

# 第八届“泰迪杯”数据挖掘挑战赛——

## B 题：电力巡检智能缺陷检测

### 一、问题背景

为了保证输电线路的安全、可靠运行，电网运行部门需要定期对输电线变电系统进行巡检、维修以及维护来确保消除故障或者隐患。随着我国经济的高速发展，对电力输电网设备等基础设施的安全运营也提出了更高的要求。

架空线路巡检作为保障输配电网正常运行的重要手段之一，一直以来都面临着网线分布广、设施布置复杂，巡线作业强度大、周期长，部分区域自然环境复杂恶劣等问题。传统的电网巡查方式是通过人工进行巡检的，该方法存在以下几个问题。

- (1) 劳动强度大，工作效率低，在危险地段会危及到巡查工人的生命安危
- (2) 人工录入数据量大，而且录入过程容易出错
- (3) 对于工人是否巡查到位无法进行有效的管理，巡查质量不能得到保障

近年来由于无人机或者智能机器人技术的飞速发展，考虑通过拍摄的大量电力设备及线路的现场图片代替人工巡检，其基本工作流程如下。

- (1) 划定无人机工作区域，设定巡查时间
- (2) 无人机飞达指定区域，进行图片拍摄（要求尽可能达到 360° 全方位），
- (3) 分析无人机拍摄的图片，并进行问题标注，反馈至调度中心
- (4) 调度中心根据问题，安排对应人员进行检修

但是由于无人机拍摄图片数目多（单个高架塔拍摄图像大于 300 张），尺寸大（4096\*2160），人工进行一张图片标注就需要 5-10 分钟，工作量巨大。同时执行标注工作的相关人员极易用眼疲劳，从而导致漏标，错标。鉴于以上情况，考虑使用图像处理与机器学习（深度学习）的方法，对图片进行标注。

### 二、任务

架空输电线路巡视主要巡视内容包括：杆塔、导线及避雷线、导线及避雷线的固定与链接、绝缘子、拉线、杆上开关设备、沿线路附近的其他工程等 7 大项内容。以上 7 大项内容中的每一项都还有子项，检查内容繁多，流程繁琐。为了能够先行探索出切实有效的步骤，本次任务以绝缘子巡视中的绝缘子自爆这一故障为目标，主要实现以下两部分内容。

#### 1、绝缘子串珠分割

由于无人机图片较大一般为（4096\*2160），绝缘子串珠占据图片中很小的一部分区域，需要设计图像分割算法，对绝缘子串珠所在的区域进行分割。参赛者需要标记出所给样例每幅图像中的标记绝缘子串珠区域的掩模图像。掩模区域仅整个绝缘子串珠。原图与掩模图像对比如图 1 所示。



图 1 原图与掩模图对比

## 2、绝缘子自爆识别和定位

根据分割图像初步识别绝缘子所在的位置，并对绝缘子串珠进行分割，而后参赛者根据所给出的标记样本的 Ground Truth 构建自爆绝缘子识别模型。参赛者利用训练模型对图像中的自爆绝缘子位置进行检测，并利用 BoundingBox 对其进行标记。自爆标记需包括自爆位置周围 2 个完好绝缘子。绝缘子自爆位置标记如图 2 所示。



图 2 样例示范

## 三、数据说明

本赛题数据总共由三部分组成。

- (1) 绝缘子自爆缺陷原图
- (2) 基于原图的标准掩模图像
- (3) 自爆绝缘子 BoundingBox 标签

绝缘子自爆缺陷原图主要是高架线路上拍摄的原图，原图中可能存在自爆点，原图的基本形式如图 3 所示。



图 3 绝缘子自爆缺陷原图

基于原图的标准掩模图像是使用绝缘子自爆缺陷原图掩模之后的图形，通过掩模的方式可以去除背景，凸显自爆点。基于原图的标准掩模图的基本形式如图 4 所示。



图 4 基于原图的标准掩模图

每张绝缘子自爆原图都会对应一个自爆绝缘子 BoundingBox 标签。这个标签以 xml 文件的形式存放，可以使用浏览器、记事本、VSCode、NotePad++等文本编辑软件打开。文件中的各项数据及其意义如表 1 所示。

表 1 自爆绝缘子 BoundingBox 标签

标签文件内容	解释
<pre>&lt;?xml version="1.0" ?&gt; &lt;doc&gt;   &lt;path&gt;.\2.JPG&lt;/path&gt;   &lt;outputs&gt;     &lt;object&gt;</pre>	可忽略
	object 内部存放标签

<code>&lt;item&gt;</code>	item 代表一个标签，可存在多个 item
<code>&lt;name&gt;1&lt;/name&gt;</code>	name 为标签名（可忽略）
<code>&lt;bndbox&gt;</code>	bndbox 内部存放标记位置
<code>&lt;xmin&gt;4045&lt;/xmin&gt;</code>	xmin,xmax,ymin,ymax 表示标记方框的 x、y，可组成标记框的 4 个点，分比为： (xmin,ymin), (xmin,ymax), (xmax,ymin), (xmax,ymax)
<code>&lt;ymin&gt;2440&lt;/ymin&gt;</code>	
<code>&lt;xmax&gt;4276&lt;/xmax&gt;</code>	
<code>&lt;ymax&gt;2700&lt;/ymax&gt;</code>	
<code>&lt;/bndbox&gt;</code>	可忽略
<code>&lt;/item&gt;</code>	
<code>&lt;/object&gt;</code>	
<code>&lt;/outputs&gt;</code>	
<code>&lt;time_labeled&gt;1577170405866&lt;/time_labeled&gt;</code>	size 用于存放图像尺寸信息
<code>&lt;labeled&gt;&gt;true&lt;/labeled&gt;</code>	
<code>&lt;size&gt;</code>	
<code>&lt;width&gt;7360&lt;/width&gt;</code>	
<code>&lt;height&gt;4912&lt;/height&gt;</code>	width 表示原始图像的宽度
<code>&lt;depth&gt;3&lt;/depth&gt;</code>	height 表示原始图像的高度
<code>&lt;/size&gt;</code>	depth 表示原始图像的深度
<code>&lt;/doc&gt;</code>	可忽略

## 四、评价方案

### 1、Dice 系数

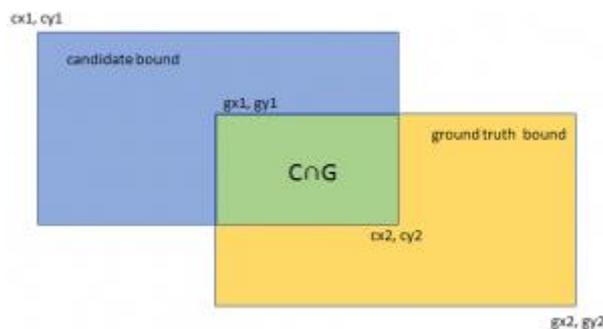
绝缘子串珠分割采用 Dice 系数进行评价，通常计算两个样本的相似度

$$dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

其中，A 为 GroundTruth 区域即专业人士标注的区域，B 为算法分割所得到的区域。Dice 系数取值范围是 [0, 1]，取值越接近 1 则越表明预测的结果与专业人士标注的结果相符合。

### 2、IOU

绝缘子自爆区域评价采用 IOU，IOU 表示产生的候选框（Candidate Bound）与原标记框（Ground Truth Bound）的交叠率或者说重叠度，也就是它们的交集与并集的比值。相关度越高该值。最理想情况是完全重叠，即比值为 1。



$$IOU = \frac{area(C) \cap area(G)}{area(C) \cup area(G)}$$

## 附录：

请仔细阅读以下说明：

### 1、关于赛题数据

- (1) 示例数据：2020年3月1日9:00:00随赛题公布。
- (2) 全部数据：2020年4月11日9:00:00公布。
- (3) 测试数据：2020年4月25日9:00:00公布。

### 2、提交作品

(1) 命名方式：论文命名为“B题”，附件命名为“作品附件”，测试结果命名为“作品测试结果”。

(2) 论文及附件内请勿出现队号、学校、学院、队员以及指导老师相关任何信息，否则该作品视为无效作品。

(3) 请参赛队于2020年4月24日16:00:00之前在竞赛官网“提交作品”处提交论文（PDF版，大小不超过50M）及附件（论文正文（Word版）、源数据（组委会提供的源数据除外）、程序的压缩包，大小不超过200M）。

### 3、公布测试数据，提交测试结果

2020年4月25日9:00:00准时放出测试数据，请在“赛题与数据”页面对应的题目右下方下载测试数据，并于2020年4月26日9:00:00前请在“提交测试结果”页面提交测试结果。