

中文引用格式:聂本武,陈述,陈云,等. 水电工程施工安全隐患图文多模态智能识别方法[J]. 中国安全科学学报, 2026, 36(3): 104-112.

英文引用格式:NIE Benwu, CHEN Shu, CHEN Yun, et al. An image-text multimodal intelligent identification method for construction safety hazards in hydropower engineering[J]. China Safety Science Journal, 2026, 36(3): 104-112.

水电工程施工安全隐患 图文多模态智能识别方法^{*}

聂本武^{1,2,3}高级工程师, 陈述^{**1,3}教授, 陈云^{1,3}副教授, 田雪琪²,
曹坤煜¹, 李智⁴正高级工程师

(1 三峡大学 水电工程施工与管理湖北省重点实验室, 湖北 宜昌 443002; 2 国家能源投资集团有限责任公司 金沙江分公司, 四川 成都 610041; 3 三峡大学 水利与环境学院, 湖北 宜昌 443002;
4 中国长江三峡集团有限公司, 湖北 武汉 430000)

中图分类号: X948

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2026.03.0881

基金项目: 国家自然科学基金资助(52479127, 52209163); 湖北省自然科学基金青年 A 类项目资助(2025AFA074);
湖北省自然科学基金创新发展联合基金资助(2024AFD153)。

【摘要】 为解决水电工程施工安全隐患识别存在单模态特征表达不完整、图文特征融合效率低等问题, 提出水电工程施工安全隐患图文多模态智能识别方法。首先, 针对水电工程施工特点, 定义 12 种施工安全隐患分类特征, 建立水电工程施工安全隐患图文多模态数据集; 其次, 利用双向变换器模型(BERT)和视觉变换器(ViT)模型分别提取隐患文本与图像特征, 引入门控融合网络(GFN), 动态调节图文特征贡献度, 捕捉多模态关联特征信息, 通过多层感知器提高对多模态分类识别精度; 最后, 通过对比试验, 检验模型识别的准确性与可靠性。结果表明: 该方法通过增强识别稳定性, 实现对多模态隐患特征的优化贡献, 多模态隐患识别准确率高达 84.99%, 较文本模型提升 1.73%, 较图像模型提升 12.24%, 隐患识别分类优于已有基准模型, 有助于提升安全隐患智能识别的鲁棒性。

【关键词】 水电工程; 施工安全隐患; 多模态; 门控融合网络(GFN); 智能识别

An image-text multimodal intelligent identification method for construction safety hazards in hydropower engineering

NIE Benwu^{1,2,3}, CHEN Shu^{1,3}, CHEN Yun^{1,3}, TIAN Xueqi², CAO Kunyu¹, LI Zhi⁴

(1 Hubei Key Laboratory of Construction and Management in Hydropower Engineering, China Three Gorges University, Yichang Hubei 443002, China; 2 Jinshajiang Branch, China Energy Investment Corporation, Chengdu Sichuan 610041, China; 3 College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang Hubei 443002, China; 4 China Three Gorges Corporation, Wuhan Hubei 430000, China)

Abstract: To address the problems of incomplete unimodal feature representation and low image-text fusion efficiency in construction safety hazard identification for hydropower projects, an intelligent image-

* 文章编号: 1003-3033(2026)03-0104-09; 收稿日期: 2025-09-14; 修稿日期: 2025-12-11

** 通信作者: 陈述(1986—), 男, 湖北英山人, 博士, 教授, 主要从事水电工程施工安全管理方面的研究。E-mail: chenshu@ctgu.edu.cn。

text multimodal intelligent identification method was proposed. First, 12 categories of construction safety hazards were defined according to hydropower construction characteristics, and an image-text multimodal dataset was established. Second, bidirectional encoder representations from transformers (BERT) model and vision transformer (ViT) model were employed to extract hazard text and image features respectively. GFN was then introduced to dynamically adjust the contribution of image and text features and capture cross-modal correlated feature information, while a multi-layer perceptron was used to improve classification accuracy. Comparative experiments were conducted to verify the model's accuracy and reliability. The results show the method optimizes the contribution of multimodal features by enhancing identification stability. The multimodal hazard identification accuracy reaches 84.99%, representing an improvement of 1.73% over the text-based model and 12.24% over the image-based model. The proposed approach outperforms existing benchmark models in hazard classification and improves the robustness of intelligent hazard identification.

Keywords: hydropower project; construction safety hazard; multimodal; gated fusion network (GFN); intelligent identification

0 引言

在当前水电工程智能建造背景下,施工现场安全隐患数据呈现多源异构特征^[1]。安全检查过程中产生的表格、文本、图像等非结构化数据^[2],具有体量大、形式多样、关联复杂的特性,增加了智能识别的技术难度。传统智能识别方法存在一定局限,单一图像识别技术虽能检测视觉显性隐患,但难以解析文本检查中隐含的语义信息;单纯依赖文本挖掘算法虽可提取结构化报告中的隐患关键词,却无法充分理解施工现场的隐患视觉关联^[3]。这种模态割裂的分析方法,忽略文本与图像模态的互补价值,加之对专家经验的主观依赖,导致隐患识别的结果存在较高误判风险。因此,如何充分利用图文信息建立多模态语义关联,实现多模态融合的安全隐患精准识别,成为当前水电工程智能建造领域亟待解决的关键问题之一。

目前,智能识别技术已广泛应用在施工安全领域,高效挖掘施工安全隐患知识成为水电领域研究热点之一。学者们通过融合计算机视觉^[4]、深度学习^[5]、多模态网络^[6]等技术,构建了多层次的安全隐患识别与分析框架。陈述等^[7]通过双向变换器(Bidirectional Encoder Representations from Transformers, BERT)预训练模型,建立基于命名实体识别的水电工程施工安全规范实体识别模型,提取规范文本中实体类别。周佳一等^[8]通过融合文本上下文语义与领域专家知识,基于注意力机制的深度学习模型,实现了隐患类别的高精度自动标注。FANG Weili等^[6]通过构建图像特征与安全规则的

映射关系,提出基于卷积神经网络(Convolutional Neural Network, CNN)的不安全行为识别方法。LIU Jiajing等^[9]采用视觉变换器(Vision Transformer, ViT)模型,实现将图像中的具体目标(如工人、脚手架)与文本描述精准对齐。田丹等^[10]结合水电工程图像局部特征与文本语义标签的联合训练,提出基于跨模态对比学习的隐患自动辨识方法,提升了复杂光照、遮挡条件下隐患识别的鲁棒性。陈岩松等^[11]针对多模态情感分析中的模态融合与特征提取问题,利用门控融合网络(Gated Fusion Network, GFN)的方法,有效提升了情感分类任务的性能。WANG Yiheng等^[12]开发了图像-文本语义相似性模型,通过对比学习机制对齐图像特征与文本描述的语义空间,实现了施工场景图像与安全隐患文本的跨模态检索,为施工安全隐患识别高效利用提供了新路径。

综上所述,现有研究多聚焦于深度学习单模态方法或部分融合多模态模型,存在跨模态特征融合不充分、模型泛化能力薄弱及领域适配性不足等问题。单模态方法局限于文本语义或图像视觉特征的独立处理,忽略了隐患描述文本与图像模态的互补价值。当前研究虽已关注多模态融合的潜在优势,但环境信息与隐患描述数据中蕴含的隐患线索尚未得到充分挖掘与利用,对跨模态表征学习机制及其在水电工程专业场景中的优化应用较少。鉴于此,笔者拟提出一种基于GFN的图文多模态智能识别方法,并通过选取对比模型,检验所提方法识别的准确性与可靠性,以期水电工程施工安全隐患智能化、精细化管理提供技术支撑。

1 基于 GFN 图文多模态智能识别方法

1.1 施工安全隐患分类特征

安全隐患的多模态表征,是在特定时空环境下,由文本、图像等异构数据源共同描述的工程风险实体^[13]。基于水电工程安全管理的实践特征,安全隐患数据主要呈现 2 种典型模式:①基于人工巡检的文本数据,即现场工程师依据专业知识体系,结合法律法规、标准规范等制度约束,通过视觉判断形成的文本记录;②基于视觉感知的图像数据,监察人员采用“随手拍”工具或人工智能(Artificial Intelligence,

AI)摄像头,基于预设的安全规则实时捕获、抓拍并存储于智能安全 APP 系统中,形成图像数据流及隐患地图。

根据《企业职工伤亡事故分类》(GB6441—1986),伤亡事故分为爆炸、车辆伤害、触电、高处坠落、火灾、机械伤害、物体打击等,将隐患图文数据按照可能导致事故的类型进行分类,这样有利于对隐患识别进行分类。此外,按照《水利水电工程施工安全管理导则》(SL721—2015)等行业标准,增加“文明施工”隐患识别的分类,可减少工序干扰、提升文明施工安全效率等施工环境的重要性。为便于研究,将隐患类型通过规范化隐患特征描述能够实现快速分类,见表 1。

表 1 水电工程施工安全隐患分类特征

Table 1 Classification characteristics of construction safety hazards in hydropower engineering	
隐患类型	隐患特征
高处坠落	存在从坠落高度基准面 2 m 以上(含 2 m)处坠落的隐患
触电事故	电流有通过人体造成伤害或死亡的隐患
物体打击	失控物体存在(如工具、材料)击中人体造成伤害的隐患
起重伤害	存在因吊具、吊物等引发的伤害的隐患
机械伤害	存在机械设备运动部件(如齿轮、皮带)导致的夹击、切割等伤害的隐患
火灾事故	易燃物堆积、电气短路或明火管理不当等发生燃烧的隐患
车辆伤害	存在机动车辆(如自卸车、挖掘机)碰撞、碾压导致伤害的隐患
爆炸事故	存在压力容器、危险化学品等因瞬间能量释放造成破坏的隐患
坍塌事故	存在土方、模板或建筑结构塌落的隐患
设备损坏	生产设备非正常损毁和机械故障、超负荷运行或维护不足引发连锁的隐患
文明施工	因现场管理混乱(如杂物堆放、通道堵塞)等脏乱差导致的间接隐患
其他事故	未涵盖在上述分类中的其他风险或隐患

按照文本、图像模态进行施工安全隐患数据收集、整理与预处理,采用结构化表格进行标准化录入,描述每张图片,确保语义一致性与可追溯性。收集某大型水电工程大量历史隐患资

料,通过筛选和清洗,每一条施工安全隐患中都包含隐患的描述、隐患的分类以及隐患的现场图片,得到施工安全隐患图文多模态数据集,见表 2。

表 2 施工安全隐患图文多模态数据集

Table 2 Image-text multimodal dataset of construction safety hazards				
序号	图片描述	隐患描述	隐患类型	隐患照片
1	挖掘机的机械臂停放在空中	挖掘机机械臂在停止作业时未放置地面	起重伤害	
2	一个打开的电源箱,未作接地连接	二级配电箱一闸多机且电缆出线未作接地连接	触电事故	
3	一个没有连接扣件的手架	边坡支护排架剪刀撑钢管与立杆相交处未使用扣件连接	坍塌事故	
⋮	⋮	⋮	⋮	⋮

续表 2				
序号	图片描述	隐患描述	隐患类型	隐患照片
2282	高处作业人员未系挂安全带	排导槽上游高处作业人员未系挂安全带	高处坠落	

1.2 施工安全隐患多模态智能识别框架

施工安全隐患图文多模态智能识别方法实现流程如图 1 所示,具体步骤为:

- 1) 数据输入。收集隐患文本、图像数据,剔除多标签和无效图像,对数据集进行设置标签、划分等预处理,建立文本与图像的一一对应关系。
- 2) 特征提取。利用 BERT 模型处理隐患文本,提取文本特征;使用 ViT 模型处理图像数据,提取图像特征。

- 3) 特征融合。融合提取的文本特征和图像特征通过 GFN,得到融合特征表达。
- 4) 特征处理。融合后的特征通过 Transformer 模型进行进一步处理,提取更高层次的特征。
- 5) 隐患分类。使用多层感知器 (Multilayer Perceptron,MLP) 对处理后的特征进行分类,最终输出各类隐患的分类结果,如起重伤害、物体打击、高处坠落等。

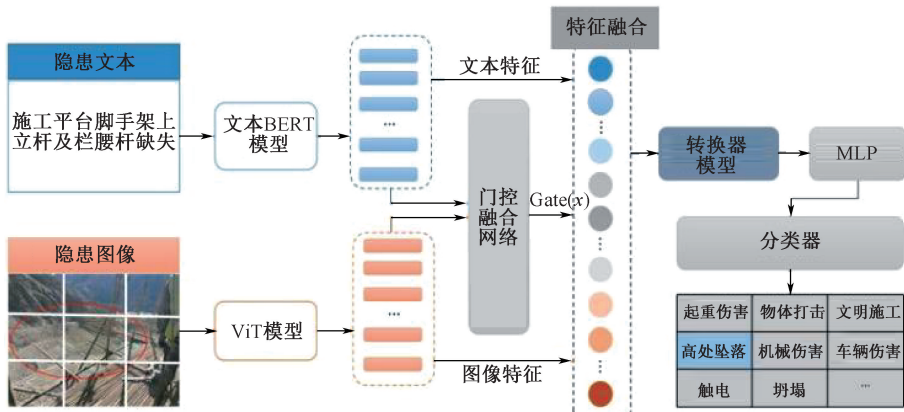


图 1 施工安全隐患图文多模态智能识别框架

Fig. 1 Framework of image-text multimodal identification for construction safety hazards

1.3 基于 GFN 的施工安全隐患多模态特征提取

1.3.1 文本特征提取

使用 BERT 模型进行微调后,作为文本编码器。与传统的单向语言模型不同,BERT 模型的编码机制能够同时从左右 2 个方向获取上下文信息,从而捕捉更加全面的语义依赖关系。BERT 模型的整体框架主要包括预训练和微调 2 个阶段。在预训练阶段,BERT 模型使用大规模的未标注语料库,借助自监督学习方式训练,主要任务包括掩码语言模型和下一句预测。在微调阶段,BERT 模型会将预训练得到的参数作为下游任务的初始化权重,并结合标注好的任务数据对整个模型参数进行端到端的优化调整,使其适配具体的任务目标。将处理后的文本输入到 BERT 模型,获得包含 768 维的向量。具体实现过程如下式:

$$t_i = [w_1, w_2, \cdots, w_n]$$
$$f_i = \text{BERT}(t_i, \theta_i) \in R^d, i \in [1, N]$$

式中: w_n 为隐患文本描述语句中的词; t_i 为隐患文本的初始序列; θ_i 为 BERT 模型的超参数; R^d 为向量空间; f_i 为隐患文本经过 BERT 模型编码后输出的向量;BERT($\cdot$) 为 BERT 模型的编码运算过程,将文本序列映射为高维向量的函数。

1.3.2 图像特征提取

为将图像识别问题转化为类比自然语言序列建模问题,使用 ViT 作为模型的图像编码器。ViT 模型结构不依赖于传统的卷积操作,而是完全基于 Transformer 中的多头自注意力机制 (Multi-head Self-Attention, MHSA) 对图像块之间的关系进行建模。与卷积运算不同,Transformer 网络通过自注意力机制处理全局图像信息,这对于在复杂施工背景下,识

别多类型隐患分布图像有明显优势。通过将图像切割成 N 个小块的 patch 作为模型的输入,在将输入传递到编码器层之前,还需要添加一个用于分类的 token,在每一个序列的首位插入一个用于识别分类的 token 向量。对于每个输入的 token,输入交叉特征融合后产生 3 个向量 query ($\mathbf{Q}$)、key ($\mathbf{K}$) 和 value ($\mathbf{V}$),通过把 n 组子向量单独计算一次注意力得到一个注意力头 h 。因此,自注意力和输入序列计算公式如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \quad (3)$$

$$\text{Encode}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{concat}(h_1, h_2, \dots, h_n) \quad (4)$$

式中: $\text{softmax}(\cdot)$ 函数的作用为归一化; $1/\sqrt{d_k}$ 为对特征的缩放,避免因维度过大导致权重接近 1 或 0。

1.3.3 多模态融合

现有研究在获得最终向量输出之后,通常以直接拼接的形式进行最终预测,这可能会无意中添加冗余信息。GFN 是利用一个可学习的门控单元,根据动态计算特征权重并加权融合,实现自适应特征整合。该机制能有效捕获多源数据的复杂依赖关系,提升多模态或复杂特征组合场景下的表示学习能力。参数主要包括输入维度、门控单元结构、Sigmoid 函数计算权重和输出特征归一化。相较于

传统融合方法(拼接、平均、注意力机制),GFN 具备动态适应性、噪声鲁棒性、计算效率高和可解释性强的优势,门控权重大小直接反映各模态贡献度,可与其他基准模型配对以增强效果^[14],如图 2 所示。文本隐患的模态和图像隐患的模态表示如下:

$$\mathbf{f}_t = \text{BERT}(t) \quad (5)$$

$$\mathbf{f}_i = \text{ViT}(t) \quad (6)$$

将基于文本隐患的模态表示为 $\mathbf{f}_t$ 和基于图像隐患的模型表示为 $\mathbf{f}_i$ 作为输入,分别输入门控循环分层网络中 $\text{Gate}(x)$ 得到文本和图像的输出门 O ,计算公式如下:

$$O_t = \mathbf{f}_t \cdot \text{Gate}(\mathbf{f}_t) \quad (7)$$

$$O_i = \mathbf{f}_i \cdot \text{Gate}(\mathbf{f}_i) \quad (8)$$

利用门控循环网络 $\text{Gate}(x)$ 以及自注意力机制,使得不同模态之间的信息得到充分交互,并通过门控机制有效消除冗余和不相关的信息,融合特征随后输入至 MLP 进行非线性映射与判别,最终输出类别概率分布:

$$\text{Gate}(x) = \mathbf{x} \cdot \text{sigmoid}(\text{MLP}(\mathbf{x})) \quad (9)$$

$$P_t = \text{softmax}(\text{MLP}(O_t + O_i)) \quad (10)$$

式中: $\mathbf{x}$ 为门控循环网络中待处理的多模态特征向量; $\text{sigmoid}(\cdot)$ 为激活函数,将输入映射到 $[0, 1]$ 区间,输出作为门控权重; $\text{MLP}(\cdot)$ 对输入 $\mathbf{x}$ 进行非线性特征变换。

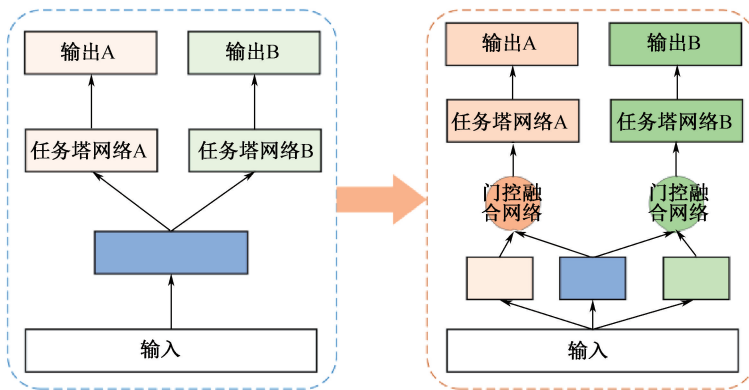


图2 多模态 GFN

Fig.2 Multimodal GFN

1.3.4 损失函数

采用交叉熵损失函数监督学习类别概率预测模型,优化目标函数 J 表示为:

$$J = - \sum_{t=1}^K y_t \ln(P_t) \quad (11)$$

式中: k 为总类别数; y_t 为第 t 个样本的真实标签(one-hot 编码),若样本属于类别 t ,则 $y_t = 1$,否则为

0; P_t 为对应的预测概率值。该损失函数表示最小化预测概率与真实标签之间的差异,从而提高模型在多分类任务中的判别性能。

采用 4 个核心指标全面评估模型性能:准确率 A 、精确率 P 、召回率 R 和 F_1 值,真阳性(True Positive, TP)、真阴性(True Negative, TN)、假阳性(False Positive, FP)和假阴性(False Negative, FN)分

别代表正确预测为正例、正确预测为负例、负例误判为正例、正例误判为负例的数量。各指标的计算公式如下:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

(12)

$$P = \frac{TP}{TP + FP}$$

(13)

$$R = \frac{TP}{TP + FN}$$

(14)

$$F_1 = 2 \times \frac{P \times R}{P + R}$$

(15)

表 3 施工安全隐患多模态试验数据样本数量

Table 3 Sample size of multimodal experimental data for construction safety hazards											
隐患类型	训练集	验证集	测试集	总数	占比/%	隐患类型	训练集	验证集	测试集	总数	占比/%
爆炸事故	57	16	8	81	4	起重伤害	85	24	12	121	5
车辆伤害	15	4	2	22	1	设备损坏	20	6	3	29	1
触电事故	505	144	72	721	32	坍塌事故	41	12	6	59	3
高处坠落	192	55	27	274	12	文明施工	246	70	35	352	15
火灾事故	205	59	29	293	13	物体打击	44	13	6	63	3
机械伤害	41	12	6	58	3	正常照片	700	200	100	1 000	—
其他事故	146	42	21	209	9	合计	2 297	656	328	3 282	100

2.2 选取施工安全隐患识别对比模型

采用预训练的 BERT 模型作为文本编码器,ViT 模型作为图像编码器。模型训练优化策略与其他主流模型保持一致。通过 Loss 曲线进行调优分析,刻画门控融合图文多模态、

2 施工安全隐患智能识别实例验证

2.1 图文多模态数据采集与预处理

水电工程历史隐患清单,是隐患识别的基础数据来源。收集某大型水电工程从 2020—2024 年历史隐患资料共 2 282 条,将试验的训练集、验证集和测试集划分比例为 7 : 2 : 1。为确保用于训练模型泛化能力,还收集整理 1 000 张非工程领域的正常图片。隐患类型的数量在一定程度上决定了训练模型的识别精度^[15],隐患的类型分布是不均匀的,见表 3,属于不平衡数据集,这是符合水电工程施工现场的特征的。

表 4 试验选取对比模型

Table 4 Comparative models selected in the experiment				
模型名称	输入类型	优点	缺点	适用领域
BERT ^[7]	文本	双向上下文预训练	计算资源消耗大、生成能力弱	文本分类、命名实体识别、阅读理解
ViT ^[9]	图像	序列化+全局自注意力机制	计算复杂度高	大规模图像分类/检测
BERT-ViT ^[12]	图像-文本	跨模态交互(交叉注意力机制/特征融合)	训练复杂、数据稀缺、模态不平衡	图文交互任务

2.3 施工安全隐患智能识别结果分析

2.3.1 与基准模型识别效果对比分析

4 类模型(文本、图像、图文、门控融合图文)在施工安全隐患识别效果上差异显著,图文结合模式可有效提升施工安全隐患识别模型性能^[16]。多模态模型 Loss 曲线对比如图 3 所示,门控融合图文模型在对照组中训练集损失值下降趋势较其他模型均有明显提升。

多模态对照组测试集结果对比见表 5。从中可

图文多模态、文本单模态及图像单模态在训练集损失值的下降趋势和收敛情况。为与安全隐患多模态识别模型进行对比,验证该模型的有效性,设置以下对比模型,其主要性能与特点见表 4。

以看出,门控融合图文模型准确率较纯文本模型提升 1.73%,较纯图像模型提升 12.24%;在召回率指标上较纯图像模型提升 23.26%;F₁ 值较纯图像模型提升 12.84%,而较图文模型提升基本持平,仅提升 0.78%。图文模型在总体上要优于单纯的文本或图像模型,而门控融合图文模型在识别准确率表现最佳,其次是图文模型。而单纯的文本模型的准确率(84.49%)略高于图像模型(76.58%),这表明:在该数据集上,文本信息的处理相对图像信息处

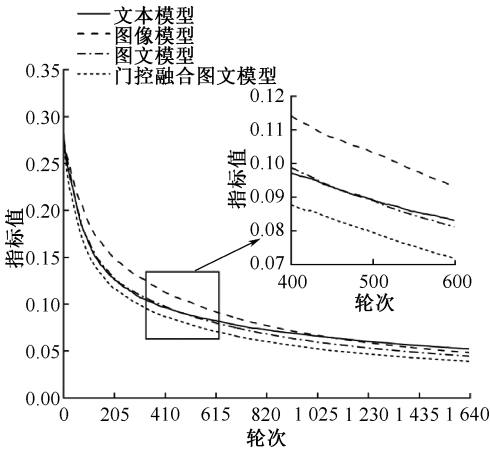


图3 多模态模型损失对比

Fig.3 Loss comparison of multimodal models

理更为准确。这是因为复杂的水电工程施工现场存在大量的人员、材料和机械设备,在现场采集到的图像数据往往在水平和垂直平面上包含大量复杂的背景信息,使得模型在提取隐患特征时容易受到干扰。但是图文模型的精确率最高,达到 87.92%,进一步说明该模型在预测正类时较为准确,误识别更少。因此,多模态模型在图像隐患识别相关任务中的提升幅度(普遍超过 20%)显著高于文本任务模型(1%~4%),从而验证了图文特征互补的有效性,其中,GFN 对图像隐患特征的优化贡献尤为突出,在召回率指标上实现单模态隐患图像模型 23.26% 的最大幅度提升。

表5 多模态对照组测试集结果对比

Table 5 Result comparison of test set for multimodal control groups				%
模型	准确率	精确率	召回率	F_1 值
文本模型	84.49	84.43	69.96	84.21
图像模型	76.58	76.82	57.61	75.32
图文模型	85.30	87.92	70.44	84.45
门控融合图文模型	85.95	85.98	71.01	84.99

2.3.2 图文多模态特征融合效果分析

GFN 在识别水电工程施工安全隐患类型时的性能评估结果见表 6。模型在火灾事故和爆炸事故识别上具有高准确性和高召回率;机械伤害、触电事故及起重伤害的识别也表现出色。这主要得益于此类隐患具有典型的视觉特征(如明火、烟雾、电气)和明确的文本描述范式(如接线、易燃物等术语),使得多模态特征能够实现精准对齐。而坍塌事故、物体打击、其他事故识别效果较低,主要受限于视觉特征模糊(如细微裂缝难以捕捉)、文本记录中常使用结构异常等非标准化表述,导致模型难以建立有效

的特征关联机制。这种性能差异本质上反映了水电工程施工安全隐患识别隐性风险难捕捉的技术挑战。

表6 图文多模态融合对隐患类型识别效果

Table 6 Effect of image-text multimodal fusion on hazard type recognition

隐患类型	精确率	召回率	F_1 值
其他事故	0.600 0	0.250 0	0.352 9
坍塌事故	0.307 7	0.307 7	0.307 7
文明施工	0.685 5	0.885 4	0.772 7
机械伤害	0.875 0	0.875 0	0.875 0
火灾事故	0.913 8	0.929 8	0.921 7
爆炸事故	0.875 0	0.875 0	0.875 0
物体打击	1.000 0	0.428 6	0.600 0
触电事故	0.905 1	0.925 4	0.915 1
设备损坏	1.000 0	0.571 4	0.727 3
起重伤害	0.892 9	0.925 9	0.909 1
车辆伤害	0.666 7	0.400 0	0.500 0
高处坠落	0.857 1	0.857 1	0.857 1

2.3.3 安全隐患识别分类效果对比分析

不同事故类别下,对照模型在测试集上的 P 、 R 和 F_1 值性能指标,如图 4 所示。GFN 在识别稳定性和识别效果上更具优势,在触电事故和火灾事故中,基于 GFN 的多模态模型识别效果一直保持较高水平,即便数据量占比出现波动,也未受明显影响,稳定性优于其他模型。机械伤害隐患识别上,门控机制多模态模型虽未达最高值,但仍显著高于图像模型,体现其在复杂事故数据处理上的优势;而在文明施工、设备损坏这类识别难度较大的隐患类型中,门控融合图文模型识别效果的下降幅度相对较小,恢复能力也更强。

利用 F_1 值指标综合评判不同模型在隐患类别识别性能见表 7。起重伤害、触电事故、火灾事故等隐患类别在各模型识别效果上较高, F_1 值在 0.9 以上,模型识别效果稳定。图文模型对多数隐患的识别均有提升,如机械伤害的识别效果优于单模态,但设备损坏与单模态持平,说明此类隐患的多模态特征增益不明显。门控融合图文模型优化了物体打击、高处坠落、爆炸等隐患的表现,说明门控融合图文模型能有效提升其特征权重分配策略,特别是物体打击和爆炸事故这 2 类隐患,较其他对照模型提升性能较大。因此,门控融合图文模型凭借其 GFN 的特性,在隐患类型识别中,整体表现出色,稳定性强,识别效果更具优势,相比传统单一模态、多模态识别模型,能更好地应对复杂隐患类型的数据特征,为施工安全隐患识别提供更可靠的方案。

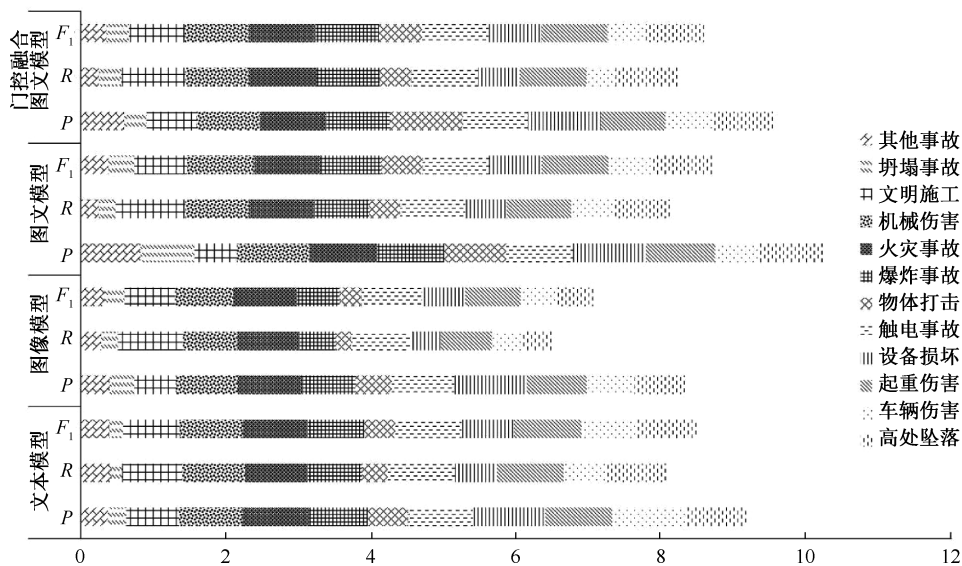


图 4 对照模型在隐患类型识别效果

Fig. 4 Recognition effect of hazard types for comparative models

表 7 对照模型在隐患类型识别的 F_1 指标差异性

Table 7 Differences in F_1 -score of comparative models

隐患类型	文本模型	图像模型	图文模型	门控融合图文模型
起重伤害	0.925 9	0.763 3	0.924 0	0.909 3
触电事故	0.908 6	0.834 7	0.909 2	0.915 2
火灾事故	0.891 7	0.875 2	0.920 4	0.921 8
机械伤害	0.875 0	0.802 4	0.936 1	0.875 0
爆炸事故	0.774 7	0.606 6	0.833 6	0.875 0
文明施工	0.776 3	0.722 8	0.757 2	0.781 2
设备损坏	0.766 2	0.676 2	0.766 2	0.766 2
高处坠落	0.845 4	0.563 7	0.850 2	0.857 1
车辆伤害	0.783 3	0.522 2	0.600 0	0.522 2
物体打击	0.449 2	0.338 1	0.619 0	0.676 2
其他事故	0.395 5	0.335 2	0.483 7	0.401 0
坍塌事故	0.198 1	0.278 9	0.444 6	0.307 7

3 结 论

1) 针对水电工程施工特点,定义 12 种施工安

全隐患分类特征,考虑文本与图像之间的关联,构建图文多模态智能识别模型;采用 BERT 模型和 ViT 模型分别提取隐患文本、图像特征,引入 GFN 进行特征融合,通过 MLP 实现对安全隐患图文多模态智能识别与分类。

2) 通过单模态、多模态的基准模型识别对照试验,发现文本在隐患描述中所承载的核心信息量要优于视觉特征;揭示了施工安全隐患图文特征互补的有效性,GFN 对图像隐患特征的优化贡献尤为突出,隐患识别准确率较纯文本模型提升 1.73%,较纯图像模型提升 12.24%。

3) GFN 可有效提升多模态模型在关键隐患类别上的识别性能,增强系统稳定性,对提升水电工程安全隐患识别模型的准确性和鲁棒性具有重要价值。下一步,将聚焦提升隐患智能分类在不平衡数据上表现不一致的问题,进一步优化门控融合机制,以全面提高施工安全隐患多模态分类性能。

参 考 文 献

[1] 卢冰,陈述,曹坤煜,等. 水电工程施工安全隐患类别辅助校正方法[J]. 水力发电学报, 2025, 44(4): 42-49.
LU Bing, CHEN Shu, CAO Kunyu, et al. Auxiliary correction methods for categories of potential safety hazards in hydropower project construction[J]. Journal of Hydroelectric Engineering, 2025, 44(4): 42-49.

[2] 陈述,王典学,杨应柳,等. 水电工程施工安全隐患语义匹配模型[J]. 中国安全科学学报, 2024, 34(12): 40-47.
CHEN Shu, WANG Dianxue, YANG Yingliu, et al. Semantic matching model of potential safety hazards in hydroelectric project construction[J]. China Safety Science Journal, 2024, 34(12): 40-47.

[3] 杨阳蕊,潘世峰,刘雪梅,等. 多模态知识图谱与大模型协同的水利工程风险应对决策推荐[J]. 水利学报, 2025,

56(4): 519–530.

YANG Yangrui, PAN Shifeng, LIU Xuemei, et al. Multimodal knowledge graph collaborated with large model for decision recommendation of water projects risk response[J]. Journal of Hydraulic Engineering, 2025, 56(4): 519–530.

- [4] 张泽辉, 张乾隆, 徐晓滨, 等. 基于视觉的工人高处攀爬不安全行为识别模型[J]. 中国安全科学学报, 2025, 35(2): 144–151.

ZHANG Zehui, ZHANG Qianlong, XU Xiaobin, et al. Unsafe behavior recognition model of high climbing workers based on vision[J]. China Safety Science Journal, 2025, 35(2): 144–151.

- [5] ZHONG Botao, SHEN Luoxin, PAN Xing, et al. Visual attention framework for identifying semantic information from construction monitoring video[J]. Safety Science, 2023, 163: DOI:10.1016/J.SSCI.2023.106122.

- [6] FANG Weili, LOVE P E D, DING Lieyun, et al. Computer vision and deep learning to manage safety in construction: matching images of unsafe behavior and semantic rules[J]. IEEE Transactions on Engineering Management, 2023, 70: 4 120–4 132.

- [7] 陈述, 张超, 陈云, 等. 基于命名实体识别的水电工程施工安全规范实体识别模型[J]. 中国安全科学学报, 2024, 34(9): 19–26.

CHEN Shu, ZHANG Chao, CHEN Yun, et al. Model of identifying entities of safety specification for hydropower engineering construction[J]. China Safety Science Journal, 2024, 34(9): 19–26.

- [8] 周佳一, 郑霞忠, 田丹, 等. 水电工程施工安全隐患多标签文本智能分类方法[J]. 水力发电学报, 2024, 43(11): 114–124.

ZHOU Jiayi, ZHENG Xiazhong, TIAN Dan, et al. Multi-label text intelligent classification method for construction safety hazards of hydropower projects[J]. Journal of Hydroelectric Engineering, 2024, 43(11): 114–124.

- [9] LIU Jiajing, FANG Weili, PETER E D L, et al. Detection and location of unsafe behaviour in digital images: a visual grounding approach[J]. Advanced Engineering Informatics, 2022, 53: DOI:10.1016/j.aei.2022.101688.

- [10] 田丹, 许仁乐, 邵波, 等. 强化特征表达的水电工程施工安全隐患自动辨识方法[J/OL]. 河海大学学报: 自然科学版: 1–11. [2025–05–28]. <https://kns.cnki.net/kcms2/article.html>.

TIAN Dan, XU Renle, SHAO Bo, et al. Automatic identification method for construction safety hazards in hydropower engineering with enhanced feature expression[J/OL]. Journal of Hohai University: Natural Sciences Edition: 1–11. [2025–05–28]. <https://kns.cnki.net/kcms2/article.html>.

- [11] 陈岩松, 张乐, 张雷瀚, 等. 基于跨模态注意力和门控单元融合网络的多模态情感分析方法[J]. 数据分析与知识发现, 2024, 8(7): 67–76.

CHEN Yansong, ZHANG Le, ZHANG Leihan, et al. Multimodal sentiment analysis method based on cross-modal attention and gated unit fusion network[J]. Data Analysis and Knowledge Discovery, 2024, 8(7): 67–76.

- [12] WANG Yiheng, XIAO Bo, BOUFERGUENE A, et al. Proactive safety hazard identification using visual-text semantic similarity for construction safety management[J]. Automation in Construction, 2024, 166: DOI:10.1016/j.autcon.2024.105602.

- [13] KIM H, YI J S. Image generation of hazardous situations in construction sites using text-to-image generative model for training deep neural networks[J]. Automation in Construction, 2024, 166: DOI:10.1016/j.autcon.2024.105615.

- [14] AREVALO J, SOLORIO T, MONTES M, et al. Gated multimodal networks[J]. Neural Computing and Applications, 2020, 32(14): 10 209–10 228.

- [15] CAO Kunyu, CHEN Shu, CHEN Yun, et al. Decision analysis of safety risks pre-control measures for falling accidents in mega hydropower engineering driven by accident case texts[J]. Reliability Engineering & System Safety, 2025, 261: DOI:10.1016/j.res.2025.111120.

- [16] TIAN Dan, LIU Hao, CHEN Shu, et al. Human error analysis for hydraulic engineering: comprehensive system to reveal accident evolution process with text knowledge[J]. Journal of Construction Engineering and Management, 2022, 148(9): DOI:10.1061/(ASCE)CO.1943-7862.0002366.



作者简介: 聂本武 (1987—), 男, 湖北孝感人, 博士研究生, 高级工程师, 主要研究方向为水电工程智能建造。E-mail: nbwpdp@sina.com。