

团体标准

T/CES XXX-XXXX

输电线路金具与设备图像样本标注要求与规范

Requirements and specifications for marking image samples of
transmission line fittings and equipment

(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国电工技术学会 发布

目 次

| | |
|----------------------|---|
| 前言..... | 1 |
| 1 范围..... | 1 |
| 2 规范性引用文件..... | 1 |
| 3 术语和定义..... | 1 |
| 4 缩略语..... | 3 |
| 5 设备类别..... | 3 |
| 6 样本基本要求..... | 3 |
| 6.1 图像视频类样本基本要求..... | 3 |
| 7 样本标注要求..... | 4 |
| 7.1 图像视频类样本标注要求..... | 4 |
| 8. 样本标注流程..... | 6 |
| 8.1 总体要求..... | 6 |
| 8.2 样本检查..... | 6 |
| 8.3 安全管控..... | 6 |
| 8.4 标注工具..... | 7 |
| 8.5 标注任务开展..... | 7 |
| 8.6 样本标注结果收集..... | 7 |
| 8.7 样本标注结果检查..... | 8 |

前 言

为规范电力人工智能平台中输电线路金具图像样本标注的基本要求，为电力人工智能平台支撑输电线路自主巡检工作开展提供指导规范，制定本文件。

本文件按照 GB/T1.1—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》给出的规则起草。

本文件由四川中电启明星信息技术有限公司提出。

本文件由中国电工技术学会标准工作委员会能源智慧化工作组归口。

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。

本文件起草单位：国网信息通信产业集团有限公司、四川中电启明星信息技术有限公司、国网重庆市电力公司、重庆大学。

本文件主要起草人：李强、宋卫平、向哲宏、吕小红、刘礼、崔秋实、赵峰、李炳森、厉仄平、徐小云、谷波、陈昌平、高攀、何亮、杨平、朱海萍、孙觉予。

本文件为首次发布。

人工智能输电线路金具图像样本标注要求与规范

1 范围

本文件规定了输电线路金具样本以及电一二次设备图像视频样本在人工智能平台标注的基本要求、标注要求和标注流程。

本文件适用于进行人工智能模型开发时的样本标注和样本入库的统一管理，包括样本的质量管控、样本标注的技术要求和流程管控。

输变电线路金具管理是电力系统设备管理的一项重要内容，通过规范设备各类缺陷的描述和定性，提高缺陷分析水平，适应标准化和精益化管理的要求，确保输变电设备的安全稳定运行。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本

GB/T 5271.28—2001 信息技术 词汇 第 28 部分;人工智能 基本概念与专家系统

Q/GDW 1560.1—2014 输电线路图像/视频监控装置技术规范 第 1 部分：图像监控装置

T/CESA 1040—2019 信息技术 人工智能 面向机器学习的数据标注规程

Q/GDW 12118.1~12118.3—2021 人工智能平台架构及技术要求

Q/GDW 1906—2013 输变电一次设备缺陷分类标准

GB37300-2018 公共安全重点区域视频图像信息采集规范

GB35114-2017 公共安全视频监控联网信息安全技术要求

GB/T 39272-2020 公共安全视频监控联网技术测试规范

GA/T1127-2013 安全防范视频监控摄像机通用技术要求

GB/T 5075-2016 电力金具名词术语

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能 artificial intelligence

一门交叉学科，通常视为计算机科学的分支，研究表现出与人类智能（如推理和学习）相关的各种功能的模型和系统。

[来源：GB/T 5271.28—2001, 定义28.01.01]

3.2

样本数据 sample data

其具备的特征能够反映总体数据情况的一部分个体数据。

[来源：Q/GDW 12118.1—2021, 定义 3.5]

3.3

标注 corpus annotation

采用人工或计算机自动方式对样本的属性或特征进行描述。

[来源：Q/GDW 1906—2013, 定义 3.5]

3.4

分辨率 resolution

衡量数字图像内所包含数据量多少的技术参数。这里指一帧图像在水平和垂直方向的有效像素数。

[来源: Q/GDW 12118.3—2021, 定义 3.1]

3.5

视频码率 video code rate

视频文件在单位时间内使用的数据流量

3.6

标签 label

标识数据的特征、类别和属性等内容, 可用于建立数据及深度学习训练要求所定义的机器可读数据编码间的联系。

[来源: T/CESA 1040—2019 定义 3.2]

3.7

边界框 bounding boxes

图像标注的一种方法, 通过在需要标注的对象的外边缘放置多个点来绘制成线进行标注。有时候, 图像中的对象由于光照或角度等原因, 其形状、大小或方向无法被很好地适配上矩形框。使用边界框可以进行更加精确的标注。

3.8

线条 line

图像标注的一种方法, 利用线条在图像的不同区域之间划定边界。

3.9

关键点 keypoint

图像标注的一种方法, 在规定位置标注上若干点, 例如人脸特征点、人体骨骼连接点等, 常用来训练面部识别模型以及统计模型。

3.10

标注工具 annotation tool

标注人员执行标注任务生成标注结果的过程中使用的工具和软件。标注工具按照自动化程度分手动、半自动和自动三种。

[来源: T/CESA 1040—2019 定义 3.5]

3.11

半自动化标注 semi-automatic annotation

基于少量人工标注、机器预标注来训练标注模型, 用于批量标注样本数据的半人工智能标注方法。

[来源: T/CESA 1040—2019 定义 3.10]

3.12

设备分层 Device layering

输变电设备分层是依据国家电网公司设备分类标准, 按结构、功能、特性等进行逐级分层, 依次分为: 设备类型、设备种类、部件、部件种类以及部位五层:

设备类型: 变电一次设备的设备类别按照国家电网公司设备分类标准, 分为线夹类、绝缘子类、塔身类、塔基、联接金具等六类。

设备种类: 设备种类为设备类别的补充, 根据设备的不同类型进行划分。如绝缘子分为瓷制绝缘子和玻璃绝缘子。

部件: 部件是设备中具有相对独立功能或作用的部分。如螺栓螺母为耐张线夹的部件。

部件种类: 部件种类为部件的补充说明, 根据部件的不同类型进行细分。

部位：是指缺陷发生的具体位置。

4 缩略语

下列缩略语适用于本文件。

BIOES: BIOES 标注模式 (b-begin, i-inside, o-outside, e-end, s-single)

COCO: 上下文中公共对象 (common objects in context)

JPEG: 联合图像专家组 (joint photographic experts group)

Json: JS 对象简谱 (javaScript pbject notation)

VOC: 视觉对象类 (visual object classes)

WAV: 窗波 (windows wave)

XML: 可扩展的标记语言 (extensible markup language)

5 设备类别

根据国家电网公司缺陷库编制会议讨论结果，输电线路金具的类别以及一二次设备在原生生产系统台账的基础上进行了适当修改，划分为：瓷质绝缘子，玻璃绝缘子，复合绝缘子，U型螺丝，直角挂板，油位计，油温计，六氟化硫开关，接地线等设备。

6 样本基本要求

6.1 图像视频类样本基本要求

6.1.1 图像视频文件存储格式要求

图像视频文件存储应满足下述要求：

图像文件应以 jpg、jpeg、png、bmp 格式进行储存；

视频文件应以 mp4、mov 格式进行存储，视频文件根据业务需求对其部分帧进行截取时，截取图片应以 jpg、jpeg、png、bmp 格式进行储存。

6.1.2 图像视频文件命名要求

图像视频文件名称应满足下述要求：

1. 图像样本名称应由三个部分组成：

- a. 巡检的线路名称；
- b. 巡检拍摄角度；
- c. 当前图片包含的金具类型；

2. 视频样本名称应由两个部分组成：

- a. 巡检的线路名称；
- b. 巡检的起始和结束位置；

3. 视频样本根据业务需求需对其部分帧进行截取时，截取图像名称应由四个部分组成：

- a. 源视频样本名称；
- b. 文件唯一性编号，从 1 开始计数；
- c. 巡检拍摄角度；
- d. 当前图片包含的金具类型；

6.1.3 图像视频类样本质量要求

图像视频样本质量应满足下述要求：

分辨率：图像样本与视频抽帧样本分辨率宜为 1920*1080 像素及以上，识别主题内容不应少于 15*15 像素要求。识别主题内容边缘清晰，无严重重影、遮挡范围不超过主题三分之二；

视频码率应满足视频包含业务相关内容，流畅、清晰，满足编解码格式需求。分辨率-码率宜为 1920*1080 像素-5Mb/s，在输电线路沿路视频质量可为 352*299 像素-512kb/s。

样本类别的全面性：样本集宜涵盖目标业务的所有类别；
 样本应从同一处金具 4 个不同角度进行拍摄；
 单类样本的多样性：单类样本集宜涵盖不同时间、天气、角度等情况；
 样本目标物体的完整性：目标区域在整个图像样本中的占比不应低于 40%。

7 样本标注要求

7.1 图像视频类样本标注要求

7.1.1 图像样本标注规则

图像样本标注规则应满足下述要求：

- 图像中出现的所有满足 6.1.3 分辨率要求的目标物体应全部标注；
- 应采用包含全部目标的情况下使用最小标注框的方法对目标物体进行标注, 在画框的过程中, 需保证所框选的图像包含命名金具的所有部位, 并且尽量保证避免在所选框内出现其他金具类型, 如果无法进行较好的避免, 则需要对误入的其他金具添加该金具名称的标注框图以便于后期电脑识别时效果更好；
- 目标物体前端遮挡不宜超过 2/3, 若目标物体存在过大比例（大于 2/3）的遮挡情况, 应进行目标物体轮廓标注后, 对被遮挡的目标物体添加“遮挡”标签；
- 标注时应使用 COCO 或 VOC 格式存储标注内容。

7.1.2 标注命名规则

在对样本图片进行标注的过程中, 标注框的命名应该严格按照:一级分类:金具类型+二级分类:金具名称+三级分类:金具部件+四级分类:缺陷名称的四级分类的方式来严格进行, 其中金具类型是指输电线路中所有金具的类别命, 包含但不限于绝缘子, 联接金具, 横担, 塔身等, 金具名称为该金具类别中的金具名称, 包含但不限于复合绝缘子, 直角挂板, 悬垂线夹等, 金具部件则是指金具上具体的部件位置, 包含但不限于销子, 螺栓, 联板, 挂板等, 而四级分类的缺陷名称是指对缺陷特征的规范化描述, 反映缺陷发生的具体部位和现象, 见表 1:

表 1 标注框的命名规则

| 一级分类 | 二级分类 | 三级分类 | 标准名称 |
|------|------|-------|------------------|
| 联接金具 | 直角挂板 | 销子 | 联接金具_直角挂板_缺销子 |
| | | 本体 | 联接金具_直角挂板_锈蚀 |
| | | 销子 | 联接金具_直角挂板_销子脱出 |
| | | 本体 | 联接金具_直角挂板_变形 |
| | | 螺母 | 联接金具_直角挂板_螺母欠扣 |
| | | 销子&螺栓 | 联接金具_直角挂板_缺销子缺螺母 |
| | | 本体 | 联接金具_直角挂板_安装不规范 |

表 2 变电一二次设备标注名称示例

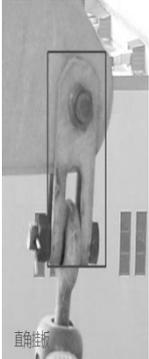
| 一级分类 | 二级分类 | 标注名称 |
|------|--------|------------|
| 变压器 | 油位计 | 变压器_油位计 |
| | 油温计 | 变压器_油温计 |
| | 六氟化硫开关 | 变压器_六氟化硫开关 |
| | 接地线 | 变压器_接地线 |

| | | |
|--|-------|-----------|
| | 高低压套管 | 变压器_高低压套管 |
| | 呼吸器 | 变压器_呼吸器 |
| | 分接开关 | 变压器_分接开关 |
| | 储油柜 | 变压器_储油柜 |

标注图像的命名规则应严格参照设备分级划分，同属一个设备的附属部件应在命名时以下划线分割上级设备与下级设备，如在进行变压器附属部件油位计标注时应标记为：变压器_油位计。

例如：在联接金具中直角挂板上的螺栓出现锈蚀的问题，标注框的名称应为：联接金具_直角挂板_螺栓_锈蚀，并在表格最后附上该缺陷的标准样本图片作为后期标注参考。见表 3：

表 3 标注框的命名规则

| | | | | |
|------|---|------------|--|-----|
| 联接金具 |  | 直角挂板_缺销子 |  | 常见 |
| | | 直角挂板_锈蚀 |  | 常见 |
| | | 直角挂板_缺螺母 |  | 不常见 |
| | | 直角挂板_安装不规范 |  | 不常见 |
| | | 直角挂板_变形 |  | 不常见 |
| | | 直角挂板_螺母欠扣 |  | 不常见 |
| | | 直角挂板_销子脱出 |  | 常见 |

7.1.3 矩形框

使用矩形框对图像样本进行标注应满足下述要求：

- a. 目标物体应全包含在标注框内，除非目标物体有过于细长伸出的部位（伸出部分占像素比例小于 5%）；
- b. 矩形框应将目标物体全部包括，矩形框边缘与目标物体的距离应小于矩形框边长 3%；

7.1.4 边界框

使用边界框对图像样本进行标注应满足下述要求：

- a. 边界框应将目标物体全部包括，边界框边缘与目标物体的距离应小于边界框边长 3%；
- b. 对混合的像素（由于运动，透明性等）如果能确认属于目标物体，也应进行标注；
- c. 边界框标注时标注点个数应大于等于 4 个。

7.1.5 线条

使用线条对图像样本进行标注应满足下述要求：

- a. 两个标注点间的距离应不小于 35 个像素；
- b. 每个标注点都应在目标物体的边界点。

7.1.6 关键点

使用关键点对图像样本进行标注应满足下述要求：

- a. 标注信息应包含关键点坐标 (x, y) 和标签信息，且标签信息应采用“是否可见-标签名”的格式；
- b. 如果标注点被遮挡，则应标注图片中距离该点最近的点，且使用二进制值“0”来表示不可见，使用二进制值“1”来表示可见。

7.1.7 旋转矩形

使用旋转矩形框对图像样本进行标注应满足下述要求：

- a. 标注信息内容应包含矩形框中心点坐标、宽、高和沿垂直方向顺时针旋转的角度；
- b. 旋转角度应采用弧度制，同一批数据集标注时应采用同一种旋转方式。对于规则的目目标物体，旋转方向宜尽量与设备轴向方向保持一致，标注方向与目标物体轴向方向的角度偏差不宜超过 5%。
- c. 标注文件命名与存储要求
- d. 标注文件应与对应标注图像命名一致，应保存为 XML 或 Json 格式。

8. 样本标注流程

8.1 总体要求

样本标注应包含样本检查、安全管控、标注工具选择、标注任务开展、标注结果收集和标注结果检查环节。

8.2 样本检查

在样本标注前应按照本文第 6 章要求对待标注样本进行检查，应根据业务需求和样本的数量采用全量检查或抽样检查，方式如下：

- a. 全量检查应对指定范围内的所有样本进行逐条检查；
- b. 抽样检查可采用随机抽样或分层抽样，方式如下：
 - 1) 随机抽样，即：针对不同业务类型的数据样本采用随机抽样进行检查；
 - 2) 分层抽样，即：针对同一业务类型的样本数据，根据样本类型不同采取分层抽样的方式进行检查。

8.3 安全管控

应满足对标注环境及标注人员的安全管控要求。具体要求包括：

- a. 标注过程应在内网环境下的指定机器中进行，机器应开启防火墙，安装杀毒软件，并禁用 USB 接口功能。机器中的所有数据文件需定期做好数据备份，不得擅自拷贝、传输，防止数据丢失或泄露；
- b. 标注人员应经过标注工作培训获得相关单位资格认证并签署样本标注保密协议后方可上岗操作。

8.4 标注工具

应使用电力人工智能平台标注工具或与其格式相兼容的标注工具进行标注。

8.5 标注任务开展

8.5.1 基本要求

应根据业务需求和标注任务难易度选择人工标注或半自动化标注。

8.5.2 人工标注

人工标注任务应按照试标注、批量标注顺序执行，具体要求如下：

- a. 试标注：
- b. 从标注任务的待标注样本中抽取试标注样本。可采用随机抽样或分层抽样方法抽取样本，抽取比例不宜低于待标注样本总量的 1%；
- c. 标注人员对抽取样本进行标注；
- d. 标注项目负责人对标注结果进行确认；
- e. 标注人员重复执行标注错误的标注任务，直至标注项目负责人确认无误。
- f. 标注人员批量执行标注任务。

8.5.3 半自动化标注

半自动标注任务应按照样本构建、模型构建、模型批量标注顺序执行，具体要求如下：
样本构建：

- a. 从待标注样本中抽取训练样本和测试样本。可采用随机抽样或分层抽样方法抽取样本，训练样本与测试样本占样本总量的比例均不宜低于 1%，训练样本与测试样本的比例宜为 7:3，训练样本与测试样本应无交集；
- b. 标注人员通过人工方式标注训练样本和测试样本；
- c. 标注项目负责人对标注结果进行确认；
- d. 标注人员重复执行标注错误的标注任务，采用 10 折交叉验证方式，直至标注项目负责人确认无误。

模型构建：；

- a. 使用标注后的训练样本建立标注模型，半自动标注工具使用 Anno-Mage；
- b. 预训练模型是基于 MS COCO 数据集的，用 RetinaNet 训练的；
- c. 使用标注后的测试样本测试标注模型。可采用召回率、精确率指标评估模型性能；
- d. 使用标注模型批量执行标注任务。

8.6 样本标注结果收集

标注结果收集应满足以下具体要求：

- a. 样本标注结果应由统一的人员进行回收和存放，防止文件外泄；
- b. 标注结果收集人员应对样本标注结果的相关信息（包括任务名称、任务类型、任务开始时间、任务结束时间、任务描述进行核对）进行检查，防止文件遗漏；
- c. 标注结果收集人员，宜按照样本类型（图像视频，语音和文字）和标注方式（图像标注包括图像分类、图像目标检测、图像分割；文本标注包括文本分类、文本标注；音频标注包括音频分类、音频标注）进行分类保存至电力人工智能平台中。

8.7 样本标注结果检查

在样本标注结果收集后应按照本文第7章要求对标注结果进行检查，应根据业务需求和样本标注的数量采用全量检查或抽样检查，方式如下：

- a. 全量检查应对指定范围内的所有样本进行逐条检查；
- b. 抽样检查可采用随机抽样或分层抽样，方式如下：
 - 1) 随机抽样，即：针对不同业务类型的数据样本采用随机抽样进行检查；
 - 2) 分层抽样，即：针对同一业务类型的样本数据，根据样本类型不同采取分层抽样的方式进行检查。