EPRI | ELECTRIC POWER RESEARCH INSTITUTE

Insulator Defect Image Dataset—Version 1.1: Documentation

3002017949

Insulator Defect Image Dataset—Version 1.1: Documentation

3002017949

Technical Update, June 2020

EPRI Project Manager

P. Kulkarni

ELECTRIC POWER RESEARCH INSTITUTE
3420 Hillview Avenue, Palo Alto, California 94304-1338 ▪ PO Box 10412, Palo Alto, California 94303-0813 ▪ USA
800.313.3774 ▪ 650.855.2121 ▪ askepri@epri.com ▪ www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

This is an EPRI Technical Update report. A Technical Update report is intended as an informal report of continuing research, a meeting, or a topical study. It is not a final EPRI technical report.

NOTE

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

ACKNOWLEDGMENTS

The Electric Power Research Institute (EPRI) prepared this report.

Principal Investigators
P. Kulkarni
T. Shaw
D. Lewis

This report describes research sponsored by EPRI.

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

Insulator Defect Image Dataset—Version 1.1: Documentation. 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 V1.1, not the imagery itself.

To download the IDID imagery: This product is available at no cost. Submit a request via email to orders@epri.com that includes your full name, mailing address, phone number, organizational affiliation, and country of citizenship."

Keywords
Insulators
Labeled Data
Defect Detection
Artificial Intelligence
JSON

CONTENTS

LIST OF FIGURES

LIST OF TABLES

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 V1.1, not the imagery itself.

To download the IDID imagery: This product is available at no cost. Submit a request via email to orders@epri.com that includes your full name, mailing address, phone number, organizational affiliation, and country of citizenship."

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.

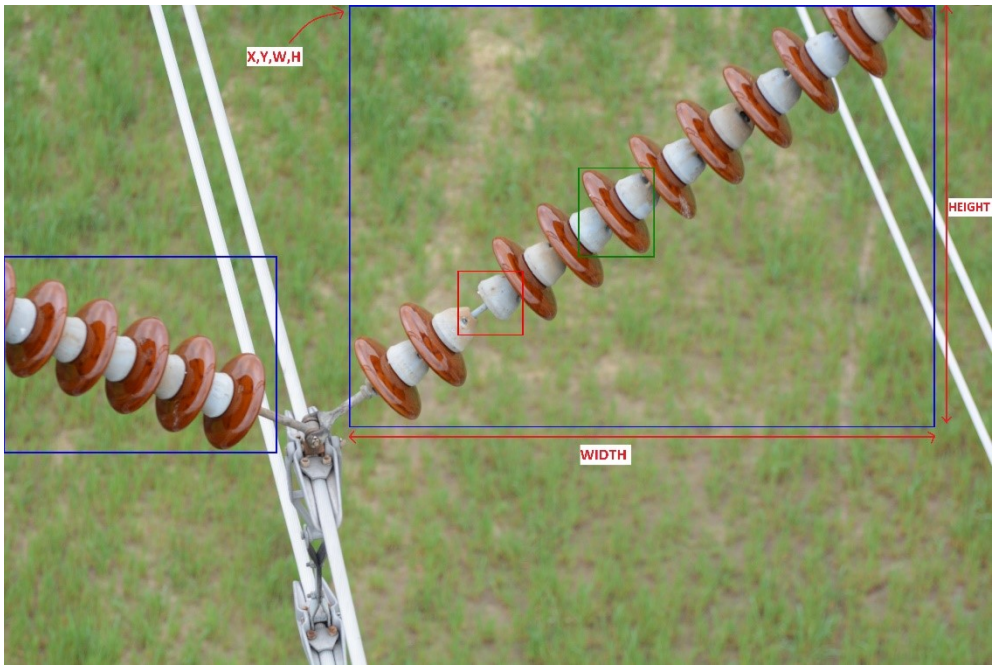

Figure 4-1
Damaged insulator string and its corresponding annotations ID: 150394.jpg

Export Control Restrictions

Access to and use of this EPRI product is granted with the specific understanding and requirement that responsibility for ensuring full compliance with all applicable U.S. and foreign export laws and regulations is being undertaken by you and your company. This includes an obligation to ensure that any individual receiving access hereunder who is not a U.S. citizen or U.S. permanent resident is permitted access under applicable U.S. and foreign export laws and regulations.

In the event you are uncertain whether you or your company may lawfully obtain access to this EPRI product, you acknowledge that it is your obligation to consult with your company's legal counsel to determine whether this access is lawful. Although EPRI may make available on a case by case basis an informal assessment of the applicable U.S. export classification for specific EPRI products, you and your company acknowledge that this assessment is solely for informational purposes and not for reliance purposes.

Your obligations regarding U.S. export control requirements apply during and after you and your company's engagement with EPRI. To be clear, the obligations continue after your retirement or other departure from your company, and include any knowledge retained after gaining access to EPRI products.

You and your company understand and acknowledge your obligations to make a prompt report to EPRI and the appropriate authorities regarding any access to or use of this EPRI product hereunder that may be in violation of applicable U.S. or foreign export laws or regulations.

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI also provides technology, policy and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI members represent 90% of the electricity generated and delivered in the United States with international participation extending to nearly 40 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; Dallas, Texas; Lenox, Mass.; and Washington, D.C.

Together…Shaping the Future of Electricity

3002017949

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com