Discussion on coding rules and labelling specifications of defect image samples of substation equipment

Sha Yang¹², Wei Xiong¹², Zhao Zhang¹², Liang Chen¹², Shuxin Huang¹²

¹Nari Group Corporation, State Grid Electric Power Research Institute, Nanjing 211106, Jiangsu, China
²Nari Technology Co. Ltd, Nanjing 211106, China
Emails: yangsha@sgepri.sgcc.com.cn

Abstract. In the image recognition of substation equipment defects based on deep learning, image sample coding rules and labeling specifications are one of the key factors to ensure the quality of sample data. At present, the image of the defect of the substation equipment is scarce, and the sample management specification is not uniform. There is a low level of image recognition for equipment defects in the substation intelligent inspection. Therefore, the paper carries out analysis and thinking on the coding rules and labeling specifications of the substation key equipment defect image samples. The sample making process is standardized through the standardization of the coding of typical substation equipment type, component type, defect type and annotation format, to some extent, which improves the accuracy of sample data quality and substation equipment defect recognition, and improves the level of defect management in intelligent inspections of substation. Finally, the results show that the standardized labeling samples help to improve the recognition accuracy and have wide application value.

1. Introduction
In 2018, at the symposium on artificial intelligence application of China Electric Power Research Institute, experts proposed that artificial intelligence had been applied in various fields and had risen to a national strategy. Substation equipment inspection work is moving towards the direction of intelligence, information and modernization. At present, in the substation intelligent operation and maintenance work, high-precision, multi-level and all-round image recognition based on Deep Learning Technology (DL) for substation equipment and specific parts is a research hotspot, which can provide data support for equipment status analysis and diagnosis. In the actual production process, the identification objects of substation inspection robot mainly include meter reading, switching breaker gate separation state, equipment temperature, equipment appearance and operating environment, etc. Among them, there are enough samples of common categories such as meter reading, on-off switch state and equipment temperature, and the recognition accuracy has reached a high level. The defect management of equipment is an important part of the operation management of substation equipment, but the accuracy of equipment defect recognition has been at a very low level. The reason is that there are many types of equipment in the field of substation, and the types of defects are various, and the required large number of scene images have a small probability and occasional occurrence in model training, and the abnormal expressions and characteristics are also various, such as oil leakage on a device, foreign object suspension, smoke, fire, etc., sample collection is difficult and technical research is very difficult.
Deep learning model training needs to process a large number of high-quality training samples to extract high-level features. The sample number and sample quality are the key factors affecting the recognition accuracy. Due to the particularity of the power industry, there are no public data sets available for power equipment. Currently, substation equipment defect image is scarcity, construction disorder, lack of standard specifications, uneven quality, etc. The formulation of image sample coding rules and labeling specifications can standardize the management of samples, which is a particularly critical step to ensure the quality of sample data.

The important significance is also reflected in several aspects. Such as (1) the inheritance of technology; (2) the improvement of efficiency; (3) the fundamental method to prevent the recurrence of problems and avoid disputes when dealing with problems, avoid or reduce problems; 4) With standardization, the operation has a basis, and the improvement has a basis.

Therefore, this paper carries out research and thinking on coding rules and labeling specifications for image samples of substation key equipment defects. Through the standardization of the typical substation equipment type, component type, defect type coding and labeling format, the sample production process has a standard specification. The quality of samples is improved, and better recognition effect is achieved in the field of substation equipment defect recognition.

2. Coding and grading of defect images of substation equipment

In order to improve the management level of substation equipment defects, formulate uniform specifications for substation equipment information and typical defects for substation, referring to classification relationship of equipment types, components, components types and parts defined in the Standard specification “substation primary equipment defects classification standards”.

2.1. The coding of equipment and components

The detection target must have a unique identifier, and both the equipment and the equipment’s components need to be uniformly named. A standardized coding rule for the sample database of substation equipment is defined, according to the classification of substation equipment components, component types and parts in the requirements of 7Q/GDW 1906.

The equipment coding rule can be described as “the initials of the equipment’s Chinese name”. Each equipment may be composed of multiple components, and the equipment coding rule can be described as “the initials of the equipment’s component’s Chinese name”. For example, the equipment of transformer can be described as “BYQ”. A transformer is made up of many parts, for example, the Body can be described “BT”, the Casing can be described “TG”, the Cooling System can be described as “LQXT”, the Tap Changer can be described as “FJKG”, the Non-Electricity Protection can be described as “FDLBH”, the Architecture and Foundation can be described as “GJJJC”.

The detailed classification is shown in the table below.

| Equipment Name | Equipment Coding | Component Name                  | Component Coding |
|----------------|------------------|---------------------------------|------------------|
| Transformer    | BYQ              | The Body                        | BT               |
|                |                  | The Casing                      | TG               |
|                |                  | The Cooling System              | LQXT             |
|                |                  | The Tap Changer                 | FJKG             |
|                |                  | The Non-Electricity Protection  | FDLBH            |
|                |                  | The Architecture and Foundation | GJJJC            |
| Reactor        | DKQ              | The Body                        | BT               |
|                |                  | The Casing                      | TG               |
|                |                  | The Cooling System              | LQXT             |
|                |                  | Tap Changer                     | FJKG             |
|                |                  | Non-electricity protection      | FDLBH            |
|                |                  | The Architecture and Foundation | GJJJC            |
| Breaker        | DLQ              | The Body                        | BT               |
2.2. Defect classification and defect coding

For the judgment criteria of equipment parts defects, refer to Q/GDW 1906 “Classification Standards for Substation of Substation”, which classify the defect level of substation equipment into three categories: critical, serious and general.

Critical Defects: Critical Defects are defects that have jeopardized the safe operation of substation equipment and may cause accidents at any time, both dangerous and urgent.

Severe Defect: Severe Defect refers to the fact that the defect has a serious threat to the safe operation of substation equipment. In the short term, the equipment can still maintain safe operation. Although the situation is dangerous, the urgency is less than the defect of the crisis.

General Defects: General defects refer to defects that pose less threat to the safe operation of substation equipment and do not affect the safe operation of the equipment for a certain period of time.

The color of the mark box of the output label image is marked with a red color according to the defect level, the critical defect of the orange mark, and the general defect of the yellow mark. From the color of the frame of the output label image, the defect level is critical, serious and general, so as to facilitate subsequent implementation of protective measures. Shown in table 2.

According to the defect grade, red color is used to mark critical defects, orange color is used to mark serious defects and yellow color is used to mark general defects. According to the selected color of the output labeled image, the defect grade is critical, serious and general, so as to facilitate the subsequent protection measures. Are shown in table 2.

| Crisis Name              | Label Color | Color description            |
|-------------------------|-------------|------------------------------|
| Critical Defects        | Red(R=0, G=255, B=0) | Red(R=0, G=255, B=0) |
| Severe Defects          | Orange(R=255, G=165, B=0) | Orange(R=255, G=165, B=0) |
| General Defects         | Yellow(R=255, G=255, B=0) | Yellow(R=255, G=255, B=0) |
Detailed classification of typical defects of substation equipment components, as is shown in Table 3:

| Defect description | Defect classification | Defect level | Defect coding |
|--------------------|-----------------------|--------------|---------------|
| Meter defect       | Dial broken           | Critical     | BJBPPS        |
|                    | Shell damaged         | Critical     | BJWKPS        |
|                    | Dial blur             | General      | BJBPMH        |
| Insulator defect   | Insulator damage      | severe       | JYZPS         |
|                    | Insulator crack       | severe       | JYZLW         |
| Respirator defect  | Silicone discoloration | severe     | GJBS          |
|                    | Oil seal damaged      | General      | YFPS          |
|                    | Oil seal missing      | General      | YFQS          |
| Foreign matter     | Hanging suspended matter | General | GKXFW        |
|                    | Nest/Hive             | General      | NC            |
| Leakage oil        | Component surface grease | severe | BBJMYW      |
|                    | Ground oil            | severe       | DMYW          |
| Metal corrosion    | Bolt corrosion        | General      | LSXS          |
|                    | Surface rust          | General      | BMXS          |
| Fireworks hazard   | Smoking               | Critical     | MY            |
|                    | Fire                  | Critical     | ZH            |
| Clip broken        | Clip Broken           | Critical     | XJPS          |

2.3. Coding rules and characteristics of sample name

2.3.1. The Coding rule of sample name. The name of the substation inspection image sample name should include substation name, equipment name, component’s name, defect name, image date and image time, and so on. The naming format of the sample image of the substation equipment is shown in Figure 1.

![Figure 1. The Coding rule of sample name](image)

The naming rules for the image sample can be described as “the initials of pinyin of the substation’s name” + “_” + “the initials of pinyin of the equipment’s name” + “_” + “the initials of the component’s Chinese name” + “_” + “the initials of the defect description’s Chinese name” + “_” + “the initials of the defect classification’s Chinese name” + “_” + “year, month, day, time” + “.jpg”.

If there are multiple defects on one image, the image sample name should include all the defects, and the devices, components, date and time information to which the defect belongs to. The defects are separated by a separator “-” to indicate one defect information. For example, It is AnShun-BYQ_TG_BJPS-BYQ_TG_JYZPS-201902011650.JPG.

2.3.2. The advantages and characteristics of the coding rule. (1) Through the sample name, you can know the substation name, equipment type, equipment component name, defect classification, defect description, date and time to capture, etc. The relationship the defect related to the equipment and parts is reflected in the sample name. The existence of one or more defects on one image can also be a good expression and traceability of affiliation. (2) It facilitates image retrieval of fine-grained information. The coding is used as the unique identifier of the sample, which can fully express the affiliation of the
defect with the equipment and the component. The intuitive and convenient features are very conducive to the sample management of large data volume in the database, also it is convenient for the re-search and development of image recognition technology, and also for the related application of debugging personnel.

2.4. Develop coding specification of sample labeling

2.4.1. The labeling area. Labeling needs to consider positive and negative samples, and an appropriate proportion of negative samples can help improve the recognition rate. Note the following points in the labeling process. (1) It is necessary to mark the visible area of the defect target, and the detection frame needs to include all the pixels of the defect object, and mark the entire area where the defect exists. For example, the type of insulator crack should mark at least the insulator region where the crack is located, and the label name is “jyz_lw”; the type of insulator rupture should mark all fracture regions and at least one adjacent insulator or metal body region. The label name is “jyz_pl”. (2) Mark the entire insulator region to which the defect belongs to. For example, the label name of insulator is "jyz", of meter is “bj”, of respirator is “hxq”. (3) Label the target or equipment without defects. For example, the label name of normal insulator is "jyz_zc", of normal meter is "bj_zc", of normal respirator is "hxq_zc".

2.4.2. The labelling name. The labeling name of image sample with defects should include information on the defect description and defect classification. The naming rules for the defect image sample labeling are shown in Figure 2.

![Figure 2. sample labeling specification](image)

When a defect classification is divided into multiple sub-categories, the general naming convention can be described as “the initials of its defect description of the Chinese name” + “the initials of its defect classification”. For example, the surface oil stain of the component of the leakage oil type can be expressed as “sly_bjbmyw”, and the nest of the foreign object type can be expressed as “yw_nc”. When there is only one type of defect, the labeling is named as: “the initials of the defect description”. For example, the type of clip breakage class can be described as “xjps”. For defects that cannot be determined, the label is named “Other”.

2.4.3. The shape and size of labeling area. It is necessary to determine the annotation shape. According to experience and practice, it is suggested to represent the output result and target range in the form of marking frame for the annotation image. It is more commonly used to mark the defect target with the minimum positive and external rectangular frame, which is conducive to the development of application research. As is shown in figure 3 (a) and (b) below.
Figure 3 is a schematic diagram of the shape of the label

The label box should include all the pixels of the object. The size of the label area is determined by the pixel size of the visible area of the defect target. According to experience, it is not less than \(28\text{px} \times 28\text{px}\).

2.5. **Output labeling requirements**

The defect level is characterized by the output color of marker box and the labeling, and the defect level is determined by the equipment, the component type, and the defect label.

The labeling file is an xml file conforming to the VOC standard, and the name of labeling file must be consistent with the original image file name.

The text format file should contain the substation name, device name, part name, defect type, defect description, defect level, image date, time, label name, marker box coordinates, marker box color, etc., and are separated by ";". When there are multiple output results in a single image, the equipment name, component name, defect type, defect description, defect level, image date, time, label name, marker frame coordinates, marker box color, etc., of each output result should be record in txt format file and separated with ";".

3. **Application analysis**

3.1. **Building the insulator data sets**

The experimental data set uses high-resolution insulators images from robot inspection and artificial shooting in the substations of Guizhou Anshun, Anhui Xuancheng, Henan Tuhui, Henan Huiji, etc., including various defect angles and various background insulator crack images, totaling 572 sheets. These images contain one or more strings of insulators, of which 273 images are positive samples and other 299 images are negative samples. In order to improve the generalization capability of the network model, the data set is divided into training sets and test sets by 4:6, and the positive and negative samples are each half in the training set and the test set.

Two kinds of labeling schemes:

1. Only label the parts of the defect, such as the insulator crack named “jyz_pl”.
2. Label the defect part, the component which the defect belongs to, and the component of the non-defective equipment. As shown in the figure, the defect part is the insulator crack named “jyz_pl”, and the part to which the defect belongs is insulator object named “jyz”, non-defective equipment parts are normal insulators named “jyz_zc”.

According to the requirements of different target and different scenes on the characteristics of the insulator defect samples, the insulator samples are labeled according to the two labeling schemes, and the training data sample library of image quality is constructed.

According to the requirements of different target and different scenes for the characteristics of the insulator defect samples, the insulator samples are marked according to the two labeling schemes. This
article uses LabelImg to label the insulator and insulator defect block locations, as is shown in Figure 4 (a) and (b).

![Figure 4. insulator sample images](image)

If the insulator is intact, mark only the position of the insulator in the image; if the insulator is defective, mark the location of the insulator defect and also mark the location of the insulator. The reason for this is that the defective block must be located on the insulator, so the insulator and the defective block are completely marked. The network can learn the correlation between the two during training, which will improve the robustness of the network. After the labeling is completed, a small insulator data set is created in accordance with the Labeling scheme 1 and Labeling scheme 2.

3.2. **Experiment analysis**

The training is carried out according to the process of input training samples, image pre-processing, algorithm construction, start training, output model weight, sample test, and output. The whole experiment process is shown in Figure 5.

![Figure 5. Experimental verification process](image)
In this experiment, YOLOV3 architecture with small target robustness is selected to detect the defect and position of the insulator. The two types of sample data sets are trained respectively according to the deep network learning model, and the trained model weights are obtained. And then some test sample images are brought into after testing the performance of the trained model, detecting the recognition effect of the sample image. The value of the mAP(Mean Average Precision) of the model is calculated through multiple times. The results are shown in Table 4.

| Labeling scheme | Mean Value of mAP | Insulator | Defect of Insulator |
|-----------------|-------------------|-----------|---------------------|
| Labeling scheme 1 | 80.0%              | 81.3      | 81.6                |
| Labeling scheme 2 | 83.0%              | 82.5      | 86.4                |

It can be concluded from the table that the mAP of the algorithm is increased by 3.0% compared with the original, in which the accuracy of the insulator is increased by 1.2%, and the accuracy of the defect is improved by 4.8%. Changes in the labeling method can better guide the network training.

4. Conclusion

Through the analysis and discussion of the coding rules and labeling specifications of typical equipment defect image samples of substation, the sample management is raised to the height of standardization construction. The sample preparation process has a standard specification and the quality of the sample can be greatly improved. It facilitates the rapid formation of efficient and high-quality sample libraries, which is conducive to the improvement of the recognition level of the defect samples image of the substation equipment. It can add strong technical support for informatization and intelligent construction of transportation inspection.

Acknowledgement

Research and Application of Transmission Line Inspection Robots” Science and technology project of NARI Technology Co. Ltd; "Research and Application of key technology of automatic inspection" science and technology project of The State Grid.

References

[1] Wang Liuwang, Zhu Yongli, Li Li, et al. Partial discharge type recognition of small sample for power transformer based on feature subsets[J]. Electrical Measurement & Instrumentation, 2015, 52 (24): 40-45.
[2] Meng Gege, Gao Qiang. A substation abnormal scene recognition algorithm based on multi-layer convolution neural network[J]. 2018, 55(5): 46-50.
[3] Yu LanYing, Yao Bo, Wu WenHai, et al. Insulator Identification Method Based on Multi-feature. Insulators and Surge Arresters. 2016, 3: 79-83.
[4] Wu Tao, Wang WeiBin, Yu Li, et al. Insulator defect detection method for lightweight YOLOV3[J]. Computer Engineering. 2019.
[5] YAO Nan, CAI Yue, MIAO Jia. Research on substation equipment insulator crack identification based on intelligent images[J]. Modern Electronics Technique. 2017, 40(22): 176-178.
[6] Chen XiaoLong. Research and application of intelligent identification and recognition technology for substation bird’s nest based on image recognition technology[J]. Electronic Testing, 2018, 16: 48-51.