

ICS 国际标准分类号
CCS 中国标准文献分类号



团 标 准

T/CES XXX-XXXX

电网设备巡检多模态大模型总体技术导则

General technical guidelines for multimodal large-scale models in power grid
equipment inspection

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国电工技术学会 发布

目 次

目 次.....	III
前 言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 符号、代号和缩略语	2
5 章节导航	2
6 数据采集	3
6.1 预训练阶段数据采集	3
6.2 微调阶段数据采集	4
7 数据清洗	5
7.1 设备巡检视觉类数据清洗	5
7.2 设备巡检规程文本类数据清洗	5
7.3 设备巡检声纹类数据清洗	5
7.4 设备巡检振动类数据清洗	6
7.5 设备巡检电气数据清洗	6
8 数据标注	7
8.1 预训练阶段数据标注	7
8.2 微调阶段数据标注	8
8.3 数据增强	9
9 模型预训练与微调	10
9.1 设备巡检多模态大模型架构	10
9.2 设备巡检模型预训练	10
9.3 设备巡检大模型微调	12
9.4 设备巡检不同模态数据融合	12
10 模型评测	13
10.1 评测维度	13
10.2 评测指标	13
10.3 评测数据	14
10.4 评测方法	14
11 模型部署	15
11.1 推理平台部署	15
11.2 模型推理加速	15
11.3 镜像化部署与服务化推理	16
附 录 A（规范性附录）输电山火识别开发示例	17
A.1 输电山火任务设置	17
A.3 输电山火数据集清洗	17
A.4 输电山火数据集标注	17
A.5 输电山火多模态大模型预训练和微调	17
A.6 输电山火多模态大模型评测	17
附 录 B（资料性附录）评测指标计算公式	18
B.1 视觉问答类评分指标计算公式	18

T/CES XXX—XXXX

B.2 图文检索类评分指标计算公式	18
B.3 图文生成类评分指标计算公式	19
B.4 其他指标计算公式	19
附录 C (资料性附录) 设备巡检标注样例	20
C.1 设备巡检视觉类标注示例	20

前　　言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规则起草。

请注意本文件的某些内容可能涉及专利，本文件的发布机构不承担识别专利的责任。

本文件由中国电工技术学会提出。

本文件起草单位：国网四川省电力公司、国网信息通信产业集团有限公司、中国电力科学研究院有限公司、广东电网有限责任公司、福建亿榕信息技术有限公司、北京国网信通埃森哲信息技术有限公司、成都搏智科技有限公司。

本文件主要起草人：张凌浩、向思屿、邝俊威、杜佩珂、毛洋、李强、赵峰、庄莉、王秋琳、李炳森、伍臣周、邱镇、刘茂凯、谈元鹏、张国梁、莫文昊、赵林林、邵彦宁。

本文件为首次发布。

电网设备巡检多模态大模型总体技术导则

1 范围

本标准规定了多模态大模型在电网设备巡检领域中的数据采集、数据清洗、数据标注、模型预训练与微调、模型评测和模型部署。

本标准适用于指导基于多模态大模型开展的电网设备智能巡检及辅助研判工作。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的，凡是注日期的引用文件，仅注日期的版本适用于本文件，凡是不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

GB/T 42755—2023 人工智能 面向机器学习的数据标注规程

TC260—003 生成式人工智能服务安全基本要求

T/CES 277—2024 电力人工智能样本增广技术架构要求

T/CES 289—2024 电力人工智能模型场景化评价指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

多模态大模型 multimodal large-scale model

一类具有大规模参数和复杂计算结构，并且能够处理和理解多种不同数据模态（如文本、图像、声音、视频等）的机器学习模型。

3.2

预训练 pre-training

一种大规模数据集上进行初步训练的模型训练策略，模型在大规模数据集上获得足够的语言理解能力，能够捕捉到语言的深层次结构和语义信息，提高模型的鲁棒性和泛化能力。

3.3

电力词表 electric power vocabulary

一种专门用于将电力领域的文本数据转换为模型可理解的序列符号工具。

3.4

长文本外推 long text extrapolation

一种解决在训练阶段模型只在有限长度的文本上学习，在推理阶段能够处理比训练时更长的文本问题，并且保持好的性能的长文本处理方法。

3.5

长文本内插 long text interpolation

一种在处理长文本时，通过某种方法将长文本分割成多个短片段，并在这些片段之间插入额外的信息或处理步骤，以提高模型对长文本的理解和处理能力的长文本处理方法。

3.6

提示词工程 prompt engineering

在输入序列中添加可学习的嵌入向量作为提示，以引导预训练模型生成特定任务的输出。

3.7

模型微调 fine tuning

一种模型训练策略，利用已知的网络结构和已知的网络参数，通过调整模型的所有或部分参数以提高模型训练效率，一般分为全量微调、增量微调。

3.8

批处理 batch

在一次迭代 (iteration) 中用于训练模型的一组样本。

3.9

拉普拉斯算子 laplace operator

一种定义在欧氏空间中的二阶微分算子，常用于图像处理中的边缘检测。

3.10

哈希算法 hash algorithm

将输入图像映射成一串短小的数字序列，该数字序列通常称为输入图像的哈希，用于图片检索，重复图片剔除，以图搜图以及图片相似度比较。

3.11

编辑距离 edit distance

一个度量两个字符序列之间差异的字符串度量标准，两个单词之间的编辑距离是将一个单词转换为另一个单词所需的单字符编辑（插入、删除或替换）的最小数量。

3.12

动态时间规整 dynamic time warping

将两个代表同一个类型的事物的不同长度序列进行时间上的“对齐”。

3.13

欧氏距离 euclidean distance

高维线性空间内向量所表示的点到点之间的绝对距离。

3.14

思维链 chain of thought

一系列相互关联的思维环节，每个环节都基于前一个环节，并为下一个环节提供了输入和刺激。

4 符号、代号和缩略语

下列符号、代号和缩略语适用于本文件。

BLEU：双语互译质量评估 (Bilingual Evaluation Understudy)

Few-shot：小样本学习 (Few-shot Learning)

FID：真实图像和生成图像的特征向量之间距离的一种度量 (Frechet Inception Distance)

LLM：大语言模型 (Large Language Model)

Lora：低秩适应 (Low-Rank Adaptation)

MFCC：梅尔倒谱系数 (Mel-scaleFrequency Cepstral Coefficients)

MLLM：多模态大模型 (Multimodal Large Language Model)

NLP：自然语言处理 (Natural Language Processing)

RAG：检索增强生成 (Retrieval-Augmented Generation)

ROUGE：基于召回率的摘要评估 (Recall-Oriented Understudy for Gisting Evaluation)

Spectrogram：谱图

Zero-shot：零样本学习 (Zero-shot Learning)

5 章节导航

本导则共分 11 章，各章节主要内容如下：

第 1 章阐述了本导则的主要内容和适用范围。

第 2 章阐述了本导则引用的规范性文件。

第 3 章阐述了本导则相关的术语及其定义。

第 4 章阐述了本导则相关的符号和代号。

第 6 章阐述了数据采集的要求规范，数据模态包括视觉类、规程文本类、声纹类、振动类以及电气量数据，用于后续清洗和标注。

第 7 章阐述了各模态数据清洗的要求规范，以确保数据质量满足后续分析需求。

第 8 章阐述了各模态数据标注与数据增强的要求，以提升模型的解析、识别和泛化能力。

第9章阐述了设备巡检多模态大模型的架构设计、预训练与微调策略，各模态编码器预训练方法及融合方式，以提升模型对电力设备巡检任务的综合理解与处理能力。

第10章阐述了设备巡检多模态大模型的评测方法，包括评测维度、指标、数据和方法，以确保模型在电力设备巡检任务中的性能和适用性。

第11章阐述了设备巡检多模态大模型的部署方法，涵盖推理平台的硬件适配、操作系统适配、安全性和隐私保护，以及模型推理加速、镜像化部署和服务化推理等关键步骤，以确保模型在实际应用中的高效性和安全性。

6 数据采集

6.1 预训练阶段数据采集

6.1.1 公开数据集采集

公开数据集采集数据量应在百万以上，数据来源可包括网络公开数据集、图片搜索引擎、主流社交媒体网站、电子化报纸期刊等途径获取的数据集和其他渠道获取的数据。

6.1.2 设备巡检预训练任务数据采集

6.1.2.1 设备巡检视觉类数据采集

设备巡检视觉类数据应涵盖但不限于以下模态：

- a) 可见光数据采集：应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集变电站相关数据时，采集对象宜包括站内安全标识、围栏、地面通道、变电设备本体（如开关柜、断路器、互感器、避雷器等）和站内外环境（包括站外警示牌、围墙、周边地形地貌等）等；
- b) 红外数据采集：应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集变电站内变压器的红外数据时，采集对象宜包括变压器接头和连接点、变压器表面、套管、变压器周围环境等；
- c) 激光点云数据采集：应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集变电站相关的点云数据时，采集对象宜包括变压器、断路器、隔离开关等设备的外观和内部结构、绝缘子、穿墙套管、电力电缆等设备。

设备巡检视觉类数据采集时应遵循以下要求：

- a) 图像目标尺寸应多样化，覆盖从近、中、远三种拍摄距离和不同拍摄视角采集的数据；姿态角度应多样化，涉及设备部件姿态、表计指针朝向、人员动作姿态等。空间位置关系需涵盖目标移位、缺失、脱落、距离过近、共面等情况；
- b) 图像背景应多样化，覆盖电网设备巡视典型背景，如输电通道场景应同时包含城区、郊区、山地、河流等场景；
- c) 对于干扰形式，应包括主要背景干扰物目标以及镜头模糊、脏污等情况。光照变化则应包含暗光、正常光照、过曝、逆光、大雾等多种拍摄环境；
- d) 红外图像应涵盖不同温度差异下的目标表现，如低温、常温、高温等情况，以及不同发射率物体的热成像特征；
- e) 点云数据应考虑不同扫描频率、分辨率下的点云数据质量和精度等；
- f) 数据量：宜在十万以上。

6.1.2.2 设备巡检规程文本类数据采集

设备巡检规程文本类数据采集应遵循以下要求：

- a) 应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集变电站变压设备相关文本数据时，采集对象宜包括设备的巡检报告、维护记录、故障日志、规章制度文本等多种数据类型；
- b) 语言形式应多样化，包括各类专业术语、行业标准表述以及不同人员可能使用的口语化描述；
- c) 信息内容应体现多维关系，包括设备运行状态、操作人员身份与职责、时间顺序及事件关联等；
- d) 异常信息应涵盖输入错误、数据缺漏、不一致信息、过时信息以及人为干扰等；
- e) 数据量：宜在十万以上。

6.1.2.3 设备巡检声纹类数据采集

设备巡检声纹类数据采集应遵循以下要求：

- a) 应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集变电站变压器设备声纹数据时，采集对象宜应包括设备运行音、故障异音、环境噪声、人员语音指令等多种声纹数据类型；
- b) 应涉及多种背景噪声环境，涵盖从安静环境、一般运行噪声到高噪声干扰环境，并且声音事件应考虑到多样的声源位置和传播路径，包括直接传导、反射以及设备之间的连锁声音影响；
- c) 异常声音应明确区分报警声、摩擦声、撞击声、震动声等不同类型的异常声响，确保数据涵盖各类可能的故障预兆和状态异常信号；
- d) 环境声音采集应覆盖多种典型电力设施现场环境，如变电站、配电房以及室外线路现场，适应不同天气状况和声学条件下的声音采集需求；
- e) 数据量：宜在十万以上。

6.1.2.4 设备巡检振动类数据采集

设备巡检振动类数据采集应遵循以下要求：

- a) 应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集输电通道振动数据时，采集对象宜包括绝缘子、开关柜、断路器、杆塔、导线、发电机、电动机等；
- b) 应同时记录设备的运行参数（如转速、负载率、温度、电流、电压等）和环境参数（如温度、湿度、粉尘、噪声等），以便与振动信号进行关联分析；
- c) 应涵盖振动加速度、振动速度、振动位移、频谱数据和时域数据，覆盖设备的高频、中频和低频振动特征；
- d) 应涵盖设备在不同工况下的振动特征，如正常运行、启动、停机及突发故障时的振动变化；
- e) 数据量：宜在十万以上。

6.1.2.5 设备巡检电气量数据采集

设备巡检电气量数据采集应遵循以下要求：

- a) 应根据场景需求采集所有相关的环境要素与目标要素数据。例如，采集变电设备电气量数据时，采集对象宜包括变压器、断路器、隔离开关、互感器、避雷器等；
- b) 数据类型应包含电压信号、电流信号、功率信号、频率信号、温度信号、局部放电信号等多种电气量数据类型；
- c) 数据应覆盖不同工况下的电气量特征，包括正常运行、启动、停机及突发故障时的信号变化；
- d) 异常信号应明确过电压、过电流、局部放电、温度异常等不同类型的故障特征，确保数据涵盖各类可能的故障预兆和状态异常信号；
- e) 数据量：宜在十万以上。

6.2 微调阶段数据采集

6.2.1 公开数据集采集

公开数据集采集应满足以下要求：

- a) 应充分考虑应用场景的特殊性，从公开数据源中筛选与目标场景高度相似的数据样本。例如输电通道山火检测，应收集城区火灾、篝火、工业排烟、过曝灯光等相似数据；
- b) 数据采集要求参照 6.1.2 节；
- c) 数据量宜在 1000 以上。

6.2.2 设备巡检微调任务数据采集

- a) 应采集与场景任务环境和目标具有一致性或相似性的数据。例如，在主变套管温度监测场景中，应采集与主变套管温度记录相关的电力设备巡检规程文本数据，包含主变设备温度监控标准、温度异常判断标准、设备故障预警措施等内容；
- b) 各模态数据采集要求应分别参照 6.1.2.1、6.1.2.2、6.1.2.3、6.1.2.4 和 6.1.2.5 节；
- c) 数据量：各类型数据量宜在 1000 以上。

7 数据清洗

7.1 设备巡检视觉类数据清洗

7.1.1 清理过曝和过暗数据

- a) 应通过图像处理工具（如 OpenCV）检测电力设备巡检图像的亮度值，过滤掉因光照条件不佳导致的过曝或过暗数据，确保设备状态清晰可见；
- b) 应对部分可用数据进行亮度调整，尝试恢复设备关键部位（如仪表盘、接头、绝缘子等）的细节，便于后续分析和故障诊断。

7.1.2 清理模糊数据

- a) 应使用清晰度检测算法（如拉普拉斯算子）评估电力设备巡检图像的清晰度，过滤掉因拍摄抖动或对焦不准导致的模糊数据，确保设备关键部位（如铭牌、连接点等）清晰可辨；
- b) 应对部分模糊数据进行锐化处理，尝试提升设备关键部位的清晰度，便于后续分析和故障识别。

7.1.3 相似数据去重

- a) 应使用哈希算法或特征比对技术识别重复的电力设备巡检图像，避免因多次拍摄相同设备导致的数据冗余；
- b) 应根据业务需求选择保留最新或最完整的巡检图像，确保数据集中每个设备的巡检记录唯一且完整。

7.2 设备巡检规程文本类数据清洗

7.2.1 去重与去噪

- a) 重复文本：应通过文本相似度算法（如编辑距离和词频-逆文档频率）查找并移除电力设备巡检规程中重复或高度相似的内容，确保规程的唯一性和准确性；
- b) 噪声文本：应移除巡检记录中的乱码、异常编码、过多非语义字符（如连续标点、表情符等），确保文本内容可读且符合电力行业规范。

7.2.2 无用或敏感内容过滤

- a) 停用词清理：应根据电力设备巡检的业务场景，选择是否保留停用词（如“的”、“是”等），确保文本简洁且重点突出；
- b) 敏感词检测：应通过替换或者标注的方法过滤敏感词汇（如设备编号、地理位置等），确保数据安全性；
- c) 空白/极短文本：应剔除不包含有效信息的文本（如长度极短的文本），确保数据质量。

7.2.3 文本一致性

- a) 统一大小写、简繁体：应根据电力行业规范，对英文大小写、简繁体汉字进行统一转换，确保巡检记录的一致性；
- b) 拼写检查：应根据业务需求，对巡检记录中的拼写错误进行校正，或使用自然语言模型检测并修正明显的拼写错误，确保文本准确性；
- c) 标点、数字处理：应统一特殊标点和数字格式（如电压值、电流值等），确保巡检记录的一致性和可解析性。

7.3 设备巡检声纹类数据清洗

7.3.1 音频质量检查

- a) 采样率：应检查电力设备巡检音频文件的采样率、位深是否符合后续模型需求，确保音频质量满足设备状态分析的要求；
- b) 音量过低/过高：应利用音频能量阈值或 RMS 电平检测，筛除因环境噪音或设备故障导致的过度嘈杂或音量过低的音频，确保音频清晰可辨；

c) 噪声背景：针对电力设备巡检场景（如变电站、输电线路等），应根据业务需求过滤或特殊保留含噪声的样本，确保音频数据适用于设备状态分析。

7.3.2 截断或音频损坏

- a) 完整性检测：应对电力设备巡检音频文件的长度、播放时长进行核对，若音频被截断或无法正常播放，则需要剔除或重录，确保音频完整性；
- b) 破损处理：应对包含静音段或高频失真的巡检音频，评估修复成本与必要性后决定保留或删除，确保音频数据质量。

7.3.3 去重和相似度判断

- a) 指纹/特征抽取：应提取电力设备巡检音频的指纹（如 MFCC、spectrogram 等），对比相似度以发现重复录音或相同的设备运行片段，确保数据集的多样性；
- b) 应根据业务需求选择保留最新或最完整的巡检音频，确保数据集中每个设备的声纹记录唯一且完整。

7.4 设备巡检振动类数据清洗

7.4.1 数据质量检查

- a) 采样频率与精度：应检查电力设备巡检振动数据的采样频率、位深是否符合后续模型需求，确保振动数据能够准确反映设备运行状态；
- b) 异常值检测：应利用统计方法（如标准差、四分位数等）识别并剔除因传感器故障或外部干扰导致的异常振动数据，确保数据的可靠性；
- c) 噪声干扰：针对电力设备巡检场景（如变电站、输电线路等），应根据业务需求过滤或特殊保留含噪声的振动数据，确保数据适用于设备状态分析。

7.4.2 数据完整性与准确性

- a) 完整性检测：应对电力设备巡检振动数据的时间序列完整性进行核对，若数据缺失或时间戳不连续，则需要进行插值或重测，确保数据完整性；
- b) 数据校准：应定期对振动传感器进行校准，确保采集到的振动数据准确无误，能够真实反映设备的振动情况。

7.4.3 去重和相似度判断

- a) 特征提取与比对：应提取电力设备巡检振动数据的特征（如频谱特征、时域特征等），对比相似度以发现重复记录或相同的设备运行片段，确保数据集的多样性；
- b) 应根据业务需求选择保留最新或最完整的巡检振动数据，确保数据集中每个设备的振动记录唯一且完整。

7.5 设备巡检电气数据清洗

7.5.1 异常数据过滤

- a) 幅值异常检测：应通过设定电气量信号的幅值阈值（如电压、电流、功率等），过滤掉因传感器故障或外部干扰导致的异常幅值数据（如电压信号的突增或突降、电流信号的异常波动等），确保电气量数据的合理性和有效性；
- b) 频率异常检测：应利用频谱分析技术，识别并剔除电气量信号中不符合设备正常运行频率范围的异常频率成分（如谐波干扰、高频噪声等），确保数据能够准确反映设备的运行特征。

7.5.2 去重与相似度判断

- a) 特征比对：应提取电气量信号的特征（如时域特征、频域特征、统计特征等），通过相似度算法（如欧氏距离、动态时间规整等）识别重复或高度相似的电气量数据，避免数据冗余；
- b) 业务规则筛选：应根据业务需求选择保留最新或最完整的电气量数据，确保数据集中每个设备的电气量记录唯一且完整。

8 数据标注

8.1 预训练阶段数据标注

8.1.1 设备巡检视觉类标注

8.1.1.1 设备巡检视觉类标注任务设计

标注任务设计应符合以下要求:

- a) 应在标注过程中覆盖多种电力场景与对象，包括变电站环境（如站内安全标识、围栏、地面通道等）、变电设备本体（如开关柜、断路器、互感器、避雷器等）、站内外环境（如站外警示牌、围墙、周边地形地貌等）以及输电设备本体（如杆塔、导线、金具等），以确保模型能够识别和理解在不同环境和作业条件下的各类信息；
- b) 具体标注任务应涵盖图文描述、图文问答、目标定位、OCR（用于读取标志牌、相序牌等）、指代目标描述、指代目标问答，以及缺陷隐患、违规动作等级判定与处置决策建议等方面的内容，应满足多样化需求，增强模型对不同类型的电力指令的理解能力。

8.1.1.2 设备巡检视觉类标注规范设计

标注规范设计应符合以下要求:

- a) 在问题提示方面，应引入相关知识，并提供相关知识库的出处。在回答方面，需要标注场景特征，明确背景特征和光照特点；标注目标种类，区分关注目标与干扰目标；
- b) 标注目标状态时应明确目标的位置、大小、动作姿态以及与其他参考物的空间状态；
- c) 应标注判别依据，结合知识库明确输入指令与判别结论之间的思维过程，并标注判别结论，提供缺陷隐患和违规动作判定等级与处置决策建议，符合《GB/T 42755-2023 人工智能 面向机器学习的数据标注规程》的要求。

8.1.2 设备巡检规程文本标注

8.1.2.1 设备巡检规程文本标注任务设计

标注任务设计应符合以下要求:

- a) 标注任务应包括文本摘要生成、关键词抽取、关键句提取、指令意图识别、文本分类、缺陷与隐患识别，以及根据规程的合规性和操作风险进行等级评估与建议等方面的内容，提高模型对电力设备巡检规程文本的理解和处理能力。

8.1.2.2 设备巡检规程文本标注规范设计

标注规范设计应符合以下要求:

- a) 在标注说明方面，应结合电力行业标准和相关知识，并给出知识来源；
- b) 在具体标注中，应标注文本的结构特征，识别章节、段落等逻辑层次；
- c) 应标注关键术语和概念，确保术语与其定义的一致性；
- d) 应标注指令内容，明确指令与操作步骤的关系；
- e) 应标注合规性，包括对潜在风险及不合规现象的识别和判断。此外，还应根据知识库提供的巡检标准和风险管理框架，标注风险评估结论，并提出相应的优化建议和决策。

8.1.3 设备巡检声纹标注

8.1.3.1 设备巡检声纹标注任务设计

标注任务设计应符合以下要求:

- a) 标注任务应包括声纹特征提取、声源识别、背景噪声识别、异常声音检测、声音与设备状态的关联分析，以及故障类型的判断和维护建议等方面的内容，提高模型对设备声纹数据的识别和分析能力。

8.1.3.2 设备巡检声纹标注规范设计

标注规范设计应符合以下要求:

- a) 在标注说明方面，应引入声纹分析的相关知识，并标明数据及知识来源；
- b) 在具体标注中，应标注声纹的基础特征，包括频率、振幅、持续时间等；区分正常运行声纹与异常声纹；
- c) 应标注声源的类型和位置，分析声音来源于哪个部件或区域；
- d) 应准确标注背景噪声的类型和干扰程度，识别异常声音的特征和原因；
- e) 应结合声纹数据与设备运行状态的关联，标注分析依据，提供准确的故障判断结论，并给出相应的维修或维护建议。

8.1.4 设备巡检振动类数据标注

8.1.4.1 设备巡检振动类数据标注任务设计

标注任务设计应符合以下要求：

- a) 标注任务应涵盖振动特征提取、设备部件振动识别、干扰振动源识别、异常振动检测、振动与设备状态的关联分析，以及故障类型的判定和维修建议等方面的内容；
- b) 应区分对不同设备部件（如电机、轴承、齿轮等）的振动模式标注，以及正常运行振动与故障振动的标注，并对振动数据中潜在的故障特征进行详细标注，以便模型能够学习到振动数据与设备健康状态之间的映射关系。

8.1.4.2 设备巡检振动类数据标注规范设计

标注规范设计应符合以下要求：

- a) 在标注说明方面，应引入振动分析的专业知识，并明确数据及知识的出处；
- b) 在具体标注中，应标注振动数据的基础特征，如振动频率、振幅、相位等；区分设备正常运行时的振动特征与故障状态下的振动特征；
- c) 应标注振动源的类型和位置，确定振动是由设备的哪个部件或哪个部位产生；
- d) 应准确标注干扰振动源的特性，如环境振动、其他设备运行引起的振动等，以及它们对目标设备振动数据的影响程度；
- e) 应结合振动数据与设备运行状态的关联，标注分析依据，提供准确的故障诊断结论，并给出针对性的维修或维护建议，如更换部件、调整设备参数等，以指导实际的设备维护工作。

8.1.5 设备巡检电气量数据标注

8.1.5.1 设备巡检电气量数据标注任务设计

标注任务设计应符合以下要求：

- a) 标注任务应包括电气量数据的特征提取（如电流、电压、功率等参数）、设备运行状态识别、异常模式检测、趋势分析与预测，以及故障类型判定与维护建议等内容；
- b) 应涵盖设备在不同运行阶段（启动、稳定运行、停机等）的电气量数据特征标注，明确区分正常工况数据与故障数据，并对数据中潜在的故障征兆（如参数波动、超限值等）进行详细标注，使模型能够捕捉到电气量数据与设备健康状态之间的关联及细微变化。

8.1.5.2 设备巡检电气量数据标注规范设计

标注规范设计应符合以下要求：

- a) 标注说明应以电气量分析的专业知识（如电力系统原理、设备运行标准等）为基础，并明确数据来源及知识依据；
- b) 具体标注应涵盖电气量数据的基础特征，如参数的周期、频率、幅值、相位、谐波分量等；
- c) 应区分设备在不同运行状态（如空载、满载、异常负载）下的电气量数据特征；
- d) 应标注数据中的突变点（如电流骤升、电压跌落）、趋势变化点（如功率缓慢衰减）等关键信息，明确其与设备运行状态或潜在故障的关联；
- e) 应标注干扰信号（如噪声、电磁干扰）的特性及其对电气量数据的影响程度；
- f) 应结合电气量数据与设备运行状态的关联性，标注分析依据并提供故障诊断结论，同时给出针对性维护建议（如参数校准、部件更换、停机检修等），以指导实际运维。

8.2 微调阶段数据标注

8.2.1 设备巡检视觉类标注

在微调阶段，标注要求应进一步细化和深化，增强模型对特定专业场景的理解与推理能力，应遵循以下要求：

- a) 应在标注阶段覆盖场景典型要素，包括目标、背景、场景等所有与任务相关的描述信息，确保模型能够准确识别和理解场景中的各类目标及其相互关系；
- b) 标注任务宜采用图文描述方式，通过图像和文字的相互补充，更全面地描述场景和目标信息，使模型能够获得更丰富的信息，提高其对场景的理解和推理能力；
- c) 描述方式宜采用结构化标注形式，使标注数据更加规范、统一，便于模型的处理和分析。

8.2.2 设备巡检规程文本标注

在微调阶段应遵循以下要求：

- a) 标注应全面涵盖规程文本中的关键要素，包括具体的操作对象（如设备名称、型号、状态等）、操作环境（如天气、光线、场地等）以及操作场景（如日常巡检、故障处理、应急抢修等）；
- b) 标注任务应采用详细的文字描述方式，对规程文本中的各类信息进行准确、清晰的表述；
- c) 参照 8.2.1 节第 c) 条。

8.2.3 设备巡检声纹标注

在微调阶段应遵循以下要求：

- a) 标注应全面涵盖声纹数据中的关键要素，包括设备的名称、型号、状态等特征，巡检环境（如噪声背景、语音清晰度等）以及巡检场景（如日常巡检、故障识别、紧急处理等）；
- b) 标注任务应采用详细的描述方式，对声纹数据中的各类特征进行准确、清晰的表述；
- c) 参照 8.2.1 节第 c) 条。

8.2.4 设备巡检振动类标注

在微调阶段应遵循以下要求：

- a) 标注应全面涵盖振动数据中的关键要素，包括振动频率、振幅、相位等特征，设备的名称、型号、部件信息（如电机、轴承、齿轮等）、运行状态（如正常运行、故障状态等）以及巡检场景（如日常巡检、故障诊断、故障预警等）；
- b) 标注任务应采用详细的描述方式，对振动数据中的各类特征进行准确、清晰的表述，区分正常振动与故障振动，标注各个设备部件的振动模式以及振动与设备状态的关联；
- c) 参照 8.2.1 节第 c) 条。

8.2.5 设备巡检电气量标注

在微调阶段应遵循以下要求：

- a) 标注应全面涵盖电气量数据中的关键要素，包括电流、电压、功率等电气参数，设备的名称、型号、运行状态（如启动、稳定运行、停机等）、以及巡检场景（如日常巡检、故障诊断、设备维护等）；
- b) 标注任务应采用详细的描述方式，对电气量数据中的各类特征进行准确、清晰的表述，区分正常工况与故障状态，标注不同运行阶段的电气量特征及其潜在的故障征兆（如参数波动、超限值等）；
- c) 参照 8.2.1 节第 c) 条。

8.3 数据增强

8.3.1 设备巡检视觉类数据增强

应采用几何变换、颜色变换、噪声添加和混合增强等方式实现数据增强，提高模型的泛化性能。

8.3.2 设备巡检规程文本类数据增强

应采用同义词替换、插入和删除、词序交换和 GPT 等语言模型生成的方式实现数据增强。

8.3.3 设备巡检声纹类数据增强

应采用声速变换、添加噪声、音调调整、混响添加等方式实现数据增强。

8.3.4 设备巡检振动类数据增强

应采用振动信号的时间缩放、频率调制、幅度调整、相位偏移和混合信号合成等方式实现数据增强。

8.3.5 设备巡检电气量数据增强

应按照《T/CES 277-2024 电力人工智能样本增广技术架构要求》的相关规范，采用电气量数据的电压扰动、电流扰动、功率缩放、相位偏移、噪声注入和多源数据融合等方式实现数据增强。

8.3.6 设备巡检多模态数据混合增强

- a) 跨模态融合：应将视觉、文本、声纹、振动和电气量信号数据进行融合，生成新的多模态样本，增强数据的多样性和丰富性；
- b) 模态内增强：在每个模态内部，应采用相应的数据增强方法，如图像的几何变换、文本的同义词替换、声纹的声速变换、振动的频率调制等，提升各模态数据的质量和数量；
- c) 模态间关联增强：通过挖掘不同模态之间的关联信息，如视觉与文本描述的对应关系、声纹与振动信号的同步变化等，生成更具代表性和一致性的多模态数据，提高模型对多模态数据的理解和分析能力。

9 模型预训练与微调

9.1 设备巡检多模态大模型架构

模型架构宜考虑电力词表、注意力机制、长文本处理、图像编码器和激活函数，每个模块应符合以下要求：

- a) 电力词表应包含电力领域的专业词汇、常用词汇及电力符号和单位；
- b) 注意力机制是大模型架构核心组成部分，应具备：
 - 1) 应支持领域适应性，注意力机制的网络设计宜考虑电力领域文本特性、电力领域图像特性，如电力书籍、标准的文本风格，电力运行序列数据的处理，电力设备状态监测量，缺陷或异物的物理特征，作业人员的着装特征等；
 - 2) 可适应电力人工智能大模型训练需要，按需选择或调整注意力网络架构，可采用自注意力网络、多头注意力网络、稀疏注意力网络等；
 - 3) 应支持电力视觉、文本、时序、结构化等多种模态数据的注意力建模。
- c) 长文本处理宜考虑长文本外推和长文本内插，应具备：
 - 1) 长文本外推应采用位置编码方法，以在电力技术报告、会议记录、规章制度等长文本中保持位置信息的准确性；
 - 2) 长文本外推应采用分段处理、滑动窗口等方法处理电力超长文本；
 - 3) 长文本内插应结合电力不同知识类型，设计电力长文本分割策略，如融合电力文档标题的句子、段落等特定长度的文本片段切割；
 - 4) 长文本内插应保持电力内容一致性，可采用语义一致性的内插方法，如针对技术标准描述，通过将标题文字内插入标准条款，保持标准条款的语义完整、语句结构清晰。
- d) 图像编码器应能提取电力领域目标的特征，能够处理电力目标的复杂性；
- e) 大模型激活函数规范性应符合以下要求：
 - 1) 应选择标准化激活函数，可根据电力人工智能大模型的具体需求和网络结构进行选择。例如，对于深度网络，ReLU 及其变体；
 - 2) 可组合不同激活函数。例如，可在不同的网络层中使用不同的激活函数，以充分利用它们的优点并减少局限性；
 - 3) 应通过实验证不同激活函数对电力专业细分场景的应用中模型性能的影响，选择性能最优的激活函数进行部署和应用。

9.2 设备巡检模型预训练

在设备巡检多模态大模型中，编码器预训练应通过各模态数据的掩码重建学习方法，提升编码器对不同类型数据（如视觉、文本、声纹、振动、电气量等）的解析和重建能力。

9.2.1 设备巡检视觉编码器预训练

具体训练策略应包括但不限于：

- a) 对不同类别的设备目标的遮挡重建：强化编码器对各类设备特征（如变压器、断路器、输电线等）的表达和识别能力，确保模型能够准确区分不同设备类型及其状态；
- b) 对不同大小目标的遮挡重建：提升编码器处理不同尺寸目标（如大型变压器与小型仪表盘）的精确性和灵活性，确保模型能够适应电力设备巡检中多样化的目标尺寸；
- c) 对目标的不同姿态和角度的遮挡重建：增强编码器对设备姿态变化（如倾斜、旋转等）的敏感度与准确度，确保模型能够应对设备在不同安装角度下的巡检需求；
- d) 结合目标与临近背景信息的遮挡重建：提升编码器对空间关系和位置逻辑的推理能力，确保模型能够准确识别设备与其周围环境（如支架、导线等）的关系；
- e) 处理干扰物和背景的遮挡重建：提高编码器的抗干扰能力及背景辨识能力，确保模型能够在复杂背景（如植被、建筑物等）中准确识别设备目标；
- f) 不同光照条件下的整体遮挡重建：包括暗光、正常光照、过曝、逆光和大雾环境，以增强编码器对多变光照条件的适应性及其对目标的识别能力，确保模型能够在各种实际巡检场景中稳定工作。

9.2.2 设备巡检声纹编码器预训练

具体训练策略应包括但不限于：

- a) 对不同类别的设备声音的掩蔽重建：强化编码器对各类设备运行声音（如变压器嗡鸣、断路器动作声、放电声等）的表达和识别能力，确保模型能够准确区分不同设备及其运行状态；
- b) 对不同频率音频的掩蔽重建：提升编码器处理不同频段声音（如低频振动、高频放电等）的精确性和灵活性，确保模型能够适应电力设备巡检中多样化的声音特征；
- c) 对声音的不同强度和动态变化的掩蔽重建：增强编码器对声音强度变化（如设备启动、运行、停止等）的敏感度与准确度，确保模型能够捕捉设备运行状态的细微变化；
- d) 结合声音与环境背景信息的掩蔽重建：提升编码器对声音来源和环境场景的推理能力，确保模型能够区分设备运行声音与背景噪音（如风声、鸟鸣等）；
- e) 处理噪声干扰和背景音的掩蔽重建：提高编码器的抗噪能力及背景音辨识能力，确保模型能够在复杂环境（如变电站、输电线路等）中准确识别设备运行声音；
- f) 不同环境声音条件下的整体掩蔽重建：包括安静环境、嘈杂环境、回声和多路径效应等，以增强编码器对多变环境声条件的适应性及其对声音信息的准确识别能力，确保模型能够在各种实际巡检场景中稳定工作。

9.2.3 设备巡检振动编码器预训练

具体训练策略应包括但不限于：

- a) 对不同类别的设备振动的掩蔽重建：强化编码器对各类设备振动特征（如电机运转振动、齿轮啮合振动、轴承磨损振动等）的表达和识别能力，确保模型能够准确区分不同设备及其运行状态；
- b) 对不同频率振动的掩蔽重建：提升编码器处理不同频段振动（如低频结构振动、高频局部放电振动等）的精确性和灵活性，确保模型能够适应电力设备巡检中多样化振动特征；
- c) 对振动的不同强度和动态变化的掩蔽重建：增强编码器对振动强度变化（如设备启动、过载、停机等）的敏感度与准确度，确保模型能够捕捉设备运行状态的细微变化；
- d) 结合振动与设备运行状态信息的掩蔽重建：提升编码器对振动与设备运行状态关联的推理能力，确保模型能够根据振动信号判断设备的运行模式和健康状况；
- e) 处理噪声干扰和背景振动的掩蔽重建：提高编码器的抗干扰能力及背景振动辨识能力，确保模型能够在复杂环境（如多设备运行、外部机械干扰等）中准确识别目标设备振动信号；

- f) 不同设备运行工况下的整体掩蔽重建：包括正常运行、轻载运行、过载运行、故障运行等，以增强编码器对多变设备工况的适应性及其对振动信号的准确识别能力，确保模型能够在各种实际巡检场景中稳定工作。

9.2.4 设备巡检电气量编码器预训练

具体训练策略应包括但不限于：

- a) 对不同类别的设备电气量的掩蔽重建：强化编码器对各类设备运行电气量特征（如电压幅值、电流波形、功率因数、局部放电特征等）的表达和识别能力，确保模型能够准确区分不同设备及其运行状态；
- b) 对不同时间尺度电气量的掩蔽重建：提升编码器处理不同时间尺度电气量（如瞬态电气量、稳态电气量、长期趋势电气量等）的精确性和灵活性，确保模型能够适应电力设备巡检中多样化的电气量特征；
- c) 对电气量的不同趋势和波动的掩蔽重建：增强编码器对电气量趋势变化（如上升、下降、平稳、突变等）的敏感度与准确度，确保模型能够捕捉设备运行状态的动态变化；
- d) 结合电气量与设备运行环境信息的掩蔽重建：提升编码器对电气量与设备运行环境关联的推理能力，确保模型能够根据电气量信号判断设备所处的运行环境和工况；
- e) 处理噪声干扰和背景电气量信号的掩蔽重建：提高编码器的抗噪能力及背景电气量信号辨识能力，确保模型能够在复杂环境（如电磁干扰、机械振动干扰等）中准确识别目标设备电气量信号；
- f) 不同设备运行周期下的整体掩蔽重建：包括设备启动初期、稳定运行期、停机维护期等，以增强编码器对多变设备运行周期的适应性及其对电气量信号的准确识别能力，确保模型能够在各种实际巡检场景中稳定工作。

9.2.5 设备巡检多模态大模型联合预训练

设备巡检多模态大模型联合预训练的目标是通过视觉编码器、中间件与大语言模型的协同预训练，使得三者之间形成高效的信息交互和融合机制。模型应具备全量参数预训练功能和冻结部分参数预训练功能。

9.3 设备巡检多模态大模型微调

设备巡检多模态大模型宜具备基于轻量化微调技术（如 LoRa）的知识注入与模型微调功能。用于微调的大模型可为通用多模态大模型或电力行业专用多模态大模型。

微调任务类型包括但不限于图文描述、图文问答、目标定位、OCR、指代目标描述与问答、缺陷隐患及违规动作等级判定、处置决策建议等。为确保模型覆盖各类电力场景，具备足够泛化性，应考虑以下方面：

- a) 多源数据覆盖：应采集并整合来自不同环境（如变电站、配电室、输电线路、巡检场景等）的图像、文本、声纹等多模态数据；
- b) 任务多样化设计：模型任务应覆盖多个场景、多个设备类型以及多个任务类型。

9.4 设备巡检不同模态数据融合

在电力设备巡检多模态大模型中，需对视觉图像、文本、声纹、振动类及时序信号等不同模态信息进行融合与对齐。模型应突破各自独立的编码器或特征提取流程，实现多模态间的有效对齐与互动，以生成综合性语义理解。

训练和评测阶段应保持融合方式一致，常见的融合方式应涵盖但不限于以下几种方式：

- a) 早期融合：在对各模态进行初步编码后立即合并特征，这种融合方式强调在特征提取的早期阶段就实现模态间的交互，从而充分利用各模态之间的互补信息；
- b) 中期融合：先分别提取各模态的高层特征，再通过 Cross-Attention、Q-Former 等中间模块实现对齐，这种融合方式强调在高层语义特征上进行模态间的交互，从而更好地理解各模态之间的关联；
- c) 晚期融合：各模态独立完成特征提取与推断后再简单合并或投票，这种融合方式强调各模态的独立性，适用于模态间关联性较弱的场景。

10 模型评测

10.1 评测维度

10.1.1 通用能力

通用能力应包含图文、文音、图音、图文音的图文检索、图片问答、图表推理、文音检索、视频异常检测等能力，可根据场景选择评测维度：

- a) 安全能力：安全性基本要求遵循《TC260—003 生成式人工智能服务安全基本要求》；
- b) 服务能力：服务能力宜选择首词元返回平均时延、平均响应时延、每百字平均响应时延迟；
- c) 视觉处理能力：通用视觉任务中，应具备OCR识别和人脸识别能力，可识别电力票据、表单信息，并进行人脸检测和搜索，以提升票据管理和身份验证的效率。
- d) 其他：其他专业推理能力评测宜考虑电气量计算、电力管理规定判断、电力信息差异性对比、电力故障缺陷推理、电力公式推导证明、电力客观知识推理、电力因果关系推理等任务；其他专业知识记忆和理解能力评测宜考虑电力事实知识、电力管理规定、电力操作流程、电力标准条款、电力机理原理、电力概念定义、电力方案方法等任务。

10.1.2 设备巡检专业能力

设备巡检专业能力指电网设备巡检多模态大模型在电力领域图文相关的能力，应包含输电巡检、变电巡检两个专业场景：

- a) 输电巡检：针对电力输送线路设备以及通道的巡检业务。应包含输电设备缺陷检测、无人机输电通道智能巡检、设备机理原理分析等任务；设备推理能力评测宜考虑设备缺陷定级、电力信息差异性对比、电力故障缺陷推理等要素；设备生成能力评测宜考虑电力故障缺陷报告生成、电力工作票生成、巡视计划编排等任务；
- b) 变电巡检：针对变电站一次及二次设备的巡检业务。应包含变电设备状态感知、设备时序状态预测、二次设备多源数据融合、设备机理原理分析等任务；设备推理能力评测宜考虑设备缺陷定级、电力信息差异性对比、电力故障缺陷推理、设备关联关系拓扑等要素；设备生成能力评测宜考虑电力故障缺陷报告生成、电力工作票生成、巡视计划编排等任务。

10.1.3 模型推理结果可解释性

在对面向电力设备巡检的多模态大模型进行评测时，除了关注其输出结果的准确性和鲁棒性之外，还应评估模型的可解释性。模型可解释性宜考虑以下几个方面：

- a) Chain-of-Thought（思维链）输出：在电力设备巡检场景中，多模态大模型应结合图像、文本、声纹等多种数据源进行综合判断。评测时宜检查模型的思维链输出是否合乎逻辑、与最终答案是否一致，以及在多大程度上展示了模型的“内在推理”；
- b) 模型中间特征可视化：中间特征可视化指的是在模型各层或各模块的输出进行可视化，帮助理解模型如何对输入进行编码、提取、变换以及最终做出决策；
- c) 参数敏感度分析：参数敏感度分析宜考虑当模型内部某些参数或网络结构发生微小改变时，对最终输出的影响程度，从而衡量参数或模块的重要程度。

10.2 评测指标

10.2.1 电网设备巡检视觉问答类多模态大模型评测指标

在对电网设备巡检视觉问答类多模态大模型进行评测时，可采用准确率和CircularEval准确率作为评测指标。

10.2.2 电网设备巡检图文检索类多模态大模型评测指标

在对电网设备巡检图文检索类多模态大模型进行评测时，可采用图像到文本检索排名前K的召回率($i2t_R@K$)、文本到图像检索排名前K的召回率($t2i_R@K$)作为评测指标。

10.2.3 电网设备巡检图文生成类多模态大模型评测指标

在对电网设备巡检图文生成类多模态大模型进行评测时，可采用 Frechet Inception Distance (FID) 和 CLIPScore 作为评测指标。

10.2.4 其他指标

电网设备巡检多模态大模型评测数据集中简答题等主观题型，可使用 Rouge、BLEU 等客观指标作为辅助，同时可采用人工评分准确率指标。

10.3 评测数据

应与模型训练所用的数据集采用相同的采集标准和标注规范，见前文第 5 节和第 7 节。

10.4 评测方法

10.4.1 推理阶段

在对设备巡检多模态大模型进行评测时，应确保输入数据形式、数据预处理、编码方式、推理超参数和推理方式等均与训练阶段设置保持一致，以确保评测结果的可信性、可复现性和实际场景适用性。

a) 提示词 (prompt) 配置应遵循以下要求：

- 1) 训练与推理时均在模型输入中包含 prompt 信息，确保模型能根据提示词进行相同方式的上下文理解；
- 2) 在使用微调模型时，文本提示词应尽量与训练时保持一致；若直接对基础模型进行测试，文本提示词则需清晰阐述相关业务背景，并对模型的输出内容、目标和格式进行严格限制；
- 3) 模型提示词配置，可使用 zero-shot 和 few-shot 两种提示词；
- 4) 基础大模型和预训练大模型可使用 few-shot 提示词；
- 5) 具备对话能力的大模型可使用 zero-shot 提示词。

b) 输入的图像、文本和声纹应满足：

- 1) 图像尺寸统一：在训练和推理阶段，图像应经过相同的预处理，包括尺寸调整、归一化等，建议保持固定分辨率或最长边策略，以匹配模型中图像编码器的输入要求；
- 2) 视觉编码器：推理时应沿用训练时使用的图像编码器及前向传播流程，避免更换编码器或额外添加预处理步骤，从而保证模型在图像理解上的一致性；
- 3) 声纹输入采样率和时间长度一致：与训练时相同的音频采样率（如 16kHz/48kHz 等）和音频时长截断/切分策略，保证声纹特征提取的一致性；
- 4) 文本输入：保持同样的分词、编码模式，使用与训练一致的分词器、词典或 tokenization 策略。

c) 训练和推理时应该保持超参一致性。

d) 在特定应用场景下，可采用 RAG 框架来构建更具针对性和实时性的智能问答或内容生成系统。

e) 除了以上因素外，推理过程还应遵循以下要求：

- 1) 时效性：推理过程中应根据软硬件及目标难度评估推理时间，保证推理时效性满足需求；
- 2) 可用性：模型格式和接口应符合广泛的标准，以便于获取和使用；
- 3) 多多样性和代表性：应涵盖不同的背景、场景、领域等，确保推理结果能够覆盖不同的使用情况；
- 4) 合规性和隐私保护：推理过程使用的数据构建应遵循适用的法规和隐私保护标准，保护用户隐私。如通过用户问卷收集、人类专家构建、权威数据集筛选等方式进行评估数据集的构建；
- 5) 评估指标完备：应为每个评价指标构建满足推理目标。

10.4.2 评分阶段

- a) 测试集准备时包含图像、文本、声纹等多模态数据，确保与训练集在分布和难度上保持一致，不同场景数据集的制作应遵循《T/CES 289-2024 电力人工智能模型场景化评价指南》。
- b) 测试题型包含主观题和客观题。
- c) 搭建推理流程

- 1) 在评测时，使用与训练时一致的输入形式和处理流程（包括 prompt、图像预处理、声纹采样率等），避免因流程不一致导致的性能偏差。
 - 2) 推理时的超参数（如 temperature、max_length、beam-search 参数等）保持与训练阶段一致，以便结果可比。
- d) 评分方法和指标
- 1) 客观题可采用模型推理过程中的随机混淆，和验证数据集 2 次混淆后平均加权分作为最终评分；
 - 2) 客观题的评分方法，应包含基于正则表达式的关键词匹配评分，基于裁判大模型的直接评分方法；
 - 3) 主观题的评分方法，可采用 BLEU 和 ROUGE 等客观指标作为验证指标，也可基于人类专家的评分方法；
 - 4) 根据任务类型选取合适的指标，不同任务类型对应不同的指标，详见附录 B。
- e) 需要纳入标准的客观评价考核点包括：
- 1) 推理时间：模型从接收到输入数据到产生输出结果的时间。推理时间越短，模型的响应速度越快，单位为毫秒；
 - 2) 吞吐量：单位时间内模型可以处理的推理任务数量。吞吐量反映了模型的整体处理能力，通常用于评估模型在批量推理中的表现。单位为每秒推理次数；
 - 3) 显存/内存占用：推理时模型占用的显存或内存资源量，以 GB（千兆字节）为单位；
 - 4) 批处理延迟：当模型处理多个输入时，批量推理所需的时间。通过批处理，可以提高吞吐量，但批处理延迟有时会增加，单位为毫秒；
 - 5) 能效：模型推理阶段的能耗。对于大规模推理任务，尤其是在边缘设备上，能效是一个关键指标；
 - 6) 响应时间与稳定性：模型对用户请求的响应速度以及其在不同负载下的稳定性。

10.4.3 评测综合分析

总结分析可基于能力维度分别汇总大模型评测分数，如有对比大模型，可展示对比分数；专项分析可基于专业上分项能力汇总大模型评测分数，如有对比大模型，可展示对比分数。

11 模型部署

11.1 推理平台部署

11.1.1 硬件适配

- a) 不同硬件平台在架构、算力和指令集上存在差异。在部署前，应确认大模型框架或推理引擎是否针对目标硬件进行了优化，或是否提供了相应的移植方案；
- b) 应评估硬件平台的 AI 加速能力，如 GPU、NPU、ASIC 等，并检查其对硬件加速库（如 CUDA、ROCM、CANN、Cambricon 库等）的支持情况。

11.1.2 操作系统适配

- a) 适配目标操作系统时，应确认操作系统内核版本、软件源兼容性、依赖库版本等；
- b) 应确保所需的驱动、编译器和依赖工具链能够在目标操作系统上正常运行。

11.1.3 安全性和隐私保护

- a) 数据安全：在部署大模型时，应确保数据的安全性，确保数据在存储、传输和计算过程中的保密性，尤其是在涉及敏感数据的场景中；
- b) 模型安全：大模型可能面临反向工程、黑盒攻击等安全威胁，部署过程中应采取加密技术、模型加固等措施，确保模型的安全性。

11.2 模型推理加速

11.2.1 高并发需求

- a) 大模型在推理阶段往往需要处理大量的并发请求，应在目标平台上做好线程数、进程数以及协程等资源使用的平衡；
- b) 宜考虑负载均衡、多实例部署、集群扩容等手段，实现分布式水平扩展，以支持大规模并发。

11.2.2 模型量化

- a) 大模型通常计算量大、参数量大，宜考虑量化（如 FP16、INT8 或更低精度）技术减小模型大小、提高推理吞吐量并降低硬件资源占用；
- b) 不同硬件对量化精度支持程度不同，应测试和评估量化对模型精度与性能的影响，选择合适的量化策略。

11.2.3 推理框架优化

- a) 应根据算力平台选择合适的推理引擎，以充分发挥硬件加速能力；
- b) 宜考虑通过编写针对硬件特性的算子实现或融合算子库，获取最佳推理效率。

11.3 镜像化部署与服务化推理

11.3.1 镜像化部署

- a) 大多数多模态大模型落地时，会采用 Docker 镜像的方式进行封装，以便快速打包、部署和迁移；
- b) 应将推理框架、依赖库、模型文件以及必要的脚本统一打包到镜像中，减轻部署时依赖不一致的问题。

11.3.2 模型服务化

- a) 应通过模型服务化（如 gRPC/RESTful API）对外提供推理接口，便于应用调用；
- b) 应使用常见的服务管理或编排平台（如 Kubernetes、Docker Swarm）进行弹性扩展，提高高并发场景下的可用性与扩展性。

11.3.3 版本管理与灰度发布

- a) 宜考虑通过容器镜像的版本化，对不同模型版本或推理加速方案进行灵活切换；
- b) 在有新版本模型或优化方案上线时，宜先进行小流量灰度测试，再逐步扩大流量范围，降低部署风险。

附录 A
(规范性附录)
输电山火识别开发示例

A.1 输电山火任务设置

- a) Prompt: 判断图中是否存在山火烟雾? 若有, 请描述具体位置, 并判断是否影响输电设施;
- b) 任务类型: 视觉问答, 输入图片和文字 prompt, 输出山火烟雾是否存在及其影响分析;
- c) 模态类型: 图像+文本 (包括可见光和红外图像)。

A.2 输电山火数据采集

- a) 公开数据集: 数据量百万以上, 来源包括网络数据集、图片搜索引擎、社交媒体等;
- b) 山火数据集: 确保数据来源多样, 结合可见光图像、红外图像、激光雷达、卫星遥感等多源数据, 重点覆盖易发生山火的输电线路区域, 考虑光照、天气等因素, 保证数据的高分辨率和鲁棒性。

A.3 输电山火数据集清洗

- a) 清理过曝/过暗数据: 使用图像处理工具去除因光照不佳导致的异常图像, 调整亮度恢复关键区域细节;
- b) 清理模糊数据: 通过清晰度检测算法移除模糊图像, 确保重要特征清晰可辨;
- c) 去重: 使用特征比对技术去重, 保证数据集的唯一性和完整性。

A.4 输电山火数据集标注

- a) 问题提示: 引用相关标准和山火防控指南;
- b) 回答要求: 标注场景特征、目标种类、目标状态和判别依据;
- c) 标注框要求: 目标物体应至少 30×30 像素, 轮廓清晰, 无间隙。

A.5 输电山火多模态大模型预训练和微调

使用公开数据集进行预训练, 并利用专门的山火数据集进行微调, 确保模型能准确识别山火场景。

A.6 输电山火多模态大模型评测

评测模型是否能判断图像中的山火烟雾, 评估模型可解释性, 并通过视觉问答类评分指标 (如准确率和 CircularEval 准确率) 进行评测。

A.6.1 输电山火多模态大模型评测注意事项

- a) Prompt 对齐: 训练和推理时使用一致的提示词;
- b) 数据对齐: 确保数据预处理和数据来源一致;
- c) 推理流程对齐: 输入形式、处理流程和超参数设置需与训练阶段一致, 确保评测结果可比。

附录 B
(资料性附录)
评测指标计算公式

B.1 视觉问答类评分指标计算公式

B.1.1 准确率

表示模型输出正确的问题数占总题目的比例。计算公式如下：

$$\text{Acc} = T / (T + F + N)$$

Acc: 准确率； **T:** 回答正确的问题数量； **F:** 回答错误的问题数量； **N:** 表示模型未给出答案的问题数量。

B.1.2 CircularEval 准确率

对于每一道题目，如有 **N** 个选项，会把这些选项向右移动 **N-1** 次，构造出 **N** 道题目相同、选项位置不同的题目。例如，原始的选项为 A:1 B: 2 C: 3，经过一次移动后，则变为 A:3 B:1 C:2，只有当这些题全部被答对，才被认为模型对这个问题回答正确。计算公式如下：

$$\text{CircularEval} = T / N$$

CircularEval: 循环评估中的预测准确率； **T:** 循环评估中预测正确的次数； **N:** 循环评估中完成的总预测次数。

B.2 图文检索类评分指标计算公式

B.2.1 i2t_R@K

表示当图文匹配任务进行图像到文本检索评测 (i2t test) 时，模型将与查询图像匹配的全部 5 个句子排名到前 **K** 名当中（取 5 句当中的最高排名计算）的平均召回率。在 Flickr30k, COCO Caption 等图文比 1:5 的数据集上的 i2t test 就是使用 i2t_R@K (**K=1, 5, 10**) 作为默认的 Recall 基本度量指标，计算公式如下：

$$\text{i2t_R@K} = \text{Number of relevant texts retrieved in top-K} / N$$

i2t_R@K: 图像到文本检索排名前 **K** 的召回率； **Number of relevant texts retrieved in top-K:** 是在模型返回的前 **K** 结果中，与查询图像相关的文本数量； **N:** 与查询图像相关的所有文本的总数。

B.2.2 t2i_R@K

表示当图文匹配任务进行文本到图像检索评测 (t2i test) 时，模型将与查询句子匹配的唯一 1 个图像排名到前 **K** 名当中的平均召回率。在 Flickr30k, COCO Caption 等图文比 1:5 的数据集上的 t2i test 就是使用 t2i_R@K (**K=1, 5, 10**) 作为默认的 Recall 基本度量指标。计算公式如下：

$$\text{i2t_R@K} = \text{Number of relevant images retrieved in top-K} / N$$

i2t_R@K: 文本到图像检索排名前 **K** 的召回率； **Number of relevant images retrieved in top-K:** 是在模型返回的前 **K** 个结果中，与查询文本相关的图像数量； **N:** 是与查询文本相关的所有图像的总数。

B.3 图文生成类评分指标计算公式

B.3.1 Frechet Inception Distance (FID)

表示同一批文本 Prompt 上的真实图像与生成图像在特征空间中分布之间的距离，衡量生成图像的质量和多样性。分布越接近，则评价结果越好，反之则越差。计算公式如下：

$$FID(Pr, Pg) = \|\mu_r - \mu_g\| + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

Pr, Pg: 多维高斯分布，参数分别为 $(\mu_r \text{ 和 } \Sigma_r)$ 和 $(\mu_g \text{ 和 } \Sigma_g)$ ；
 μ : 均值向量和； Σ : 协方差矩阵；
Tr: 矩阵的迹。

B.3.2 CLIPScore

表示生成的图像和 prompt 放入到特征提取器 (CLIP) 中获取到的 embedding 之间的余弦距离。计算公式如下：

$$\text{CLIPScore}(I, T) = \cos(EI, ET)$$

I: 生成的图像；T: 给定的文本 (prompt)；EI: 图像 I 通过 CLIP 模型提取的特征嵌入 (embedding)；ET: 文本 T 经 CLIP 模型提取的特征嵌入 (embedding)。

B.4 其他指标计算公式

B.4.1 Rouge

表示生成文本与参考文本之间的重叠程度，侧重于召回率。常见的变体包括 ROUGE-N、ROUGE-L 等，以 Rouge-L 为例。计算公式如下：

$$\text{Rouge-L} = \text{LCS}(X, Y) / \text{Len}(Y)$$

Rouge-L: 生成文本和参考文本之间的相似度；LCS(X, Y): 是 X 和 Y 的最长公共子序列的长度；
Len(Y): 参加文本的长度。

B.4.2 BLEU

表示机器生成的翻译与人工参考翻译之间的相似度。计算公式如下：

$$\text{BLEU} = \text{BP} \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right)$$

BLEU: 生成文本与参考文本的匹配程度； p_n : n-gram 精度； w_n : n-gram 的权重，通常取均匀权重 $(1/N)$ ；BP: 简短惩罚因子。

附录 C
(资料性附录)
设备巡检标注样例

C.1 设备巡检视觉类标注示例

任务背景：对变电站变电设备进行渗漏油检测，通过多模态大模型对设备图像进行分析，识别设备是否存在渗漏油问题，并评估其风险等级。任务要求对渗漏油进行详细标注，包括设备状态、缺陷类型、风险等级及处置建议，以便运维人员快速响应和处理。

a) 示例图像：[插入示例图像]

b) 标注结果：

- 1) 场景类型：变电站环境；
- 2) 目标名称：变压器表面油污；
- 3) 目标位置：[矩形框标注]；
- 4) 目标状态：异常，存在油污；

依据：变压器表面发现疑似油液渗漏的区域，该区域范围较大，且区域颜色较深，并存在往周围扩散的迹象，表明存在渗漏油现象，可能影响设备正常运行。

5) 图文描述：图中所示为变电站，变电站内右侧区域设有一台变压器，变压器处于正常运行状态，但表面发现明显油渍，表明存在渗漏油问题。

6) 图文问答：

问题：图中是否存在油液渗漏的迹象？

答案：存在油液渗漏的迹象。

c) 缺陷隐患判定：

1) 缺陷类型：油液渗漏；

依据：变压器底部发现明显油渍，油渍范围较大且颜色较深，符合渗漏油的特征。

2) 判定等级：中等风险；

依据：渗漏油可能导致变压器绝缘性能下降，存在短路或火灾风险，但尚未影响设备正常运行，因此判定为中风险

3) 处置建议：安排专人进行维护，检查渗漏油原因并及时处理，避免问题进一步恶化。

依据：根据变电站设备维护规程，渗漏油问题需立即处理，以防止设备故障或安全事故。