科学技术与教育 2026年第6期
ISSN: 3079-4455

面向配电站房运维人员安全风险识别的检索增强多模态大模型方法研究

周静*
国网湖北省电力有限公司武汉供电公司，湖北武汉，中国
*通讯作者

【摘要】针对配电站房运维中误操作、越界接近、PPE 佩戴不规范及环境—设备耦合异常难以统一识别的问题，提出一种检索增强多模态大模型方法。该方法融合现场图像/视频、作业文本和环境传感数据，引入安全规程知识库，通过跨模态注意力与规则一致性约束实现风险识别、分级输出和可解释预警。基于 5 类典型作业风险模拟数据集进行验证，并与 SVM、CNN-BiLSTM、ViT-BERT 及多模态 Transformer 对比。结果表明，本文方法在准确率、F1 值、误报率和漏报率等指标上均表现更优，可提升复杂场景下多源信息融合与安全风险判别能力。

【关键词】配电站房；安全风险识别；检索增强；多模态大模型；跨模态融合；规则一致性约束

1.引言

配电站房是配电网末端供电保障与设备运维的重要单元，其内部通常同时存在带电设备密集、通道狭窄、作业空间受限、照明条件不均、环境变量波动以及多类工器具交叉使用等特点，一旦出现误操作、越界接近、PPE 佩戴不规范或环境设备异常叠加，极易演化为人身伤害与停电事故。近年来，围绕配电站房智能化，研究者已在智能配电房总体架构、升级改造、AR/MR 辅助巡检、站房巡检优化、知识图谱运检以及室内无人机自主巡检等方面取得了一定进展[1-6]。这些研究显著提升了站房”可观、可测、可视、可管”能力，但现有方法多聚焦设备状态监测、巡检路径优化或单任务识别，对”作业票文本—人员行为视频—环境传感数据—安全规程知识”之间的跨模态关联建模仍显不足，难以有效应对开放场景下的语义歧义、时序演化和复合风险识别问题。与此同时，大模型在电力行业知识服务、辅助决策、故障诊断和预测分析中的应用快速发展，相关研究指出，提示词工程、检索增强生成和领域微调正成为电力场景落地的关键路径[7]。因此，面向配电站房运维现场，构建融合视觉、文本、传感器与安全知识的检索增强多模态大模型，对实现人员安全风险的实时识别、分级预警与可解释输出具有重要研究意义。

2.大模型算法原理

2.1 总体思路

考虑到 Transformer 自注意力机制适合建模长程依赖关系，BERT 适合提取作业票和操作文本语义，CLIP 类图文对齐思想适合处理开放词汇视觉理解，而电力领域研究又表明 RAG 与领域微调是大模型落地的重要技术路径[7-10]，本文提出一种检索增强多模态大模型（Retrieval-Enhanced Multimodal Large Model，RE-MLLM）用于配电站房运维人员安全风险识别。其核心思想是：先对视频帧、作业文本和环境传感量分别编码，再从安全规程知识库中检索与当前场景最相关的规则片段，最后通过跨模态注意力实现联合推理并输出风险类别与风险分值。

2.2 问题定义

设单个样本表示为

$$x_i = \{V_i, T_i, S_i\} \quad (1)$$

其中，$V_i = \{I_{i1}, I_{i2}, ..., I_{iT}\}$ 为作业视频帧序列，T_i 为作业票、操作描述或语音转写文本，$S_i \in \mathbb{R}^m$ 为环境与设备状态传感向量，包含温湿度、局放强度、门禁状态、电流、电压、烟雾浓度、人员定位偏差等特征。对应标签为

$$y_i \in \{1, 2, ..., C\} \quad (2)$$

本文设置 $C = 5$，分别表示：正常作业、PPE 缺失、越界入侵、误操作、环境—设备耦合风险。

模型目标是学习映射函数

$$f_\theta: \{V_i, T_i, S_i\} \rightarrow \hat{y}_i \quad (3)$$

并输出风险概率向量

$$p_i = [p_{i1}, p_{i2}, ..., p_{iC}] \quad (4)$$

2.3 多模态编码

2.3.1 视觉编码

对视频序列中每一帧 I_{it}输入视觉编码器

$E_v(\cdot)$，得到帧级特征：

$$h_{it}^{v} = E_v(I_{it}),\ t = 1, 2, ..., T \quad (5)$$

进一步采用时序注意力得到视频整体表征：

$$\beta_{it} = \frac{\exp\left(w_v^{T} h_{it}^{v}\right)}{\sum_{t=1}^{T} \exp\left(w_v^{T} h_{it}^{v}\right)} \quad (6)$$

$$\tilde{h}_i^{v} = \sum_{t=1}^{T} \beta_{it}\, h_{it}^{v} \quad (7)$$

这里，β_{it} 用于衡量第 t帧对安全风险判断的重要程度。

2.3.2 文本编码

将作业文本T_i输入文本编码器$E_t(\cdot)$，得到语义向量：

$$h_i^{t} = E_t(T_i) \quad (8)$$

为适应配电站房场景，可将”停电许可”“验电接地”“工器具确认”“禁止跨越安全边界”等领域语义通过领域微调嵌入文本编码器。

2.3.3 传感数据编码

对传感向量 S_i进行标准化处理：

$$\tilde{S}_i = \frac{S_i - \mu}{\sigma} \quad (9)$$

其中 μ和 σ分别为训练集均值和标准差。随后通过多层感知机映射到统一隐空间：

$$h_i^{s} = \phi\left(W_2 \phi\left(W_1 \tilde{S}_i + b_1\right) + b_2\right) \quad (10)$$

其中 $\phi(\cdot)$为非线性激活函数。

2.4 安全知识检索增强

为降低大模型幻觉并增强场景可解释性，构建配电站房安全知识库 $K = \{d_1, d_2, ..., d_N\}$，其中每个 d_j表示一条规程、标准条款、典型违章案例或专家经验片段。

首先构造查询向量：

$$q_i = W_q[h_i^{v}; h_i^{t}; h_i^{s}] \quad (11)$$

其中 $[\cdot;\cdot]$表示向量拼接。然后通过余弦相似度计算与知识库条目的相关度：

$$a_{ij} = \frac{q_i^{T} d_j}{\|q_i\| \|d_j\|} \quad (12)$$

选取 Top-k 个知识片段后，得到检索增强表示：

$$\alpha_{ij} = \frac{\exp(a_{ij})}{\sum_{j \in \text{Top-k}} \exp(a_{ij})} \quad (13)$$

$$h_i^{k} = \sum_{j \in Top\text{-}k} \alpha_{ij}\, d_j \quad (14)$$

该步骤的意义在于：当视觉上出现”未戴绝缘手套”、文本中出现”检修未结束”以及传感器显示”门禁异常开启”时，模型不仅依赖统计学习，还能显式调用相应安全规程进行辅助判断。

2.5 跨模态融合与分类

将视觉、文本、传感和知识检索特征拼接成初始联合表示：

$$H_i^{0} = \left[\tilde{h}_i^{v}; h_i^{t}; h_i^{s}; h_i^{k}\right] \quad (15)$$

随后采用多头注意力机制进行融合：

$$\text{Attention}(Q, K, V0 = \text{softmax}\left(\frac{QK^{T}}{\sqrt{d}}\right)V \quad (16)$$

经过 L层跨模态融合后，得到最终隐向量 z_i：

$$z_i = \text{FusionTransformer}\left(H_i^{0}\right) \quad (17)$$

风险类别概率为：

$$p_i = \text{softmax}(W_o z_i + b_o) \quad (18)$$

进一步定义综合风险分值：

$$R_i = \sum_{c=1}^{C} \omega_c\, p_{ic} + \lambda r_i \quad (19)$$

其中，ω_c 表示各类风险的严重度权重，r_i 为规则冲突得分，λ 为调节系数。通过 R_i可将识别结果转化为低风险、中风险和高风险三级预警。

3 实验测试

3.1 模拟数据集构建

由于配电站房人员安全风险的真实标注数据获取成本高、隐私约束强、危险场景难以大规模复现，本文采用模拟数据进行算法验证。构建思路如下：

基于配电站房典型作业流程，设定”入场核验—设备巡视—停送电操作—工器具使用—退出复核”5 个阶段；

为每个样本同步生成视频行为片段、作业文本和传感器变量；

通过规则注入方式随机制造 PPE 缺失、误入危险边界、工序颠倒、异常温升、烟雾波动等风险事件；

采用高斯噪声、遮挡扰动和传感漂移模拟真实现场的不确定性。

每个样本包含 8 帧视频片段、1 段作业文本、12 维传感向量。共生成 12000 个样本，划分为训练集、验证集和测试集。表 1 给出样本分布。

表 1. 模拟数据集类别分布

类别	总样本	训练集	验证集	测试集
正常作业	2400	1680	360	360
PPE 缺失	2000	1400	300	300
越界入侵	2400	1680	360	360
误操作	2800	1960	420	420
环境-设备耦合风险	2400	1680	360	360
合计	12000	8400	1800	1800

3.2 对比算法与评价指标

选取以下方法作为对比基线：

SVM：以人工提取特征为输入的传统分类器；

CNN-BiLSTM：卷积提取局部视觉特征，BiLSTM 建模时序关系；

ViT-BERT：分别提取图像与文本特征后进行晚融合；

多模态 Transformer：不引入知识检索与规则约束的标准多模态模型；

本文方法：检索增强多模态大模型。

评价指标包括准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F1 值、误报率和漏报率。

3.3 对比结果分析

表 2. 各算法在测试集上的对比结果

模型	准确率	精确率	召回率	F1	误报率	漏报率
SVM	84.11%	84.23%	84.01%	84.00%	13.89%	6.46%
CNN-BiLSTM	88.67%	88.65%	88.57%	88.56%	11.11%	4.51%
ViT-BERT	90.67%	90.61%	90.60%	90.57%	8.89%	3.75%
模态 Transformer	92.67%	92.59%	92.60%	92.57%	6.94%	2.92%
本文方法	95.28%	95.20%	95.24%	95.21%	3.89%	1.88%

由表 2 可知，本文方法在各项指标上均优于对比算法。与性能次优的多模态 Transformer 相比，准确率提升了 2.61 个百分点，F1 值提升了 2.64 个百分点，误报率下降了 3.05 个百分点。这说明仅依赖多模态特征融合仍不足以支撑复杂安全场景识别，而引入安全知识检索与规则一致性约束后，模型对复合型风险的判别能力明显增强。尤其在”视频行为与文本描述不一致”或”环境变量异常但视觉证据较弱”的场景下，本文方法的优势更为明显。

3.4 混淆矩阵分析

为进一步分析各类风险识别效果，给出本文方法在测试集上的混淆矩阵，如表 3 所示。

表 3. 本文方法的混淆矩阵

真实类别 \ 预测类别	正常作业	PPE 缺失	越界入侵	误操作	环境-设备耦合风险
正常作业	346	7	1	4	2
PPE 缺失	9	282	2	4	3
越界入侵	7	1	346	4	2
误操作	2	11	5	399	3
环境-设备耦合风险	9	3	2	4	342

从表 3 可以看出，模型对”误操作”和”越界入侵”识别效果较好，主要原因在于这两类行为在视觉轨迹和文本逻辑上具有较强的模式特征。相对而言，”PPE 缺失”与”正常作业”之间仍存在少量混淆，这说明当遮挡、低照度或摄像头视角不佳时，个体防护细节仍可能被部分削弱；”环境-设备耦合风险”与”正常作业”的少量混淆，则反映出仅凭弱异常传感波动难以完全拉开风险边界，这也是未来需要结合更强时序建模与异常先验的重点方向。

4 结论

本文面向配电站房运维人员安全风险识别需求，提出了一种检索增强多模态大模型方法。该方法融合现场视觉、作业文本、环境传感数据及安全规程知识，通过跨模态注意力、规则一致性约束和时序平滑机制，实现了对典型安全风险的识别与分级预警。基于模拟数据的对比实验表明，所提方法在准确率、F1 值、误报率和漏报率等指标上均优于 SVM、CNN-BiLSTM、ViT-BERT 及标准多模态 Transformer 等方法，体现出较强的多源信息融合与复杂场景判别能力。研究结果表明，该方法可为配电站房智能化安全运维提供有效支撑。

参考文献

[1] 张雪峰，陈红州，熊小伏，马政. 一种智能配电房设计方法[J]. 智能电网，2020，10（4）：137-147. DOI：10.12677/SG.2020.104015.

[2] 郑培昊，王满意，李建伟，沙博. 智慧配电房升级改造建设研究及应用[J]. 电力信息与通信技术，2019，17(12)：73-77.

[3] 何明，陈莹莹，张斌，李思尧. AR 技术在配电站房巡检业务中的应用研究[J]. 现代信息科技，2019，3（1）：175-176.

[4] 裴超，王大磊，杨占刚，黄宇翔，张杰恺. 考虑时空分布的配电网站房巡检策略[J]. 电气技术，2023，24（1）：86-90.

[5] 曹捷，阙小生，李慎兴，范永学，李兴，宋文志. 基于知识图谱技术的配电站房智能运检[J]. 吉林大学学报（信息科学版），2023，41（3）：474-483.

[6] 曹锦云，卢其芳. 配电站房室内无人机自

主巡检关键技术研究及其应用[J]. 农村电气化， 2024， （10）： 29-32.
[7] 严新荣， 高翔， 林达， 等. 大模型在电力行业的应用与挑战[J]. 发电技术，2025，46（4）： 637-650.
[8] Vaswani A， Shazeer N， Parmar N， et al. Attention Is All You Need[EB/OL]. arXiv：1706.03762， 2017.
[9] Devlin J， Chang M W， Lee K， Toutanova K. BERT： Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of NAACL-HLT. 2019： 4171-4186.
[10] 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. PMLR，2021，139：8748-8763.