Insulator Defect Image Dataset - Version 1.2: ReadMe

3002017949

This publication is a corporate document that should be cited in the literature in the following manner:

Insulator Defect Image Dataset - Version 1.2: Readme EPRI, Palo Alto, CA: 2020. 3002017949.

ABSTRACT

High quality labeled image data for defective assets is limited in the electric utility industry. For training an Artificial Intelligence / Machine Learning model, standalone images cannot suffice. EPRI, through its Artificial Intelligence initiative, has released a curated and labeled dataset 'Insulator Defect Image Dataset Version 1.0 (IDID V1.0)'.

This readme file provides the dataset overview, characteristics, file structure and labeling format for IDID V1.2, not the imagery itself.

This publication is a corporate document that should be cited in the literature in the following manner:

Insulator Defect Image Dataset - Version 1.2: Readme EPRI, Palo Alto, CA: 2020. 3002017949.

1
DATASET OVERVIEW

The Insulator Defect Image Dataset (IDID) consists of labeled high quality images of transmission line insulators. The images have insulator string as the primary subject and parent class. These images contain 3 sub-classes:

1. Flashover damage insulator shell
2. Broken insulator shell
3. Good insulator shell.

This readme document provides the dataset overview, characteristics, file structure and labeling format for IDID V1.2, not the imagery itself.

2
DATASET DETAILS

IDID is divided in two sets. A training set and a test data set.

Training Set

Details of training dataset are provided in Table 2-1.

Table 2-1
Train set details

Train set	
Total number of images	1596
Total labeled assets	7568
Labeled asset split	
Insulator strings	1788
Good insulator shell	2636
Broken insulator shell	1140
Flashover damage insulator shell	2004

Test Set

Table 2-2 provides characteristics of the test set held by EPRI. The test set is not available for download.

Table 2-2
Test set details

Test set	
Total number of images	88
Total labeled assets	403
Labeled Asset split	
Insulator strings	103
Good insulator shell	147
Broken insulator shell	64
Flashover damage insulator shell	89

3
FILE STRUCTURE

The dataset contains 2 folders named Flashed and Broken. Table 3-1 describes the contents of the two folders.

Table 3-1
Folder structure

Folder	Contents
Flashed	The folder contains a JSON file for labels and a subfolder with the images containing Insulator strings with Good and Flashover Damage insulator shell
Broken	The folder contains a JSON file for labels and a subfolder with the images containing Insulator strings with Good and Broken insulator shells

4
LABELS

The labels/annotations are stored in a Java Script Object Notation (JSON) format to be easily imported and viewed. It is a standard format that allows easy ingestion in ML pipelines. The labels are bounding boxes which are rectangular in shape. Label files have the following structure:

```
[
   {
      "filename":"100009h.jpg",
      "Labels":
      {
         "objects":[
            {
               "name":"insulator",
               "material":"porcelain",
               "type":"unknown",
               "manufacturer":"unknown",
               "conditions":{
                  "glaze":"Flashover damage"
               },
               "string":1,
               "bbox":[
                  X,
                  Y,
                  W,
                  H
               ],
               "comments":"none"
            },
            …..                                          (annotated objects continued)

   …..                                                   (filenames continued)

]                                                        (EOF)
```

Explanation of the fields in the JSON file is provided in Table 4-1.

Filename	Corresponds to the filenames in the dataset
Labels	A dictionary indicating the annotated objects it contains
Objects	A list containing all the annotated objects in that image
Name	A tag to identify the Asset
Material	The material of the Insulator
Type	The configuration of the Insulator
Manufacturer	The manufacturer of the Insulator
Conditions	The condition associated to the asset or a part of the asset
String	A Boolean value that indicates whether the Insulator is a string (1) or an individual disc(0)
bbox	A list containing the Bounding box value of the asset
X	X pixel coordinate of top left corner
Y	Y pixel coordinate of top left corner
W	Width of the asset
H	Height of the asset
Comments	Any additional comments used while Labeling

Table 4-1
Description of the tags in JSON

Figure 4-1 shows the bounding boxes drawn on a damaged insulator string.

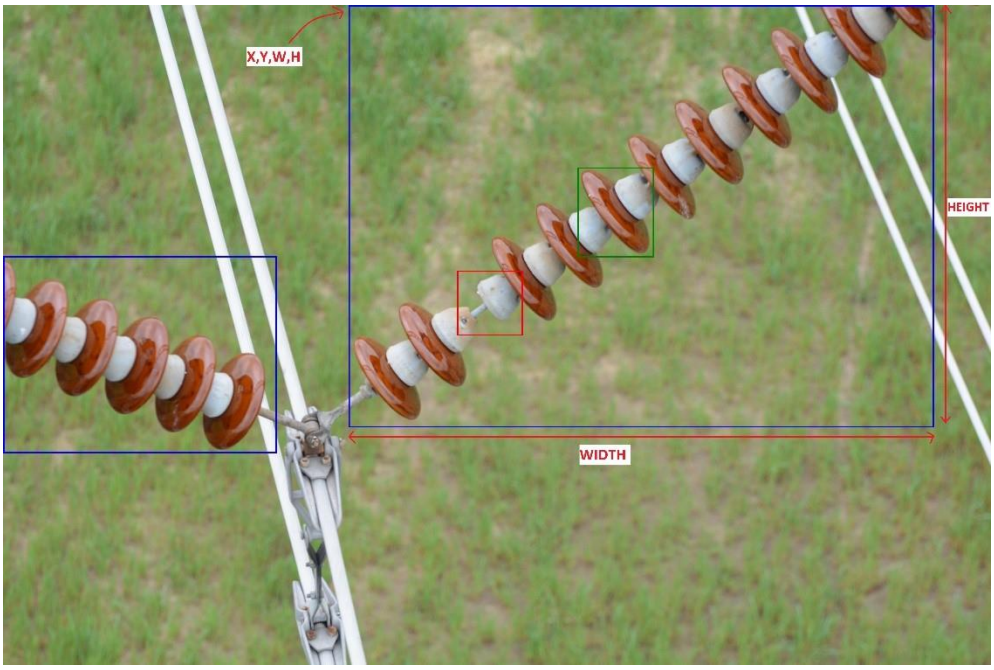

Figure 4-1
Damaged insulator string and its corresponding annotations ID: 150394.jpg