蚀等复杂环境中，容易产生锈蚀、破损和部件缺失等多类缺陷。金具作为配电线路中的关键连接与承载部件，其运行状态直接关系到线路结构安全和供电可靠性。若金具缺陷不能被及时发现并处理，可能引发连接失效、机械强度下降及运行风险增大等问题。因此，面向复杂巡检环境开展高精度、高效率的配电线路金具缺陷检测研究，具有重要的理论意义和工程应用价值[3-4]。

近年来，随着无人机巡检和电力视觉技术的发展，基于深度学习的目标检测方法已广泛应用于输配电线路设备识别与缺陷检测任务中[5]。从早期基于卷积特征的快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)、单次前向目标检测网络(you only look once, YOLO)、单次多框检测器(single shot multibox detector, SSD)系列检测器，到近年来基于注意力机制的检测变换器(detection transformer, DETR)及其改进模型，目标检测技术在检测精度和推理效率方面不断提升[6-9]。在电力巡检场景中，已有研究围绕输配电线路部件识别、缺陷分类和复杂场景适应开展了较多探索。文献[10]将双注意力机制嵌入 Faster R-CNN，用于航拍输电线路螺栓缺陷检测，增强了模型对复杂背景中螺栓缺陷区域的关注能力，但该方法仍属于典型的封闭集检测框架，对类别语义关系和未知缺陷泛化能力考虑不足。文献[4]提出基于共有特征评分的解耦知识蒸馏方法，用于输电线路金具检测，通过教师—学生网络中的知识迁移提高了检测模型的轻量化部署能力，但其重点在于部件目标检测效率提升，对锈蚀等弱显著缺陷的低对比纹理和边界扩散特征建模仍不充分。文献[11]将剪枝与知识蒸馏相结合，面向配电线路设备部件检测实现了模型压缩和性能保持，为巡检模型端侧部署提供了有效思路，但轻量化模型在处理低对比、弱纹理和遮挡干扰目标时，仍可能存在局部细节特征表达不足的问题。

针对电力部件缺陷识别任务，部分研究进一步引入多模态信息、缺陷先验知识或两阶段检测机制，以提升模型对复杂缺陷形态的识别能力。文献[12]利用多模态对比学习开展输电线路螺栓缺陷分类，通过图像特征与语义信息之间的对比约束增强了缺陷类别判别能力，但其任务主要面向图像级分类，难以直接满足配电线路金具锈蚀场景中对缺陷区域精确定位的需求。文献[13]提出锈蚀知识引导的配电线路金具及其缺陷双阶段检测方法，将锈蚀先验知识引入检测流程，有助于提升金具及锈蚀缺陷的识别效果。然而，该类方法仍主要依赖封闭类别监督和阶段式检测流程，对开放语义描述、未见类别泛化以及细粒度锈蚀文本语义的利用仍然有限。总体来看，现有电力巡检检测方法在部件识别、轻量化部署和特定缺陷检测方面已取得进展，但多数方法仍主要依赖视觉外观差异进行判别。当配电线路金具锈蚀区域呈现低对比度、弱纹理、边界模糊和腐蚀扩散等特征时，缺陷响应容易被复杂背景、光照变化或金具本体结构淹没，从而导致轻微锈蚀漏检、正常金具误判以及相似缺陷类别混淆等问题。

除视觉表征不足外，语义监督信息有限也是制约配电线路金具锈蚀检测性能的重要因素。传统目标检测方法通常采用固定类别标签进行训练，类别标签多为离散符号，难以充分刻画“轻度锈蚀”、“锈蚀扩散”、“结构破损”、“保护罩缺失”等具有专业属性的细粒度缺陷信息。为缓解单一视觉监督中语义表达不足的问题，视觉语言模型通过将图像区域特征与文本语义嵌入映射到统一特征空间，为目标检测提供了更强的语义建模能力和开放词汇泛化能力[14]。在开放词汇目标检测领域，OVR-CNN 通过利用图像描述信息扩展检测类别，使模型能够从文本语义中获得未见类别知识[15]；ViLD 将视觉语言模型中的语义知识蒸馏到目标检测器中，提升了模型对开放类别的识别能力[16]；GLIP 将目标检测与短语定位统一到语言—图像预训练框架中，通过大规模图文数据学习区域与文本之间的对应关系，增强了检测模型的语言理解能力[17]；Grounding DINO

通过语言引导查询选择和跨模态特征交互提升了开放集目标检测性能[18]。这些方法推动了目标检测从封闭类别识别向开放语义理解拓展，但其预训练数据多来自自然场景图文数据，对配电线路金具锈蚀、腐蚀扩散和结构异常等专业概念的语义表达仍存在领域偏差；同时，部分高性能视觉语言检测模型依赖深层跨模态融合或复杂解码结构，计算开销较大，难以满足配电线路巡检对实时检测和工程部署的需求。

为兼顾开放语义检测能力与推理效率，YOLO-World 等方法将 YOLO 系列检测器与视觉语言建模相结合，通过区域-文本对比学习实现实时开放词汇检测[19-20]。WeDetect 进一步将开放词汇检测视为高效检索过程，在保持双塔架构推理效率的同时提升了开放类别检测能力[21]；LLMDet 则利用大语言模型监督增强开放词汇检测器的类别语义学习能力[22]。然而，上述方法主要面向通用视觉场景，缺乏针对电力巡检领域知识的显式建模。在配电线路金具锈蚀检测任务中，锈蚀区域往往同时具有低对比视觉特征和专业语义属性，若仅依靠通用文本提示或固定文本嵌入，模型难以准确理解“腐蚀产物”“锈蚀斑块”等领域概念，也难以在轻微锈蚀与正常金具之间形成稳定的语义判别边界。因此，有必要将配电工程领域知识转化为文本提示，并结合视觉区域特征对文本嵌入进行动态优化，从而实现视觉特征与专业语义信息的有效对齐。

综上所述，配电线路金具锈蚀缺陷检测仍面临 3 方面关键挑战。(1) 缺陷表征困难。锈蚀区域通常呈现低对比度、弱纹理和边界不清等特征，在复杂背景、光照变化及遮挡干扰下，其局部视觉响应容易被周围结构淹没，导致模型难以稳定捕获有效缺陷特征。(2) 语义信息不足。传统检测方法大多依赖固定类别标签进行监督，难以充分表达锈蚀程度、扩散形态和结构状态等细粒度语义信息，限制了模型对专业缺陷知识的理解与利用。(3) 类别边界模糊。轻度锈蚀与正常金具、不同锈蚀阶段之间往往仅存在细微外观差异，少数类缺陷样本数量有限又进一步加剧了模型对多数类的学习偏置，使得模型在复杂场景下容易出现误检和漏检。因而，如何从视觉表征增强、领域语义注入、跨模态判别约束等层面协同提升模型性能，成为配电线路金具锈蚀检测亟待解决的关键问题。

针对上述挑战，本文借鉴双塔式视觉语言检测架构的思想，提出一种领域知识引导的配电线路金具锈蚀检测方法(domain knowledge-guided detection, DKDetect)。该方法在保持高效推理的前提下，从视觉增强与语义约束两个维度协同优化。首先，设计低对比特征增强模块，通过多尺度方向特征提取、边界引导调制和通道自适应重标定，增强低对比锈蚀区域的视觉表征能力；其次，构建知识引导表征模块，将配电领域知识转化为文本提示，并利用视觉条件提示适配器实现文本嵌入的动态优化，提升模型对锈蚀语义的理解能力；最后，引入跨模态语义对齐策略，从整体分布层面约束视觉特征与文本语义的一致性，缓解类别混淆问题，提高检测结果稳定性。实验结果表明，所提方法在闭集、开放语义和开放词汇泛化任务中均优于基线模型，能够在保持较高实时性的同时有效提升配电线路金具缺陷检测精度。

本文的主要贡献如下：

1）提出一种面向配电线路金具锈蚀检测的领域知识引导视觉语言检测框架，在保持双塔架构高效推理能力的同时，提升复杂巡检场景下金具锈蚀缺陷的检测精度。

2）设计低对比特征增强模块，通过多尺度方向特征提取、边界引导调制和通道自适应重标定，增强低对比锈蚀区域的结构表征能力。

3）构建知识引导表征模块和跨模态语义对齐策略，通过领域知识文本提示、动态文本嵌入优化和语义分布约束，提升模型的专业语义理解能力与跨模态判别稳定性。

1 本文方法与网络架构

1.1 配电线路金具缺陷视觉语言检测框架

如图 1 所示，本文提出一种面向配电线路金具锈蚀检测任务的领域知识驱动视觉语言目标检测框架 DKDetect，旨在解决复杂电力场景下金具锈蚀检测面临的缺陷表征困难、语义信息不足和类别边界模糊等关键问题。该框架采用了双塔体系结构，其中语言分支以 XLM-RoBERTa[23]作为编码器初始化模型参数，视觉分支则由 ConvNeXt[24]主干网络、CSPRepBiFPAN Neck[25] 特征聚合结构和 YOLO-World[19]对比检测头组成。

低对比度特征增强模块(dual-level low-contrast feature enhancement module, DLFE)处于主干网络和候选区域生成阶段中间，其目的是提升低对比锈蚀区域的特征表征能力。该模块利用多尺度方向卷积

来提取锈蚀纹理的各向异性结构特征，并利用边界引导机制对原始特征进行自适应增强。知识引导表征模块(knowledge-guided representation module, KGRM)将领域知识注入到视觉语言对齐过程中，并提出视觉条件提示适配器实现动态提示优化。候选区域的视觉特征要经由轻量级适配网络产生提示偏移量，这个偏移量会针对原始文本嵌入进行内容感知式调整，使得语义表征能随图像特征自适应更新，以此提升类别判别的语义一致性。在对齐阶段，通过跨模态语义对齐策略(cross-modal semantic alignment strategy, CMSA)将视觉区域特征与文本嵌入被映射到共享嵌入空间中，通过区域-文本对比分类损失实现语义匹配约束。在保留原有损失的基础上，引入语义对齐正则项提高类别预测的稳定性。DKDetect 在颈部网络并未采用额外的融合层，其分类过程仍通过图像特征和文本嵌入间的点积实现。这样一种轻量化跨模态交互形式既保证了推理效率，又实现了领域知识引导下的动态语义增强。

DKDetect 保留了双塔体系结构原有的高效性，通过 DLFE 和 KGRM 实现结构增强和语义约束的有机统一，构建一种配电线路金具锈蚀缺陷检测的视觉语言模型检测架构。

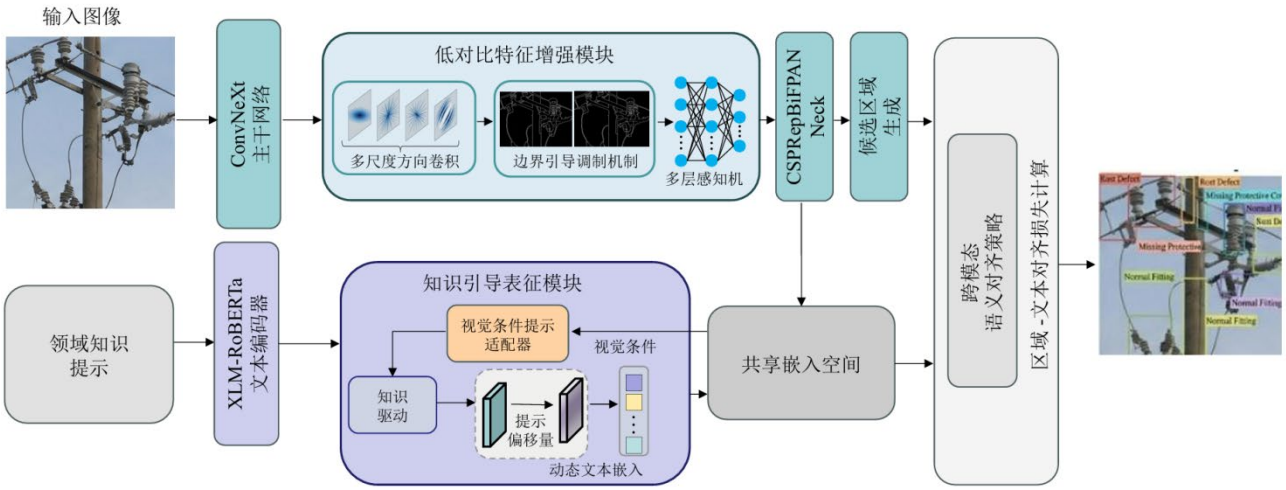


图 1　本文模型整体网络架构

Fig.1　Overall network architecture of the proposed model

1.2　低对比特征增强模块

在配电线路金具锈蚀缺陷检测任务中，锈蚀缺陷在实际巡检环境中易受光照变化、背景干扰等因素影响，通常具有低对比度、边界模糊等特点。锈蚀与正常金具表面灰度差异小，这就导致两者在深层特征空间中的差异不显著。

针对上述问题，本文提出低对比特征增强模块，用于在视觉特征提取阶段提升低对比腐蚀区域特征提取能力。如图 2 所示，DLFE 整体结构由多尺度方向特征提取单元，边界特征提取增强和通道自适应重标定单元 3 部分组成，其被嵌入于主干网络输出与候选区域生成模块之间，在不改变整体双塔检测结构情况下，仅对视觉分支进行增强处理。图中，Conv 3×3、Conv 5×5、Conv 7×7 和 Conv 1×1 分别表示对应卷积核尺寸的卷积操作；GAP、MLP、Σ 和 ⊙ 分别表示全局平均池化、多层感知机、逐元素求和和按空间位置广播后的逐元素乘法。

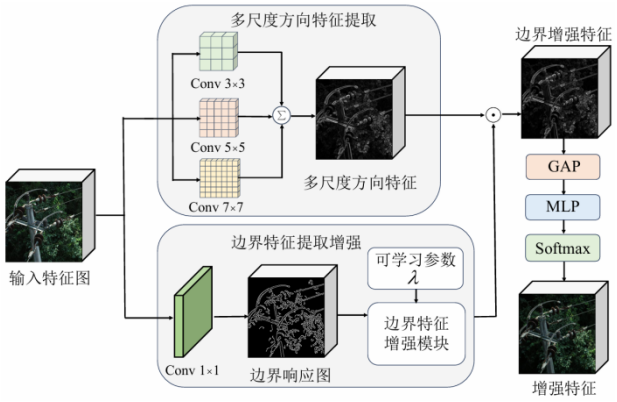


图 2　低对比特征增强模块

Fig.2　Dual-level low-contrast feature enhancement module

设主干网络在第 l 层输出的特征图为：

$$\boldsymbol{F}^{(l)} \in \mathbf{R}^{C_l \times H_l \times W_l} \tag{1}$$

式中：C 表示特征图的通道数；H 表示特征图高度；W 表示特征图宽度。

对于单尺度输入特征 $\boldsymbol{F}$，DLFE 的总体结构可以形式化表示为：

$$\boldsymbol{F}_{\mathrm{v}} = \boldsymbol{F}_{\mathrm{enh}}\left(\boldsymbol{F}_{\mathrm{ms}}(\boldsymbol{F})\right) \tag{2}$$

式中：$\boldsymbol{F}_{\mathrm{ms}}$ 表示经过多尺度卷积融合后的特征；$\boldsymbol{F}_{\mathrm{enh}}$ 表示经过边界调制增强后的特征；$\boldsymbol{F}_{\mathrm{v}}$ 表示最终输出的增强视觉特征。

为捕获锈蚀区域在不同尺度和方向上的结构差异，DLFE 采用不同感受野大小的卷积核对输入特征进行并行处理。首先，在输入特征 $\boldsymbol{F}$ 上采用不同核尺寸的卷积操作，即：

$$\boldsymbol{F}_k = \mathrm{Conv}_k(\boldsymbol{F}), \quad k \in \{3,5,7\} \tag{3}$$

式中：$\mathrm{Conv}_k(\cdot)$表示核尺寸为 $k\times k$ 的卷积操作；$\boldsymbol{F}_k$ 表示经过不同尺寸卷积核处理后的特征，较小卷积核用于提取细粒度纹理变化，而较大卷积核用于捕获更广域的结构差异。

随后对多尺度特征进行融合：

$$\boldsymbol{F}_{\mathrm{ms}} = \sum_k \boldsymbol{F}_k \tag{4}$$

通过该多尺度特征叠加机制，可以增强锈蚀区域在不同尺度上的响应强度，提高模型对微弱纹理变化的辨识能力。在实际检测过程中，锈蚀缺陷往往会和金具表面腐蚀扩散边缘相关联。为提升模型对边界信息的关注度，DLFE 引入了边界预测分支。首先通过一层卷积生成边界响应图：

$$\boldsymbol{B} = \sigma(\mathrm{Conv}_b(\boldsymbol{F})) \tag{5}$$

式中：$\sigma(\cdot)$为 Sigmoid 激活函数；$\mathrm{Conv}_b(\cdot)$ 表示边界预测分支中的卷积操作；b 表示边界分支；$\boldsymbol{B} \in [0,1]^{H\times W}$ 表示空间维度上的边界响应图；H 和 W 分别表示特征图的高度和宽度。

随后利用边界响应图对多尺度特征进行调制：

$$\boldsymbol{F}_{\mathrm{enh}} = \boldsymbol{F}_{\mathrm{ms}} \odot (1 + \lambda \boldsymbol{B}) \tag{6}$$

式中：λ 为可学习参数，用于控制边界增强幅度；$\odot$ 表示按空间位置广播后的逐元素乘法。

为了进一步增强关键特征通道的表达能力，DLFE 在空间增强之后引入通道重标定机制。首先通过全局平均池化获取通道统计信息：

$$z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} \boldsymbol{F}_{\mathrm{enh}}(c,i,j) \tag{7}$$

式中：c 是通道索引；i 和 j 分别表示特征图在高度和宽度方向上的空间位置索引；$\boldsymbol{F}_{\mathrm{enh}}(c,i,j)$ 表示增强后的特征图在第 c 个通道、位置(i,j)处的特征值；z_c 表示第 c 个通道的全局统计量，随后进行通道转置并将各通道统计量组成通道统计向量：

$$\boldsymbol{z} = [z_1, z_2, \cdots, z_c]^{\mathrm{T}} \tag{8}$$

式中：$\boldsymbol{z}$ 表示通道统计向量。

然后通过多层感知机(multilayer perceptron, MLP)生成通道权重向量：

$$\boldsymbol{w} = \mathrm{Softmax}(\mathrm{MLP}(\boldsymbol{z})) \tag{9}$$

式中：$\boldsymbol{w}$ 表示通道权重向量；MLP(·) 表示多层感知机；Softmax(·) 表示 Softmax 函数。

最终得到增强特征为：

$$\boldsymbol{F}_{\mathrm{v}} = \boldsymbol{w} \odot \boldsymbol{F}_{\mathrm{enh}} \tag{10}$$

该步骤能抑制冗余特征通道，提高与锈蚀缺陷相关通道的响应强度。

本文提出的低对比特征增强模块通过多尺度方向特征提取、边界特征提取增强以及通道自适应重标定对视觉特征进行结构化增强。该模块有效提升了锈蚀区域在特征空间里的可区分性，为后续工作提供了更加稳定的视觉表征基础。

1.3 知识引导表征模块

在配电线路金具锈蚀缺陷检测任务中，数据标注往往只有“锈蚀”和“正常”两种标签。模型学习目标需依靠视觉外观差别来进行特征鉴别。实际上，锈蚀缺陷形态多变，仅靠视觉特征进行检测易受光照、背景干扰，导致特征表达不稳定。

视觉语言模型为目标检测注入领域知识提供了新的技术思路。通过自然语言描述锈蚀缺陷，将其文本特征和视觉特征投影到共享特征空间中，从而加强模型对锈蚀缺陷的理解能力。

基于此，如图 3 所示，本文提出了知识引导表征模块(knowledge-guided representation module, KGRM)，该模块既没有改动检测任务原先设定的标签，又在文本特征方面加入知识引导信号，从而实现对锈蚀缺陷特征的稳定理解。KGRM 将金具锈蚀缺陷的领域知识转为成自然语言形式的文本提示，依照腐蚀程度、结构状态或者缺陷表现形式来形成描述性语句，把原本抽象的类别标签拓展成带有语义信息的文本表达。设缺陷类别集合为 $Q = \{q_1, q_2, \cdots, q_N\}$，其中 N 为类别总数。每个类别都会经过知识库获取对应的文本提示 T_i。

$$T_i = \{T_i^1, T_i^2, \cdots, T_i^K\} \tag{11}$$

式中：T_i 表示第 i 个类别的文本提示集合；T_i^K 表示第 i 个类别的第 K 条文本提示。

随后，将提示集合中的每条文本提示送进语言编码器中，获得细粒度的文本嵌入表示为：

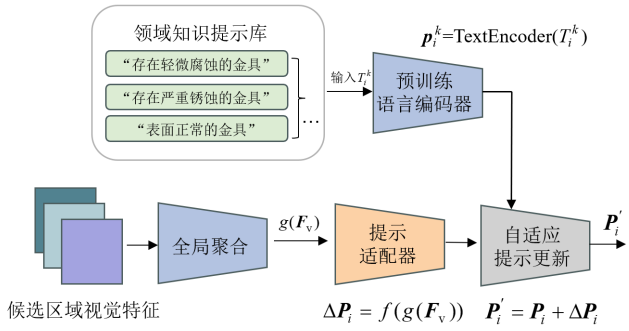


图 3 知识引导表征模块

Fig.3 Knowledge-guided representation module

$$\boldsymbol{p}_i^k = \text{TextEncoder}(T_i^k),\quad k=1,2,\cdots,K \tag{12}$$

式中：TextEncoder(·)表示文本编码器；$\boldsymbol{p}_i^k$ 表示第 i 个类别的第 k 条提示对应的文本嵌入表示。

最终，为了融合库内多维语义信息，本文引入加权聚合机制来生成该类别的语义原型 $\boldsymbol{P}_i$：

$$\boldsymbol{P}_i = \sum_{k=1}^{K} \alpha_i^k \boldsymbol{p}_i^k \tag{13}$$

式中：$\boldsymbol{P}_i$ 表示类别 q_i 的语义原型；α_i^k 表示第 k 条提示的聚合权重。本文采用平均聚合方式，即 $\alpha_i^k = 1/K$。

通过这种显式的原型聚合机制，原本抽象、孤立的类别标签被成功扩展为富含多维形态特征的静态语义空间。然而，在复杂场景下，不同图像样本中缺陷表现形式存在差异，若采用固定文本嵌入进行匹配，可能导致语义对齐不充分。为此，KGRM 引入视觉条件提示适配机制，使文本嵌入能够根据当前候选区域特征进行自适应更新。

因为配电线路锈蚀缺陷存在类内差异大的特点，采用固定文本嵌入会导致视觉语言匹配度差。所以，KGRM 采用了视觉条件提示适配机制，使得文本嵌入根据候选区域的特征进行更新。经视觉增强模块输出的候选区域特征为：

$$\boldsymbol{F}_v \in \mathbf{R}^{d_v} \tag{14}$$

式中：d_v 表示视觉特征向量的维度。

首先对视觉特征进行全局汇聚，得到紧凑表示向量 $g(\boldsymbol{F}_v)$。随后通过轻量级多层感知机生成提示偏移量 $\Delta\boldsymbol{P}_i$：

$$\Delta\boldsymbol{P}_i = f(g(\boldsymbol{F}_v)) \tag{15}$$

式中：$g(\cdot)$ 表示对候选区域视觉特征的全局汇聚操作；$f(\cdot)$ 表示适配网络。

动态文本嵌入 $\boldsymbol{P}_i'$ 表示为：

$$\boldsymbol{P}_i' = \boldsymbol{P}_i + \Delta\boldsymbol{P}_i \tag{16}$$

KGRM 在保持原有检测框架简洁性的前提下，实现了跨模态信息融合，有效提升了模型的语义对齐能力和领域适配能力，为配电线路金具锈蚀缺陷检测任务提供了可解释性的语义建模机制。一方面，领域知识经过文本化处理之后，类别语义空间得到扩充，模型就能学会类别间潜藏的语义关系。另一方面，视觉条件提示适配机制优化了文本嵌入的动态适应能力，提高了语义对齐能力。

1.4 跨模态对齐策略

在跨模态对齐阶段，模型的关键目标是创建视觉区域特征和文本语义嵌入间的稳定映射联系。针对这一目标，本文引入了跨模态语义对齐策略(cross-modal semantic alignment strategy，CMSA)分支所输出的候选区域特征为 $\boldsymbol{F}_v \in \mathbf{R}^{d_v}$，经过知识引导后的文本嵌入为 $\boldsymbol{P}_i' \in \mathbf{R}^{d_t}$，通过线性投影层将其映射至嵌入空间。其中，$d_t$ 表示文本嵌入向量的维度。引入视觉投影矩阵 $\boldsymbol{W}_v \in \mathbf{R}^{d\times d_v}$ 与文本投影矩阵 $\boldsymbol{W}_t \in \mathbf{R}^{d\times d_t}$，其中，$d$ 表示共享嵌入空间的统一特征维度，分别对视觉特征与文本嵌入进行变换：

$$\boldsymbol{Z}_v = \boldsymbol{W}_v\boldsymbol{F}_v,\quad \boldsymbol{Z}_{t,i} = \boldsymbol{W}_t\boldsymbol{P}_i' \tag{17}$$

式中：$\boldsymbol{Z}_v, \boldsymbol{Z}_{t,i} \in \mathbf{R}^d$ 表示映射至统一维度 d 的跨模态特征表示。

为消除不同样本间向量模长差异对相似度计算的影响，通常对投影后的特征进行 L2 归一化处理，从而保证相似度计算具有稳定数值范围。在此基础上，视觉区域特征与文本嵌入之间的匹配程度通过归一化点积进行度量，即余弦相似度：

$$S_i = \frac{\boldsymbol{Z}_v^{\mathrm{T}}\boldsymbol{Z}_{t,i}}{\|\boldsymbol{Z}_v\|_2\|\boldsymbol{Z}_{t,i}\|_2} \tag{18}$$

每个候选区域都要同所有类别文本嵌入进行相似度计算，由此产生类别得分向量，并通过区域—文本对比损失进行优化。视觉区域特征会逐渐趋近与其对应的文本嵌入，从而在嵌入空间中形成清晰的类别分布。假设真实类别标签为 y，那么分类损失就可以表示成：

$$L_{\text{cls}} = -\ln\frac{\exp(S_y/\tau)}{\sum_{i=1}^{N}\exp(S_i/\tau)} \tag{19}$$

式中：τ 为温度系数；N 表示类别总数。

引入语义对齐正则项，约束视觉特征分布和文本语义分布二者整体结构的一致性，而非仅着眼于

单个样本间的匹配关系，通过最大化其余弦相似度来实现：

$$L_{\text{align}}=\frac{1}{M}\sum_{j=1}^{M}\| \boldsymbol{Z}_{\text{v}}^{(j)}-\boldsymbol{Z}_{\text{t},y}^{(j)}\|_2^2 \tag{20}$$

式中：M 表示批次中正样本区域数量；$\boldsymbol{Z}_{\text{v}}^{(j)}$ 表示第 j 个正样本区域映射后的视觉跨模态特征表示；$\boldsymbol{Z}_{\text{t},y}^{(j)}$ 表示第 j 个区域真实类别对应的文本嵌入。该正则项从整体分布角度约束跨模态特征的一致性，有助于减小类别间边界模糊问题，提高模型在复杂背景下的判别稳定性。

最终，模型的总体优化目标函数可表示为：

$$L=L_{\text{cls}}+L_{\text{bbox}}+\gamma L_{\text{align}} \tag{21}$$

式中：γ 为语义对齐正则项的权重系数，其是随网络训练共同优化的自适应可学习参数，用于动态调节语义对齐正则项在总损失中的作用强度。实验中，γ 初始化为 0.1，并在训练过程中通过反向传播自动更新。L_{bbox} 表示目标检测中的边界框回归损失。跨模态语义对齐策略示意图如图 4 所示。

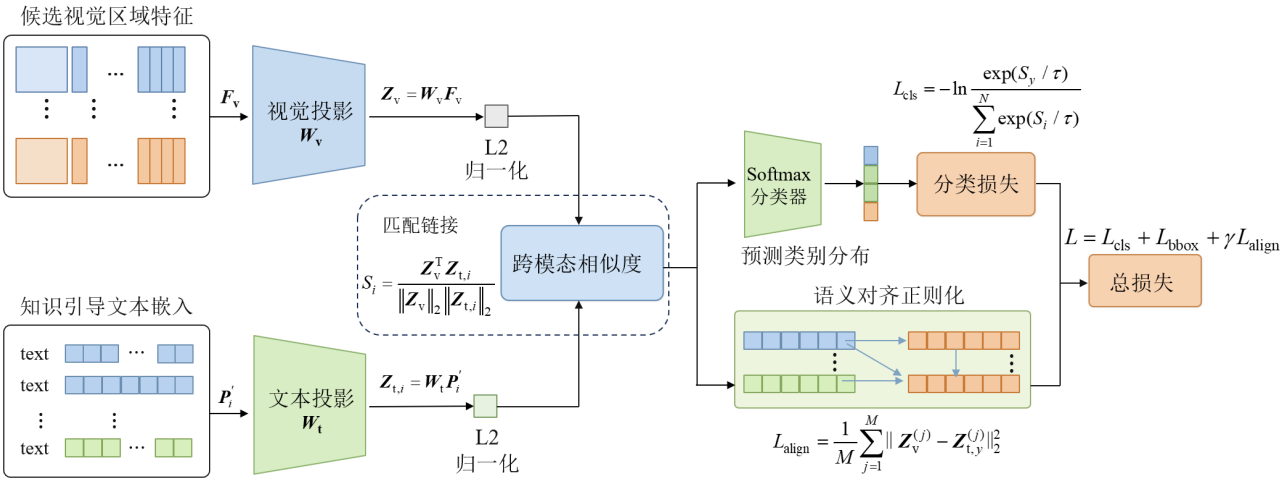


图 4 跨模态语义对齐策略示意图

Fig.4 Cross-modal semantic alignment strategy

跨模态对齐机制不仅能够提升类别预测的判别能力，还能够增强模型在小样本和开放语义场景下的鲁棒性。其原因在于，文本语义空间中不同类别之间并非相互孤立，而是存在一定的结构关系和语义相似性。通过引入语义对齐正则项，模型能够在视觉特征与文本特征之间建立更加一致的对应关系，使同类样本在嵌入空间中更加聚集，不同类别样本之间保持更清晰的语义间隔。相比仅依赖视觉特征进行分类的方式，该机制能够利用类别描述中的先验语义信息，缓解小样本条件下视觉表征不足和类别边界模糊的问题。同时，语义对齐约束有助于形成更平滑、可解释的特征分布，降低背景干扰或相似金具类别造成的误检风险，从而为配电线路金具锈蚀缺陷检测提供更稳定、可信的判别依据。

2 实验结果分析

2.1 实验环境设置

本文实验环境见表 1，模型设置的批次大小为 8，初始学习率为 0.01，训练 100 个轮次。

表 1 实验环境设置

Table 1 Experimental environment

名称	型号
操作系统	Ubuntu20.04.6
CPU	Silver 4210R CPU
GPU	RTX 3080
CUDA	11.6
Python	3.8
PyTorch	1.13.1

2.2 数据集及评价指标

本文所用数据均来源于某巡检部门近几年采集的配电线路巡检图像，具有较强的工程实际背景，具体数据集划分如表 2。为满足目标检测任务需求，本文对图像中的金具目标及其缺陷进行了人工标注，并统一整理为上下文通用目标数据集(common objects in context，COCO)格式。该数据集共包含 1 503 张图像、6 765 个标注实例，按照 7:3 的比例划分为训练集和验证集，其中训练集包含 1 052 张图像，验证集包含 451 张图像。COCO 格式数据集

主要用于闭集条件下的检测性能对比实验和消融实验，可为模型训练提供标准化的边界框与类别监督信息。

表 2 配电线路金具及其缺陷检测数据集

Table 2 Dataset for power distribution line fittings and their defect detection

配电线路金具	实例数量
螺栓型耐张线夹正常	487
拉杆型耐张线夹正常	2 482
螺栓型耐张线夹锈蚀	740
拉杆型耐张线夹锈蚀	1 956
螺栓型耐张线夹保护罩缺失	1 100

为验证模型在复杂自然语言描述下的跨模态语义对齐能力，本文构建了目标检测与视觉定位(object detection and visual grounding，ODVG)格式数据集用于开放语义标注验证实验。与传统 COCO 格式采用孤立、离散的类别编号不同，ODVG 格式在边界框标注的基础上，为每个实例赋予了包含丰富物理属性的自然语言描述文本，从而构成了高质量的视觉—语言配对监督信号。该数据集的图像来源与划分比例与前述 COCO 数据集保持一致，其核心目的在于评估模型能否有效利用文本空间中的领域先验知识替代传统的线性分类器。两种格式的标注结构对比如表 3 所示。需要特别指出，表 3 仅列举了各类别标注文本的部分典型示例。在实际构建 ODVG 数据集时，本文依托配电领域专家先验建立了一个多样化的领域知识提示库，为同一类别构建了涵盖不同锈蚀程度、扩散形态和结构特征的丰富描述样本，以确保模型语义学习的全面性与鲁棒性。

表 3 ODVG 格式标注文本典型示例

Table 3 Typical annotation examples in ODVG format

类别	标注文本示例
螺栓型耐张线夹正常	螺栓型耐张线夹表面光洁完好，螺栓紧固无松动
拉杆型耐张线夹正常	拉杆型耐张线夹结构完整，表面无异常
螺栓型耐张线夹锈蚀	螺栓型耐张线夹表面出现红棕色腐蚀产物
拉杆型耐张线夹锈蚀	拉杆型耐张线夹表面有锈蚀斑块
螺栓型耐张线夹保护罩缺失	螺栓型耐张线夹保护罩缺失，螺栓裸露

本文采用目标检测领域的标准评价指标：交并比(intersection over union，IoU)阈值为 0.5 时的平均精度均值(mean average precision at an IoU threshold of 0.5，mAP50)和 IoU 阈值为 0.5~0.95 时的平均精度均值(mean average precision，mAP)，作为模型检测精度的量化指标。模型推理速度通过每秒处理帧数(frames per second，FPS)衡量，模型规模通过参数量(number of parameters，Params)衡量。

2.3 消融实验

为验证 DKDetect 各模块的有效性，本节以基线模型为起点，在基线上分别单独引入 DLFE、KGRM 和跨模态对齐策略进行消融实验，实验结果如表 4 所示。

表 4 DKDetect 消融实验

Table 4 Ablation study of DKDetect

DLFE	KGRM	CMSA	mAP50/%	mAP/%
			84.2	62.5
√			85.3	63.9
	√		86.1	64.4
		√	84.9	63.3
√	√	√	87.1	66.5

引入 DLFE 使 mAP50 从 84.2%提升至 85.3%，mAP 从 62.5%提升至 63.9%，这种提升源于多尺度方向特征提取对锈蚀区域纹理结构的加强和边界特征提取增强对锈蚀边缘特征强化。同时，DLFE 凭借通道重标定机制抑制多余特征通道，加强金具锈蚀缺陷相关特征响应，进而为后续候选区域的生成提供了更具判别力的视觉表征基础。引入 KGRM，mAP50 值为 86.1%，mAP 值为 64.4%，其增益幅度在 3 个模块里最大。这说明将领域知识当作文本提示来运用，并凭借视觉条件提示适配器进行动态优化，对加强模型的语义判断能力有着最明显的效应。一方面，领域知识文本提示把原本只是带有“锈蚀”“正常”等抽象类别标签的语义空间扩展成带有腐蚀程度、外观特征等丰富语义信息的向量空间，从而让模型在该嵌入空间里形成类别间的结构联系；另一方面，视觉条件提示适配器能让文本嵌入按照当下候选区域的视觉特征自动调整更新，防止了固定文本嵌入在复杂情况下存在语义对齐不足的情况发生。KGRM 的增益高于 DLFE，这说明在当前数据集中语义方面的限制对于改善检测性能的作用比视觉层面的特征优化更明显。引入跨模态对齐策略，mAP50 值为 84.9%，mAP 值为 63.3%，增益比较小。原因在于语义对齐正则项大多在整体分布上约束视觉特征和文本语义的一致性，该正则项的作用重点在于使类别边界更清晰并抑制误检，并非直接改善特征表征能力。若没有 DLFE 提供的高质量视觉特

征以及 KGRM 所提供的动态语义嵌入，则跨模态对齐策略的潜力也就无法得到完全发挥。三者同时引入时，mAP50 和 mAP 分别为 87.1%和 66.5%。相较于基线模型，各自改进了 2.9%和 4.0%。

综合来看，DLFE 从视觉底层增强了锈蚀区域的结构显著性，为 KGRM 提供了更清晰的视觉特征输入；KGRM 进一步结合领域知识约束，使文本嵌入能够更准确地适配视觉特征；跨模态对齐策略则从特征分布层面强化视觉与语义的一致性。三者形成了“视觉增强—语义建模—对齐约束”的递进式协作链路。实验中 mAP 增益高于 mAP50，说明所提模块在更严格 IoU 阈值下仍能保持优势，不仅提升了目标定位精度，也改善了类别判别能力，对配电线路金具锈蚀缺陷的精确检测具有实际意义。

2.4 对比实验

为全面评估 DKDetect 的检测性能，本实验首先在 COCO 格式数据集上进行封闭集对比，结果如表 5 所示。传统单阶段检测器中 YOLOv9c 表现较好，其 mAP50 达到了 78.7%。在 Transformer 类检测器中，Deformable-DETR 经由可变形注意力机制把 mAP50 提升到 81.2%，但推理速度偏低，仅有 19.7 帧/s。视觉语言检测器 YOLO-World-L 和 YOLOE-8-L 虽取得较高 mAP50，但因缺乏领域知识致使检测能力受限。Grounding DINO 精度达到了 86.2%，但其 6.0 帧/s 的推理速度难以满足配电巡检的即时性需求。基线模型 WeDetect 的 mAP50 为 84.2%，推理速度达到 35.1 帧/s，兼顾了精度与效率。本文 DKDetect 在此基础上，mAP50 与 mAP 进一步提升至 87.1%和 66.5%。与 Grounding DINO 相比，DKDetect 不仅精度高出 0.9%，其 34.9 帧/s 的推理速度更是前者的 5.8 倍，实现了精度与效率的最佳平衡。

此外，针对实际巡检数据中存在的类别不均衡问题，表 6 给出了各类别独立检测指标。需要说明的是，实际配电线路巡检数据中，不同金具类型和缺陷类型的出现频率具有天然差异，类别不均衡本身反映了真实工程场景的数据分布特征。为保证实验结果对真实巡检场景的代表性，本文保留原始数据分布，并通过交并比阈值为 0.5 时的平均精度(average precision at an intersection over union threshold of 0.5，AP50)分析模型在不均衡条件下的检测能力。由表 6 可知，实例数量较少或外观差异较弱的目标检测难度相对更高。例如，YOLO-World-L 在螺栓型耐张线夹保护罩缺失类别上的 AP50 为 73.3%。相比之下，DKDetect 在该类别上的 AP50 达到 79.1%，较基线模型提升 4.0%；在螺栓型耐张线夹锈蚀类别上，也较基线模型提升 3.7%。这主要得益于 DLFE 对低对比缺陷特征的增强、KGRM 对领域语义信息的引入以及 CMSA 对类别边界的约束。上述结果验证了本文模型在类别不平衡场景下依然具备较强的鲁棒性。

表 5 目标检测模型闭集对比实验

Table 5 Comparative experiments of object detection models

算法	mAP50/%	mAP/%	FPS/(帧·s^{-1})	Params/10^6
YOLOv8s[26]	76.3	55.3	121.4	11
YOLOv9c[27]	78.7	57.4	99.7	25
YOLOv10s[9]	77.9	56.7	171.2	7
DETR[8]	78.5	49.2	28.6	41
DAB-DETR[28]	80.3	57.3	21.0	43
Deformable-DETR[29]	81.2	58.8	19.7	40
YOLO-World-L[19]	82.5	59.5	54.6	48
YOLOE-8-L[20]	83.0	60.2	55.9	45
Grounding DINO[18]	86.2	65.6	6.0	172
LLMDet[22]	85.5	64.8	6.0	172
基线模型[21]	84.2	62.5	35.1	176
本文模型	**87.1**	**66.5**	**34.9**	**176**

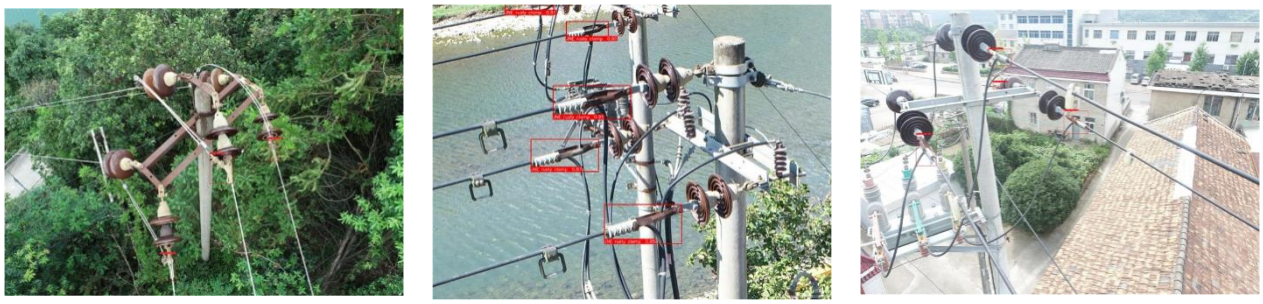
表 6 各类别独立检测精度(AP50)对比实验

Table 6 Per-category precision (AP50) comparison

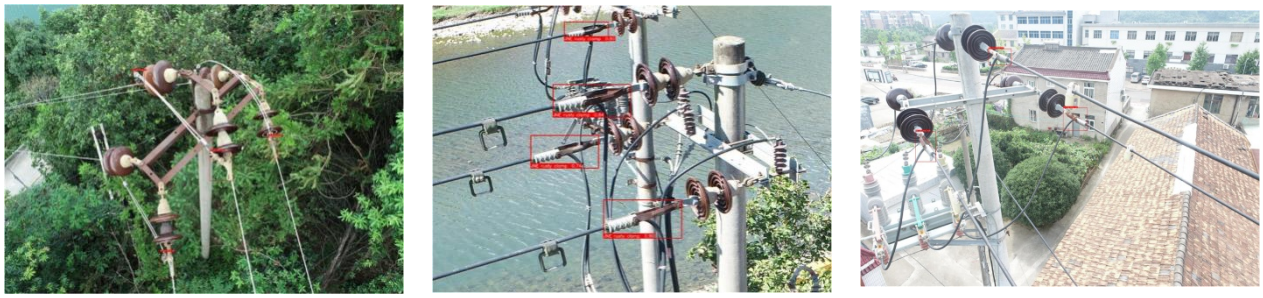
金具缺陷类别	AP50/%			
	YOLO-World-L	Grounding DINO	基线模型	本文模型
螺栓型耐张线夹正常	80.2	83.7	81.5	85.2
拉杆型耐张线夹正常	90.4	93.6	92.3	94.1
螺栓型耐张线夹锈蚀	77.1	81.2	78.6	82.3
拉杆型耐张线夹锈蚀	91.5	94.3	93.5	94.8
螺栓型耐张线夹保护罩缺失	73.3	78.2	75.1	79.1

2.5 可视化结果分析

为了进一步定性验证上述定量实验的结论，直观展示模型在复杂场景下的检测能力，图 5 给出了本文模型(DKDetect)与基线模型(WeDetect)的可视化对比结果。在图 5(a)、(d)对比中，该组锈蚀形态丰富多样，WeDetect 明显存在类别混淆的情况，把轻微锈蚀金具错判成正常金具。但 DKDetect 依靠 DLFE 模块，让低对比度的锈蚀区域在特征空间具有更高区分度，并未产生错检，漏检现象。在图 5(b)、



(a) 复杂植被背景，本文模型 (b) 密集金具场景，本文模型 (c) 建筑背景，本文模型



(d) 复杂植被背景，基线模型 (e) 密集金具场景，基线模型 (f) 建筑背景，基线模型

图 5 本文模型可视化结果

Fig.5 Visualization results of the proposed model

(e)对比中，图像边缘处存在缺陷目标，特征模糊，对比度低。WeDetect 对此类样本存在明显的漏检情况，特别未能检测到图像边缘的锈蚀目标。DKDetect 经由 DLFE 对区域特征执行结构化加强，再结合 KGRM 针对文本实施动态调整，即使在特征质量变差时，仍保住了视觉和语义的匹配精度，从而捕捉到边缘区域的锈蚀目标。在图 5(c)、(f)对比中，该组轻度锈蚀较多，其视觉特征和正常金具差别很小，WeDetect 对这种低显著性样本既会漏检又会错检。但 DKDetect 依靠 DLFE 捕捉细粒度视觉特征变化，再利用 KGRM 进行领域知识文本提示。在面对低显著性样本时，其可以凭借语义空间里的结构化联系来做出精准判断，跨模态对齐策略进一步稳定了视觉特征和文本语义的一致性，有效提升了低显著性样本的检测精度。综上所述，面对缺陷弱表征、语义信息弱和类内差异大 3 大问题，DKDetect 均保持了较为可靠的检测性能。

2.6 开放语义标注验证实验

为验证本文方法在开放语义标注下的跨模态对齐性能，本文基于 ODVG 格式数据集开展评估，并将其作为后续开放词汇检测的基础。由表 7 可知，相比闭集实验，各模型精度均出现不同程度下降，主要原因在于自然语言描述较固定类别标签具有更强的语义多样性和模糊性，模型需要从文本中提取更复杂的类别判别信息。YOLO-World-L 和 YOLOE-8-L 在开放语义场景下的 mAP50 分别为 72.5%和 74.1%，较闭集结果下降约 9%，说明其文本嵌入分类能力对自然语言表达变化的适应性仍有限。Grounding DINO 凭借深度跨模态融合取得 78.6%的 mAP50，性能优于非融合方法，但推理效率难以满足即时巡检需求；LLMDet 性能与 Grounding DINO 接近，但同样存在推理开销较大的问题。

WeDetect 基线模型在 ODVG 格式的数据集上，mAP50 为 76.8%，mAP 为 56.5%，其推理速度为 35.1 帧/s，这体现出双塔框架在开放语义场景下兼具精准度与效率的优势。DKDetect 以 WeDetect 作为基础，并利用带有领域知识的视觉优化和语义建模机制，其 mAP50 值升至 81.2%，mAP 值也增至 60.4%，分别比基线高出 4.4%和 3.9%，增幅高于闭集环境中的收益，这说明 DKDetect 所提出的知识引导表征模块在开放语义情形下起到更大作用，而且视觉条件提示机制有效改善了多元化自然语言表达而产生的语义对齐难题，使模型能从不同形式的表述更为稳定地获取类别信息。相较于 Grounding DINO，DKDetect 的 mAP50 高出 2.6%，而且推理速度为其 5.8 倍。这再次说明在未采用重型跨模态融合层时，经由领域知识提升达成高精度开放词汇检测是可行的。

表 7 开放语义目标检测对比实验

Table 7 Comparative experiments of object detection models

算法	mAP50/%	mAP/%	FPS/(帧·s^{-1})	Params/10^6
YOLO-World-L[19]	72.5	52.8	54.6	48
YOLOE-8-L[20]	74.1	54.3	55.9	45
Grounding DINO[18]	78.6	58.2	6.0	172
LLMDet[22]	77.3	57.1	6.0	172
基线模型[21]	76.8	56.5	35.1	176
本文模型	**81.2**	**60.4**	**34.9**	**176**

2.7 开放词汇泛化能力实验

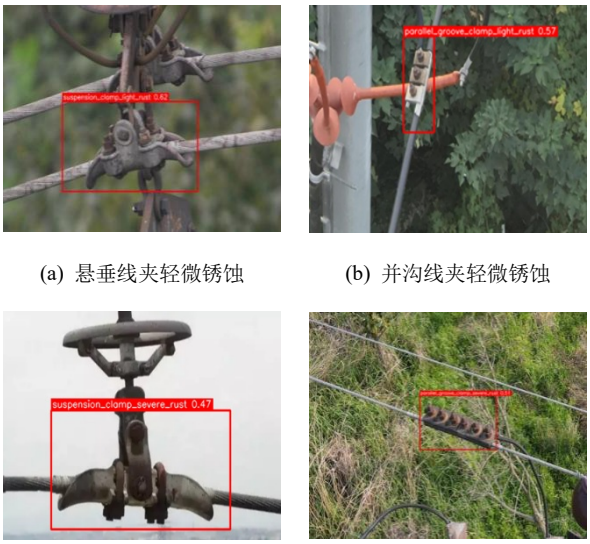
为进一步验证 DKDetect 的开放词汇泛化与细粒度语义理解能力，本实验额外收集了两种未参与训练的金具(悬垂线夹、并沟线夹)共计 410 张图像进行零样本测试。推理时引入“轻微锈蚀”与“严重锈蚀”的细粒度文本提示，mAP50 结果如表 8 所示。面对全新的金具结构，各模型精度相较于闭集均明显下降，体现了零样本细粒度检测的极大挑战。特别是在“轻微锈蚀”上，基线模型由于过度拟合训练集的外观特征，在新类别上泛化困难，如悬垂线夹轻微锈蚀仅 25.8%。通用大模型 Grounding DINO 也难以将宽泛的文本概念与具体未知金具的微小纹理相对齐，其轻微锈蚀精度仅在 28%以下。相比之下，本文的 DKDetect 在所有未见类别和细粒度划分上均取得最优表现，整体 mAP50 达到 44.2%，较基线模型提升了 5.5%，并超越 Grounding DINO 3.2%。这证明 DKDetect 能依靠 DLFE 模块提取通用的底层锈蚀纹理，并利用 KGRM 模块将细粒度的专家先验转化为文本提示，有效实现跨类别的动态语义迁移。即便在极易漏检的未见类别轻微锈蚀场景下，本文方法依然能保持 30%以上精度，验证了其在复杂开放场景下良好的泛化能力与鲁棒性。

表 8 开放词汇泛化能力实验

Table 8 Open-vocabulary generalization experiment

金具缺陷类别	AP50/%			
	YOLO-World-L	Grounding DINO	基线模型	本文模型
悬垂线夹轻微锈蚀	23.5	27.6	25.8	31.2
悬垂线夹严重锈蚀	50.3	53.7	51.5	56.4
并沟线夹轻微锈蚀	23.0	28.1	25.5	31.6
并沟线夹严重锈蚀	51.4	54.6	52.1	57.5

如图 6 可视化结果所示，对于未参与训练的悬垂线夹和并沟线夹(图 6(c)和 6(d))，模型能够较好地定位严重锈蚀区域，体现出一定的跨类别泛化能力。在轻微锈蚀场景下(图 6(a)和 6(b))，DKDetect 在领域知识文本的引导下，基本识别出了未知金具表面的局部锈蚀特征。需要指出的是，在图 6(a)的检测中，尽管模型标出了主要的锈蚀区域，但金具上方及边缘处的微弱锈蚀斑块仍存在局部漏检。这反映出在零样本且缺陷视觉显著性较低的条件下，实现锈蚀区域的完整召回仍具挑战。总体而言，模型在未见类别上提供了较为合理的缺陷定位，验证了本文方法在开放环境下的有效性。



(a) 悬垂线夹轻微锈蚀 (b) 并沟线夹轻微锈蚀

(c) 悬垂线夹严重锈蚀 (d) 并沟线夹严重锈蚀

图 6 开放词汇泛化能力实验可视化结果

Fig.6 Visualization results of open-vocabulary generalization

3 结论

1）本文针对配电线路金具锈蚀区域低对比、弱纹理和边界模糊导致的表征不足问题，设计了低对比特征增强模块，通过多尺度方向特征提取、边界引导调制和通道自适应重标定，增强锈蚀区域的纹理、边界及关键通道响应，提高了模型对轻微锈蚀和低显著性缺陷的特征提取能力。

2）针对固定类别标签语义有限及通用文本提示存在领域偏差的问题，构建知识引导表征模块，将配电工程领域知识转化为文本提示，并利用视觉条件提示适配器动态优化文本嵌入，提高了模型对锈蚀形态和金具状态的语义理解能力。

3）针对轻微锈蚀与正常金具类别边界模糊的问题，引入跨模态语义对齐策略，通过区域-文本对比分类损失和语义对齐正则项约束视觉特征与文本语义的一致性，增强相似类别的判别能力，形成“视觉增强—语义建模—对齐优化”的协同检测机制。

4）基于上述设计，DKDetect 在自建数据集上的 mAP50 和 mAP 分别达到 87.1%和 66.5%，较基线提高 2.9%和 4.0%，推理速度为 34.9 帧/s；在开放语义检测和零样本泛化实验中的 mAP50 分别达到 81.2%和 44.2%，验证了其在检测精度、泛化能力和推理效率方面的优势。未来将进一步研究细粒度区域文本对齐与小样本自适应方法，提升模型对未见类别轻微锈蚀及复杂巡检环境下缺陷的检测能力。

参考文献 References

[1] 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.

[2] 李道兴，王晓辉，李黎，等. 面向电力无人机巡检图像分析处理的自动化深度学习系统：架构设计与关键技术[J]. 电力信息与通信技术，2024，22(4)：38-54.
LI Daoxing, WANG Xiaohui, LI Li, et al. The architecture and key technologies of an automatic deep learning system for image analysis in UAV transmission line inspection[J]. Electric Power Information and Communication Technology, 2024, 22(4): 38-54.

[3] NGUYEN V N, JENSSEN R, ROVERSO D. Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning[J]. International Journal of Electrical Power & Energy Systems, 2018, 99: 107-120.

[4] 赵振兵，吕雪纯，苗思雨，等. 基于共有特征评分的解耦知识蒸馏输电线路金具检测方法[J]. 高电压技术，2024，50(12)：5229-5237.
ZHAO Zhenbing, LÜ Xuechun, MIAO Siyu, et al. Inspection method for decoupled knowledge distillation transmission line fittings based on shared feature scoring[J]. High Voltage Engineering, 2024, 50(12): 5229-5237.

[5] 葛召，李洪文，刘海峰，等. 基于 YOLO-GSS 的输电线路边缘端实时缺陷检测方法[J]. 高电压技术，2025，51(2)：669-677.
GE Zhao, LI Hongwen, LIU Haifeng, et al. Real-time defect detection method for edge-end of transmission line based on YOLO-GSS[J]. High Voltage Engineering, 2025, 51(2): 669-677.

[6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015: 91-99.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 21-37.

[8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020: 213-229.

[9] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024: 3429.

[10] 戚银城，武学良，赵振兵，等. 嵌入双注意力机制的 Faster R-CNN 航拍输电线路螺栓缺陷检测[J]. 中国图象图形学报，2021，26(11)：2594-2604.
QI Yincheng, WU Xueliang, ZHAO Zhenbing, et al. Bolt defect detection for aerial transmission lines using faster R-CNN with an embedded dual attention mechanism[J]. Journal of Image and Graphics, 2021, 26(11): 2594-2604.

[11] 梁振锋，梁家钰，马富齐，等. 联合剪枝和知识蒸馏的配电线路设备部件检测方法[J]. 高电压技术，2025，51(11)：5448-5459.
LIANG Zhenfeng, LIANG Jiayu, MA Fuqi, et al. Detection method for power distribution line equipment components combined with pruning and knowledge distillation[J]. High Voltage Engineering, 2025, 51(11): 5448-5459.

[12] 张珂，郑朝烨，石超君，等. 基于多模态对比学习的输电线路螺栓缺陷分类[J]. 高电压技术，2025，51(2)：630-641.
ZHANG Ke, ZHENG Zhaoye, SHI Chaojun, et al. Transmission line bolt defects classification based on multi-modal contrastive learning[J]. High Voltage Engineering, 2025, 51(2): 630-641.

[13] 赵振兵，唐辰康，张靖梁，等. 锈蚀知识引导的配电线路金具及其缺陷双阶段检测方法[J]. 智能系统学报，2026，21(1)：167-178.
ZHAO Zhenbing, TANG Chenkang, ZHANG Jingliang, et al. Rust knowledge-guided dual-stage detection method for distribution line fitting and defect detection[J]. CAAI Transactions on Intelligent Systems, 2026, 21(1): 167-178.

[14] ZHANG J Y, HUANG J X, JIN S, et al. Vision-language models for vision tasks: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8): 5625-5644.

[15] ZAREIAN A, DELA ROSA K, HU D H, et al. Open-vocabulary object detection using captions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021: 14388-14397.

[16] GU X Y, LIN T Y, KUO W C, et al. Open-vocabulary object detection via vision and language knowledge distillation[C]//Proceedings of the 10th International Conference on Learning Representations. OpenReview. net, 2022.

[17] LI L H, ZHANG P C, ZHANG H T, et al. Grounded language-image pre-training[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 10955-10965.

[18] LIU S L, ZENG Z Y, REN T H, et al. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection[C]//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer, 2024: 38-55.

[19] CHENG T H, SONG L, GE Y X, et al. YOLO-world: real-time open-vocabulary object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2024: 16901-16911.

[20] WANG A, LIU L H, CHEN H, et al. YOLOE: real-time seeing anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, USA: IEEE, 2025: 24591-24602.

[21] FU S H, SU Y K, RAO F Y, et al. WeDetect: fast open-vocabulary object detection as retrieval[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, USA: IEEE, 2026: 20377-20387.

[22] FU S H, YANG Q Z, MO Q J, et al. LLMDet: learning strong open-vocabulary object detectors under the supervision of large lan-

guage models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2025: 14987-14997.

[23] CONNEAU A, KHANDELWAL K, GOYAL N, et al. Unsupervised cross-lingual representation learning at scale[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020: 8440-8451.

[24] WOO S, DEBNATH S, HU R H, et al. ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 16133-16142.

[25] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv: 2209.02976, 2022.

[26] JOCHER G, CHAURASIA A, QIU J. YOLO by ultralytics[EB/OL]. GitHub, 2023[2026-02-29]. https://github. com/ultralytics/ultralytics.

[27] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[C]//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer, 2024: 1-21.

[28] LIU S L, LI F, ZHANG H, et al. DAB-DETR: dynamic anchor boxes are better queries for DETR[C]//Proceedings of the 10th International Conference on Learning Representations. OpenReview. net, 2022.

[29] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[C]//Proceedings of the 9th International Conference on Learning Representations. OpenReview. net, 2021.



赵振兵
1979—，男，博士，教授，博导
研究方向为电力人工智能、电力视觉技术、知识表达与推理、机器学习、数字孪生等
E-mail: zhaozhenbing@ncepu.edu.cn

ZHAO Zhenbing
Ph.D., Professor



田本喜
2001—，男，硕士生
主要研究为电力视觉等
E-mail: 220242215120@ncepu.edu.cn

TIAN Benxi



李浩鹏(通信作者)
1991—，男，博士，讲师，硕导
研究方向为电力人工智能、图像信息处理等
E-mail: 51552550@ncepu.edu.cn

LI Haopeng
Ph.D.
Corresponding author

收稿日期 2026-04-28 修回日期 2026-06-19 编辑 卫李静