# Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks

Xian Tao, Dapeng Zhang, Zihao Wang, Xilong Liu, Hongyan Zhang, and De Xu, *Senior Member, IEEE*

*Abstract*—As the failure of power line insulators leads to the failure of power transmission systems, an insulator inspection system based on an aerial platform is widely used. Insulator defect detection is performed against complex backgrounds in aerial images, presenting an interesting but challenging problem. Traditional methods, based on handcrafted features or shallow-learning techniques, can only localize insulators and detect faults under specific detection conditions, such as when sufficient prior knowledge is available, with low background interference, at certain object scales, or under specific illumination conditions. This paper discusses the automatic detection of insulator defects using aerial images, accurately localizing insulator defects appearing in input images captured from real inspection environments. We propose a novel deep convolutional neural network (CNN) cascading architecture for performing localization and detecting defects in insulators. The cascading network uses a CNN based on a region proposal network to transform defect inspection into a two-level object detection problem. To address the scarcity of defect images in a real inspection environment, a data augmentation method is also proposed that includes four operations: 1) affine transformation; 2) insulator segmentation and background fusion; 3) Gaussian blur; and 4) brightness transformation. Defect detection precision and recall of the proposed method are 0.91 and 0.96 using a standard insulator dataset, and insulator defects under various conditions can be successfully detected. Experimental results demonstrate that this method meets the robustness and accuracy requirements for insulator defect detection.

*Index Terms*—Aerial image, convolutional neural network (CNN), data augmentation, defect detection, insulators.

## NOMENCLATURE

| | |
|---|---|
| CNN | Convolutional neural networks. |
| UAV | Unmanned aerial vehicles. |

| | |
|---|---|
| OAD-BSPK | Orientation angle detection and binary shape prior knowledge. |
| BRISK | Binary robust invariant scalable keypoints. |
| VLAD | Vector of locally aggregated descriptors. |
| SVM | Support vector machine. |
| WMRA | Wavelet multiresolution analysis. |
| CPLID | Chinese power line insulator dataset. |
| RPN | Region proposal network. |
| ReLU | Rectified linear units. |
| ILN | Insulator localizer network. |
| DDN | Defect detector network. |
| ROI | Region of interest. |
| VOC2007 | Visual object classes challenge 2007. |
| DOST | Discrete orthogonal s-transform. |
| IoU | Intersection over union. |
| ACF | Aggregated channel features. |
| SSD | Single shot multibox detector. |
| YOLO | You only look once. |
| GAN | Generative adversarial nets. |

## I. INTRODUCTION

INSULATORS are essential equipment used to electrically isolate and mechanically secure wires in high-voltage power transmission systems [1], [2]. Insulator failure is a direct threat to the stability and safety of power transmission lines. Statistically, accidents caused by insulator defects account for the highest proportion of power system failures. Therefore, intelligent and timely detection of insulator defects is particularly important.

In the last decade, many studies have investigated automating the task of visually inspecting power transmission systems [3], [4]. As aerial platforms such as helicopters and UAVs have become available, and demonstrate high efficiency, accuracy, and safety, they have become an important tool in the inspection of power equipment (see Fig. 1) in recent years. A camera mounted on the platform is able to obtain many aerial images that include insulator information. At present, insulator defect inspection is performed manually based on such aerial images, and as a result, highly efficient and accurate detection is difficult. To overcome the limitations of human inspection, automated defect detection techniques [5]–[7] need to be developed to assist or replace human decisions.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2                                                                                                    IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS



Fig. 1.   Two main types of aerial inspection platforms: (a) helicopters and (b) UAVs.

Images captured by aerial inspection platforms often include cluttered backgrounds resulting from the presence of towers, mountains, rivers, grasslands, and farmland. In an actual detection environment, aerial images capture different views of changing environmental and illumination conditions. Therefore, processing these images is complicated, which can easily lead to false fault detections. Insulator inspection is usually divided into two steps: 1) insulator localization and 2) defect recognition. Locating an insulator in an image requires overcoming several unfavorable factors. Second, the conditions of the insulators must be diagnosed while determining their specific defect locations in order to facilitate maintenance. Although many types of defects exist, this paper focuses on the most common defect of outdoor insulators: missing caps.

Several studies have focused on insulator localization, such as [8], that use OAD-BSPK to locate multiple insulators at different angles in complex aerial images. However, the method is restrictive since one must have prior knowledge of the insulator shape, and it can only function under certain scenarios, e.g., with untextured backgrounds or with a camera facing the sky. Thus, it does not work well for the highly complex backgrounds used in this paper. Li *et al.* [9] used the Otsu thresholding algorithm to detect insulators, but their approach is only capable of detecting insulators via a strong contrast, limiting its applicability considerably. In contrast to such methods that are not robust, Liao and An [10] proposed an insulator detection algorithm based on local features and spatial orders in aerial images. It is specifically designed for localization, which is not necessary and not possible in defect detection as the proportion of insulators in aerial images is usually too small. Zhao *et al.* [11] proposed a method for insulator string detection, and presented an apparent representation for insulator strings in infrared images based on the BRISK and VLAD methods. The method uses proposed features to construct a training classifier. The localization of the insulator is achieved by embedding the classifier into a multiscale sliding detection frame. The infrared images are usually of low resolution and with less complex backgrounds compared to visible images. Further, the features identified via shallow learning techniques are not sufficiently discriminative for complex aerial images. Oberweger *et al.* [12] proposed an insulator detection algorithm based on discriminative training of local gradient-based descriptors and a subsequent voting scheme for localization. This required the insulator caps to be clear of obstruction and a higher angle to clearly view the insulator.

A completely different method was proposed by Wu *et al.* [13] who used an active contour model to segment aerial insulator images with texture inhomogeneity, which is more invariant to clutter. This method suffers from high computational costs and its models cannot be initialized automatically.

To detect insulator defects, [9] and [12] proposed methods for finding missing caps from aerial images. These methods split an insulator into several parts using profile projection, but the partitioning was not suitable for all sizes of insulators. Murthy *et al.* [14] proposed a condition analysis system for overhead insulators using combined SVM and WMRA. Though the SVM classifier could provide the defect category information, the defect location was not obtained. An alternative method proposed monitoring the condition of the insulators [15], using *k*-means clustering to locate insulators within bounding boxes. The DOST and SVM were used to estimate the condition of the insulator with a DOST. This method needs insulators that are nearly the same size, and uses insulator feature classification to determine its condition, but also does not provide a specific defect location. Zhao *et al.* [16] developed a multipatch CNN feature extraction method for insulator status inspection. The method divides an insulator by histogram valley detection, and a deep CNN is used for feature extraction of the insulator caps. The method does not consider differently sized or partially visible insulators which would be hard to divide into caps.

The insulator defect detection methods discussed above usually locate the object first, and then extract its features with a classifier to determine its status. Location determination methods include using shape information [8], a sliding window [11], color information [12], clustering [15], and histogram projection [16]. Such methods convert the object detection task into a classification problem. However, they still suffer from the drawback of using limited, handcrafted, or shallow-learning-based features and cannot provide the specific location of the defects. Moreover, these methods typically rely on several simplifying assumptions, particularly the insulator size, shape, and background. In real applications, however, insulators are extremely diverse in shape, appearance, and size, and their backgrounds vary widely, including towers, mountains, rivers, grasslands, and farmland. Fig. 2 shows four background types with insulators in the foreground. The size of these insulators is not the same. Until now, no studies that can be used as an evaluation benchmark have been published, and most existing studies only offer a practical demonstration of a method. Therefore, our goal is to provide a well-established benchmark for insulator defect detection.

This paper aims to develop more general approaches for detecting insulator faults. A complete solution is proposed for combining insulator localization and defect detection by using cascading CNNs. An RPN approach is used to first locate an insulator in a given image. Once the insulator has been located, a DDN identifies the insulator and any defects. It can give the specific location of the defect, not just relying on classification to determine whether defects exist. The network is trained end-to-end with a large dataset consisting of training data, including synthetic images created using data augmentation.

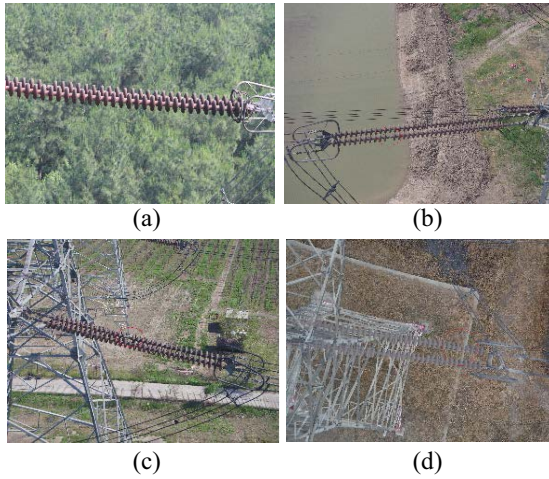The main contributions of this paper are as follows.

Fig. 2. Four background types of insulator. (a) Forest background. (b) River background. (c) Land background. (d) Tower background.

1) While many existing studies focus on insulator localization and defect detection separately, we propose a new cascading CNN which combines these two insulator inspection tasks. To the best of our knowledge, we are the first to consider the insulator defect inspection problem as a two-level object detection problem based on CNN.
2) A data acquisition method is proposed that obtains many defect samples, with some samples being synthetic. We introduce a methodology to enhance the volume and diversity of defect insulator datasets, and exploit the enhanced datasets to train a deep CNN.
3) Successful defect detection using the proposed method is demonstrated using a real-world dataset. We construct a CPLID and make it available online for public use.[1]

The remainder of this paper is organized as follows. Section II introduces the system framework. The proposed cascading CNN architecture is illustrated in Section III. In Section IV, we explain our data acquisition methods in detail. In Section V, we describe the experiments conducted to evaluate the method. Finally, the conclusions are presented in Section VI.

## II. SYSTEM FRAMEWORK

An automatic insulator inspection system consists of three parts: camera, which can be installed on UAV, the 4G communication network, and the ground monitoring center, where the detection algorithm is embedded into the ground server, as shown in Fig. 3. The camera has an autofocus function that adjusts the size of the insulator in the field of view. The images of the insulator string are first captured, and sent to the ground monitoring center by 4G communication network. The insulator defects can be recognized and located from the images in the monitoring center. The client's interface displays the status of the current insulators in real time.
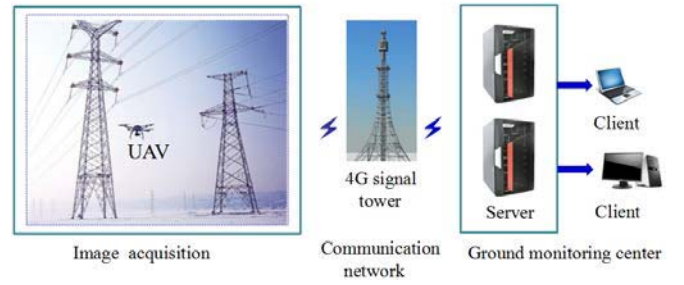
Fig. 3. Whole system framework.

## III. PROPOSED DETECTION METHOD

The method proposed here is designed to automatically detect missing insulator caps in complex aerial images. A novel detection architecture for both localizing insulators and detecting their defects is presented that leverages CNN. The proposed model transforms defect detection to a two-level detection problem, directly uses object detection to locate insulator position, and then uses the object detection method again to locate the defect position on the insulator. This makes full use of the strong representation and regression performance of CNN. Fig. 4 shows the entire structure of our model.

### A. Architecture of CNN

CNNs are a type of feedforward neural network widely applied in image-related analysis, and typically consist of convolutional, pooling, and fully connected layers. The convolutional layers apply a number of filters on local regions of the input, obtaining feature maps of the input image. Suppose the number of filters is $k$, $W^i$ denotes the weight of filter $i$, $b$ is the bias of filter $i$, $x_s$ stands for a small image patch, and $\sigma$ is the activation function. The convolution of $x_s$ given filter $i$ is

$$f_{i,s} = \sigma\left(W^i x_s + b^i\right) \qquad (1)$$

where $x_s$ is a patch window that slides across the image with stride size $s$. Through the convolutional layers, it is possible to obtain the feature maps for representing the image information. The pooling layers perform down-sampling along spatial dimensions, and are normally applied after the convolutional layers to reduce feature dimensionality in order to avoid overfitting. The fully connected layers normally constitute the last few layers of a CNN. The structure of the fully connected layers is the same as that of a traditional neural network. Different image-related tasks can be implemented by defining different loss functions behind the fully connected layer, such as pixel classification and pixel coordinate position regression. Various deep CNNs normally consist of alternating convolutional and pooling layers, followed by fully connected layers.

### B. Cascading Architecture

The proposed cascading model includes two networks. The first network ILN detects all the insulators in the images. Insulators in the form of rectangular boxes are cropped from the original image, while other parts are discarded. The second

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4

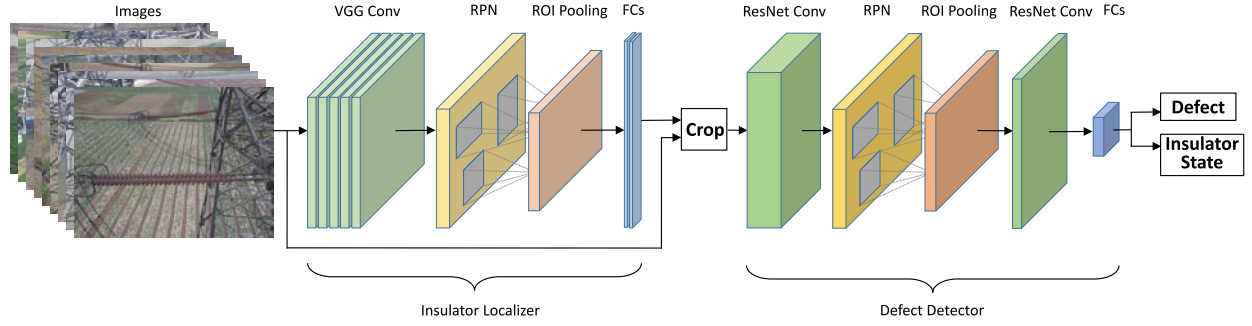IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS



Fig. 4.    Cascading insulator and defect detection architecture.

network DDN detects all the missing caps from the cropped images. We introduced RPN [17] to our model to compute proposals effectively. RPN is designed to quickly produce a collection of detection boxes that may contain objects in an image. In the first network, this detection box may contain the region of the insulator, and in the second network such a box may contain the area of the missing cap. The RPN is inserted after the last convolutional layer. A small window slides across the convolutional feature map in order to classify objects and move the regressing bounding box. In the RPN, three scales and three aspect ratios are used, resulting in nine different object proposals. It is more efficient to obtain the detection box directly on the feature map than to determine it from an input image. To train the RPN, the loss function for an aerial image can be defined as

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \tag{2}$$

where $i$ is the object proposal index, and $p_i$ represents the ground truth label. The label $p_i^*$ is the $i$th object's proposal, and if the proposal is for an object, $p_i^* = 1$, but otherwise is $p_i^* = 0$. $t_i$ is a four-dimensional vector that represents the coordinates of the detection box, and $t_i^*$ is the labeled detection box. $L_{\text{cls}}$ and $L_{\text{reg}}$ are the logarithmic loss and regression loss functions, respectively, and $N_{\text{cls}}$ and $N_{\text{reg}}$ are the number of classes and detected boxes, and with $\lambda$, these three parameters are used to normalize the two loss functions.

*ILN:* In the front-end of the model, we use the convolution layers of VGG-16 [18], which consists of a stack of $3 \times 3$ convolution and $2 \times 2$ max-pooling layers, to generate the feature maps. These feature maps are fed into the RPN. In the RPN, each $3 \times 3$ area of the feature maps becomes a feature with 512 dimensions by applying a convolution layer and ReLU. The 512-dimensional features are fed into two fully connected layers to perform box classification and regression. Based on the box information, another pooling is applied to the feature map produced from the VGG-16 convolution layers, called ROI pooling. From this step, the features extracted in the VGG-16 convolution layers can be shared. In the end, the pooling result is equipped with two fully connected layers for box classification and more precise box border regression. In the inference stage of the ILN, a number

of image patches are cropped from the original images. The sizes of these cropped images are expected to be the smallest rectangular area, which contains one or more complete insulators.

*Crop Module:* The output of the ILN is a confidence score, along with the left-top and bottom-right corner vertex coordinates of the boxes containing insulators. The insulators are cropped out of the input images at these coordinates. Then, these cropped images are sent to the DDN to find the missing cap defects. The ILN is powerful in that a tiny insulator may be found in the image. However, these tiny insulators may be so small that determination of any abnormality may be difficult. Hence, before sending to the DDN, the size of cropped images must be determined to be sufficiently large. To this end, the equivalent edge length ratio $R$ is defined as

$$R = \sqrt{\frac{H \times W}{(x_2 - x_1)(y_2 - y_1)}} \tag{3}$$

where $H$ and $W$ are the height and width of the input image, $[x_1, y_1]$ and $[x_2, y_2]$ are the coordinates of the bottom-right and top-left vertices of the box that is output from the insulator localizer. Then, we define a threshold $R_t$, that is used to decide whether to crop the insulator. This procedure accelerates the following defect detection process.

*DDN:* The original images are cropped according to the detected box in the ILN as the inputs of the DDN. The network at this stage is aware of its previous iteration. In this stage, the defects are smaller relative to the insulator patches, allowing a more accurate network to be designed in the DDN. ResNet-101 [19] convolution layers are used instead of those from VGG-16 because of the former's more numerous available network layers. In contrast to the ILN, which is based on VGG-16 and only has fully connected layers followed by RPN and ROI Pooling layers, the DDN inserts its RPN and ROI Pooling layers between two sets of convolution layers. This enables improved detection accuracy at a cost of additional computing time. For this reason, ResNet was chosen as the front end for the DDN. The output of the DDN must not only to provide the defect location, but also mark the corresponding defective insulator. Thus, two outputs are generated: a defect box (with confidence) to locate the defect, and an insulator annotated with its status.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TAO *et al.*: DETECTION OF POWER LINE INSULATOR DEFECTS USING AERIAL IMAGES ANALYZED WITH CNNs 5
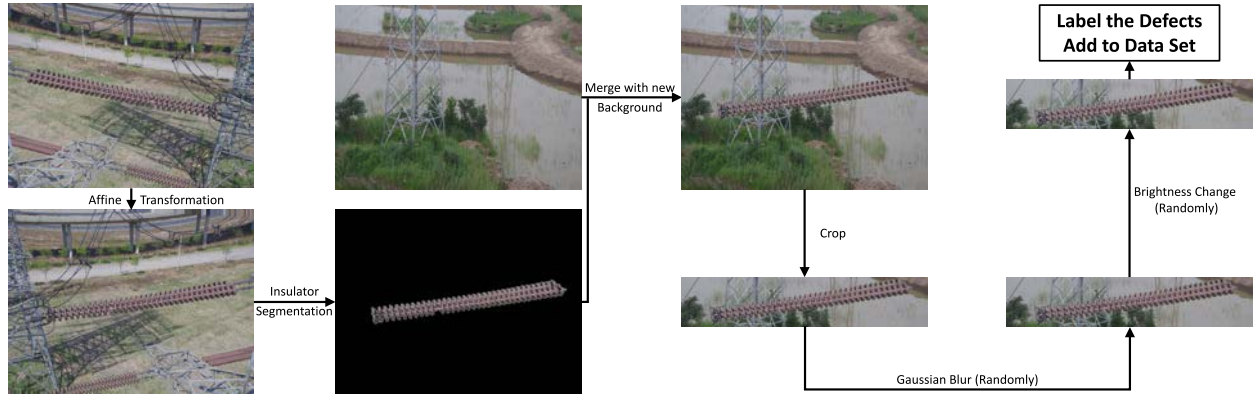


Fig. 5. Data augmentation process. An affine transformation is applied to the image initially, and then the insulator is separated from the original background, and merged with the new background. Then, the region containing the insulator is cropped from the artificial image. Blur and brightness changes may be added to the image before labeling the defects and adding to the dataset.

## IV. DATA AUGMENTATION

The State Grid Corporation of China provided aerial images of insulators. A common problem in defect detection is that defect datasets are normally too small to train CNN. Owing to the limited size of the training samples, augmenting the data was necessary in order to avoid overfitting. Data augmentation generates new datasets using existing data. A pipeline demonstrating how augmented samples are generated is shown in Fig. 5, and includes four operations: 1) affine transformation; 2) insulator segmentation and background fusion; 3) Gaussian blur; and 4) brightness transformation. Each step simulates different imaging environments, including changes in perspective, background, blur, and lighting.

### A. Affine Transformation

A color aerial image can be described by a function defined on a two-dimensional finite region. Mathematically, we have

$$I : \Omega \rightarrow [r, g, b] \in \{0, \ldots, 255\}^3$$
$$[x, y]^T \rightarrow I(x, y) \tag{4}$$

where

$$\Omega = \left\{ [x, y]^T : 0 \leq x < W, 0 \leq y < H \right\} \tag{5}$$

is the image region, $W$ is the width, and $H$ is the height of image $I$. We use $p = [x, y]^T$ to denote the image point or pixel at location $(x, y)$ and $I(x, y) = I(p) = [r, g, b]_p$ to denote the intensity of the color image at point $p$. For convenience, homogeneous coordinates $\bar{p} = [x, y, 1]^T$ describe point $p$ in the image. An affine transformation for an image can be defined as

$$\bar{p}' = A\bar{p} \tag{6}$$

where

$$A = \begin{bmatrix} M & t \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & t_1 \\ m_{21} & m_{22} & t_2 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

Here, $\bar{p}'$ are the homogeneous coordinates of the point after transformation and $\bar{p}' = [x', y', 1]^T$. $M$ controls the scaling, shear, rotation, and reflection transforms of the image, and $t$ controls the translation transform.

Transformations applied to the images include scaling, shearing, rotating, flipping, and translating, replicating how a UAV might capture photographs of the same insulator from different angles. With the affine transformation, we can change the size, direction, and position of insulators in the original image. These newly generated images can be used as training samples. In this paper, the affine transformation is applied randomly within a certain range. After the transform, points lying outside the boundaries of the input are filled with the values of their nearest points.

### B. Insulators Segmentation and Background Fusion

An aerial image containing insulators can be divided into two segments by human vision: 1) the foreground (insulators) and 2) the background (other structure). However, the limitation of an affine transformation is that it cannot change the relative position of insulators and the background. To solve this problem, we generate samples by adding insulators to background images.

Background images are collected from real aerial scenes without any insulators. This portion of the algorithm segments the insulator contours from the complex aerial images, and then fuses the insulator back into the background images without insulators.

*1) Insulator Segmentation:* Since the background of an aerial image is complex, an insulator cannot be directly segmented using prior knowledge of colors or gradients. The TV-based algorithm [20] is an image segmentation method that is based on an energy functional and a total variation model, and is an interactive method, as the foreground and background must be marked manually, as shown in Fig. 6. When the TV-based method is used to segment the insulator image, the final segmentation result is closely related to the detail level of the mark. As more marks are applied, better results can be obtained. Thus, significant effort is spent in marking each image.

The fully convolution network U-Net [21] is applied for automatically segmenting the insulators. The U-Net is

---

**Algorithm 1** Insulator Segmentation

**Input:** Image set $\{\mathcal{I}_{org}\}$ that each image contains the defective insulator.

**Output:** Image set $\{\mathcal{I}_{ins}\}$ that each image contains only the defective insulator with black background.

1: Divide the set $\{\mathcal{I}_{org}\}$ into two subsets $\{\mathcal{I}\}_s$ and $\{\mathcal{I}\}_r$, ensure $\{\mathcal{I}_{org}\} = \{\mathcal{I}\}_s \cup \{\mathcal{I}\}_r$ and $|\{\mathcal{I}\}_s| \ll |\{\mathcal{I}\}_r|$

2: **for** Image $\mathcal{I} \in \{\mathcal{I}\}_s$ **do**

3:    Use white color to mark the insulator in $\mathcal{I}$;

4:    Use black color to mark background in $\mathcal{I}$;

5:    Generate constrain mask image $\mathcal{I}_c$ by the marks.

6:    **while** $\mathcal{I}_c$ not converge **do**

7:      Calculate dual variable $\boldsymbol{p}$ via $\nabla\mathcal{I}$ and $\mathcal{I}_c$.

8:      Update constrain image $\mathcal{I}_c$ by $\boldsymbol{p}$;

9:    **end while**

10:   Make $\mathcal{I}_c$ to binary image $\mathcal{I}_b$.

11:   Use $\mathcal{I}_b$, as mask image: $\mathcal{I}_t = \mathcal{I}\&\&\mathcal{I}_b$;

12:   Add $\mathcal{I}_t$ to $\{\mathcal{I}\}_{train}$;

13: **end for**

14: Use $\{\mathcal{I}\}_{train}$ to training u-net.

15: Use u-net to segment rest of image $\{\mathcal{I}\}_r$, get $\{\mathcal{I}_{ins}\}$.

16: **return** $\{\mathcal{I}_{ins}\}$

---



Fig. 6. For the interactive segmentation algorithm, the foreground and background were marked with white and black.

a convolutional neural network for fast and precise image segmentation. Initially, a few images are marked manually using the TV-based algorithm to obtain the preliminary segmentation. Then, these images and their segmented results are inserted into the U-Net for training. Finally, the U-net segments the remaining images, obtained by affine transform. The whole process is described in Algorithm 1. The use of the U-Net saves significant time during pixel annotation while improving segmentation. As shown in Fig. 7, for the same image, using the trained U-Net for segmentation performs better than the TV-based method. The U-Net segmentation result is also more complete, although the image inserted into the U-Net for training was generated by the TV-based algorithm.

*2) Background Fusion:* Using segmentation, many nonbackground insulator patches were obtained. The new images $I_{new}$ are generated by merging patches and backgrounds as follows:

$$I_{new}(p) = \begin{cases} \alpha I_{fg}(p) + (1-\alpha)I_{bg}(p) & \text{if } I_{bg}(p) \neq 0 \\ I_{bg}(p) & \text{otherwise.} \end{cases} \quad (8)$$

$I_{fg}$ and $I_{bg}$ indicate the insulator patch and the background image, and $\alpha$ are the coefficients of image fusion.

### C. Gaussian Blur and Brightness Transformation

In order to change the ambiguity and brightness of the image for simulating real imaging environments, blur and brightness variance were added to a portion of the image.

*1) Gaussian Blur:* The images were filtered using an $s \times s$ Gaussian kernel with standard derivation $\delta$. The images were convoluted with different Gaussian kernels, and the degree of blur of the images was variable. Since aerial images are captured by drones, this filtering simulates images taken by a camera with imprecise focus.

*2) Brightness Variance:* Similarly, photographs taken in real-world scenarios have changing illumination, and hence brightness variance must be added to a portion of the images. We first map the image from RGB space to HSL space, and then the image brightness is adjusted by changing the L component in HSL space.

## V. EXPERIMENTS

In this section, we evaluate our method for insulator defect detection using our standard insulator dataset. First, the dataset, experimental configuration, the construction of training and test sets, defect detection criteria, and implementation details are provided. Then, detection accuracy and speed of the proposed method are evaluated and compared by the experiments. Second, we compared results of defect detection with other competing methods. Besides, the advantage of the methodology of data augmentation is discussed. Finally, parts of the detection results of the CPLID are presented by images.

### A. Experiment Description

*1) Dataset:* The experiment dataset is 1956 high resolution images which include 900 positive samples and 1056 negative samples (as shown in Fig. 8). Positive samples are defect-free insulator images, which captured by an UAV with a DJI M200 camera at a resolution of $4608 \times 3456$ pixels and stored in BMP format. All positive samples contain one or more defect-free insulators. Negative samples are defective insulator images which consist 60 UAV captured images and 996 synthetic samples by data augmentation. These images contain one or more defective insulators.

*2) Experimental Configuration:* All the experiments were conducted based on the deep learning framework MXNet [22], and executed on a computer with an Intel Core i7-8700k CPU, NVIDIA GTX-1080 with GPU memory of 8 GB, and 32 GB RAM. Our implementation uses Ubuntu 16.04 operating system.

*3) Construction of Training Sets and Test Sets:* One thousand one hundred and eighty-six images in the experiment dataset for training and 770 images for testing were carried out in the ILN for insulator location. For defect detection (location), since there are no defects in positive samples, the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TAO *et al.*: DETECTION OF POWER LINE INSULATOR DEFECTS USING AERIAL IMAGES ANALYZED WITH CNNs                7
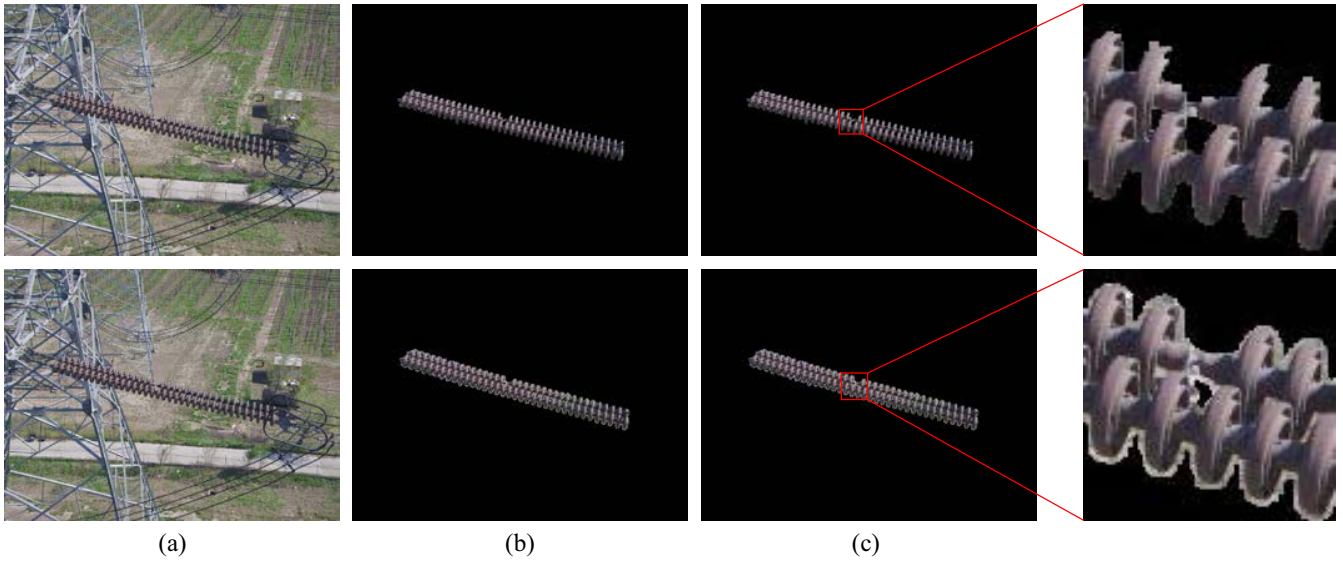


Fig. 7.   Image segmentation comparison. The top row used the TV-based method, and the bottom row used the U-Net. (a) Original image to be segmented. (b) Insulator segmentation results. (c) Detail of the segmentation result. (c) Edge of the insulator is not complete when using the TV-based algorithm.
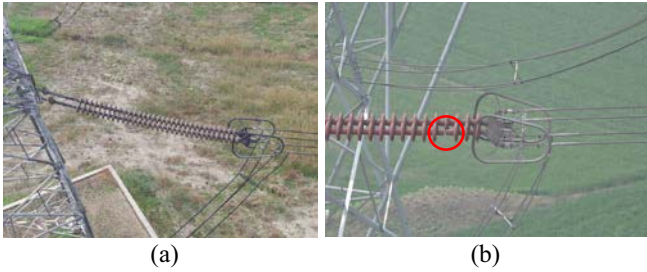


(a)                                (b)

Fig. 8.     Examples of training samples. (a) Positive sample. (b) Negative sample.



Fig. 9.   Examples of defect box.

Fig. 9, the defect boxes start with the adjacent cap on the left of the missing cap and end with the neighboring cap on the right of the missing cap. The defective image of this labeled mode is very similar to the dumbbell structure. In a complex aerial image, it is more representative than the purely labeled missing area.

*4) Defect Detection Criteria (Evaluation Criteria):* We adopted three widely used measures to quantitatively evaluate the performance of our defect detection method: precision ($P_r$), recall ($R_e$), and $F_1$. A higher evaluation value indicates better performance and vice versa

$$P_r = \frac{T_P}{T_P + F_P} \tag{9}$$

$$R_e = \frac{T_P}{T_P + F_N} \tag{10}$$

$$F_1 = \frac{2 * P_r * R_e}{P_r + R_e} \tag{11}$$

where $T_P$ (True Positive) and $F_P$ (False Positive) denote the number of correctly and incorrectly located defects, respectively. $T_P + F_P$ is the total number of located defects, and $T_P + F_N$ is the total number of actual defects. $F_1$ is the harmonic mean of precision and recall.

The region of the missing cap is small compared with the entire aerial image, so we calculate the above criteria under the IoU threshold. The IoU is defined as

$$IoU(G_T, P_M) = \frac{\text{Area}(G_T \cap P_M)}{\text{Area}(G_T \cup P_M)} \tag{12}$$

where $G_T$ is the ground truth and $P_M$ is the located box, respectively. For each located box, its IoU scores to all ground

TABLE I
PARAMETER SETTING OF DATA AUGMENTATION

| Parameter name | Value |
|---|---|
| rotation range | 0.2° |
| width and height shifts range | 5% |
| shear range | 0.05 rad( counter-clockwise direction) |
| scale range | 5% (horizontal flip) |
| coefficients of image fusion | 0.95 |
| Gaussian kernel size | 3 |
| standard derivation | 0.8409 |

cropped images of all negative samples were used for training and testing the DDN. Seven hundred and ninety-two images for training and 264 images for testing were carried out in the DDN for defect detection (location). Most of the negative sample images are obtained through the method of data augmentation in Section IV. The parameter settings of data augmentation are shown in Table I.

For the ground truth in training sets and test sets, detection regions (insulator strings and defects) were labeled with "insulator" and "defect" by the boxes, respectively, which are determined by the coordinates of the left-top corner and right-bottom corner. These labeled results are in the VOC2007 format [23]. The insulator boxes are the rectangular areas, which contain the complete insulator areas. As shown in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS

TABLE II
TRAINING PARAMETER SETTING OF ILN

| Parameter name | Value |
|---|---|
| Batch size | 2 |
| Epoch | 20 |
| IoU threshold | 0.8 |
| Initial learning rate | 0.001 |
| Momentum value | 0.9 |

TABLE III
TRAINING PARAMETER SETTING OF DDN

| Parameter name | Value |
|---|---|
| Batch size | 1 |
| Epoch | 25 |
| IoU threshold | 0.5 |
| Initial learning rate | 0.001 |
| Momentum value | 0.9 |



Fig. 10. Location results of aerial images. (a) Defect-free insulator strings. (b) Defective insulator strings.

TABLE IV
PERFORMANCE OF ILN AND DDN

| Network | Pretrained Model | Precision Rate | Recall Rate | $F_1$ Rate | Testing Speed |
|---|---|---|---|---|---|
| ILN | VGG-16 | 0.904 | 0.966 | 0.934 | 115 ms |
| | ResNet-50 | 0.897 | 0.957 | 0.926 | 165 ms |
| | ResNet-101 | 0.899 | 0.92 | 0.909 | 180 ms |
| DDN | VGG-16 | 0.415 | 0.629 | 0.5 | 80 ms |
| | ResNet-50 | 0.849 | 0.928 | 0.887 | 138 ms |
| | ResNet-101 | 0.91 | 0.958 | 0.933 | 149 ms |

truths are calculated, and the score and corresponding class are recorded. A located box with an IoU score larger than the threshold is assigned as the corresponding class (defect).

*Implementation Details:* The goal of our cascading network is to detect missing caps. The first network ILN detects all of the potential defective insulators in the images. The rectangular insulator regions for defect analysis are obtained using the crop module, while other small insulator regions are discarded. Then, the missing caps are located from the second network DDN based on the cropped images. The training parameter settings of the ILN and DDN are shown in Tables II and III, respectively. The threshold $R_t$ in the crop module is set to 0.15. The best performance is obtained with the pretrained model VGG-16 and ResNet-101.

Fig. 10 shows the insulator defects detection results of aerial images. In Fig. 10(a), the blue box is the output of the cascading network, and the green boxes are the manually labeled insulators. Since there is no missing cap in Fig. 10(a), the output of the network is the detected insulator strings that are the same as the output of the ILN. The insulator string behind the tower was ignored for labeling, but the network can still locate it, showing that the proposed network learned the features of the insulators sufficiently to meet the requirement of robust insulator positioning. In Fig. 10(b), defective insulators and defect are bounded by the red box and the defect-free insulator is bounded by the blue box. As can be seen from Fig. 10(b), the smaller missing cap under a complex scene is effectively located by the cascading structure.

We also used three different pretrained models to test the cascading model: VGG-16, ResNet-50, and ResNet-101. Table IV shows the comparison of different pretrained models. For the ILN, the accuracies of the three pretrained models were nearly the same. However, the VGG-16 model was 30%–36% faster than ResNet. For the DDN, the accuracy performance comparison allows us to choose ResNet-101 as the only pretrained model. As can be seen from Table IV, the single ILN network may miss or misdetect some objects, but these objects are too small and will be removed through the crop module. This ensures that the cropped images input into DDN network are all potential defective insulators. The values of precision

and recall of the DDN are 0.91 and 0.958, respectively. The entire detection time cost of a single defective image is approximately 360 ms. These results show that our method enables the real-time location of insulator defects with high accuracy.

### B. Performance of Cascading Architecture

To evaluate the advantage of the proposed cascading architecture quantitatively, we compared it with other four competitive methods. The evaluation criteria for comparative experiments are those presented in (9)–(11). Since the located objects are the missing caps, the training and test sets for comparative experiments comprise 1056 negative samples.

*1) Faster R-CNN [17]:* It is a typical end-to-end object detection framework that is a combination of RPN and fast RCNN. Reference [24] uses Faster R-CNN's object detection network to detect dead end body of insulator. Therefore, in the comparative experiment, only the missing-cap area is labeled and located instead of insulator strings. We directly detected insulator defects on aerial images with a noncascading structure.

*2) ILN+CNN:* Zhao *et al.* [16] considered the defect detection of insulators as a detection and classification problem. However, the dataset used here is different, with this paper not making their implementation publicly available. Thus, in the comparative experiment, a detection module was built based on our ILN and the classification module followed this idea, using pretrained model VGG-19 to classify the cropped image into two categories: 1) normal insulators and 2) defective insulators. Owing to the method of classification, only the category of the cropped image is given and the position of the missing cap cannot be obtained. The defect detection criteria in this method is the same as in our proposed model. However, $T_P$ and $F_P$ denote the number of the cropped images (located insulator strings) that are correctly and incorrectly classified, respectively.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TAO *et al.*: DETECTION OF POWER LINE INSULATOR DEFECTS USING AERIAL IMAGES ANALYZED WITH CNNs

9

TABLE V
DETECTION PERFORMANCE FOR DIFFERENT METHODS

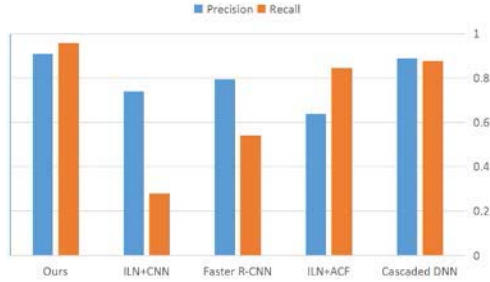| Model Name | $F_1$ | Speed |
|---|---|---|
| Ours | **0.933** | 359 ms/image |
| ILN+CNN [16] | 0.406 | 265ms/image |
| Faster R-CNN [17] | 0.643 | 180 ms/image |
| ILN+ACF [25] | 0.727 | **135 ms/image** |
| Cascaded DNN [27] | 0.885 | 387 ms/image |



Fig. 11. Effect analysis of the cascade architecture.

*3) ILN+ACF Detector [25]:* To compare with a traditional machine learning method, we also trained the ACF and Adaboosting models to detect and locate the missing cap. In the comparative experiment, the detection module was also built based on our ILN to locate insulator strings and an ACF-based detector was used to locate the missing cap. An ACF-based detector is a traditional state-of-the-art method in the field of pedestrian detection. Piotr's Computer Vision MATLAB Toolbox [26] was used to implement the ACF detector as a baseline. The ACF detector was trained with a sliding window size of $48 \times 48$ pixels and 4096 weak classifiers.

*4) Cascaded DNN Detector:* In [27], the cascaded DCNN-based detection method was proposed in a coarse-to-fine manner, including two detectors (SSD and YOLO) to sequentially localize the cantilever joints and their fasteners. This cascaded DNN detector was used to detect the missing cap with the Python implementation in the comparative experiment.

Table V and Fig. 11 show the quantitative comparison of five methods applied to the defective insulator images. The best performance is highlighted in bold. It can be seen that our proposed method achieved the best defect detection accuracy. The values of precision, recall and $F_1$ of the DDN are 0.91, 0.958, and 0.933, respectively. The cascaded DNN detector also exhibits good performance in both precision and recall. Compared with the two cascaded methods, the single object detection framework (Faster R-CNN) and classification method have achieved poor detection results. The ILN+ACF architecture is superior in detection-time consumption, but it shows very poor performance in terms of precision. Meanwhile, the recall of the ILN+CNN method is the lowest among the five methods.

Owing to the complicated backgrounds and different attitudes of the insulators, locating the insulator defects is beyond the capacity of traditional machine learning methods. Generally, the ACF detector has a large speed advantage. It uses aggregated multichannel features (including gradients, colors, etc.) and Adaboosting classifiers to implement the sliding-window detector. By using the feature pyramid to accelerate the detection process, it can reach 30 ft/s on a single core processor. However, its detection accuracy does not meet our requirements. Since the aggregated channel feature is a hand-crafted feature, it is not as robust as the learning feature. Adaboosting models are essentially shallow networks compared to deep learning networks, so the ACF detector performs poorly in the expressing the features of insulator defects. Meanwhile, distinguishing between normal or defective insulators was also beyond the capacity of the classification method. The feature expression of missing caps in the cropped images was weak, since the missing caps were too small. The shapes of the insulators in different cropped images were also different. Owing to the limitation of the number of insulator samples, it is also not possible to train classification networks from scratch. As a result, many defective insulators are difficult to distinguish from normal insulators, which results in the majority of defective insulators being classified as defect-free.

Unfortunately, Faster R-CNN shows poor performance in terms of precision and recall. This is not surprising since it is difficult to distinguish the $40 \times 40$-pixel objects in the $4608 \times 3456$-pixel raw input image. Owing to the pooling layer in the CNN shrunking the image, the low-level layers contain less semantic information about the objects. This will cause the missing cap area to be easily misdetected or missed using the single network. The precision and recall obtained by the method in [27] are close to ours, and a cascaded object detection framework was also used. However, both of its detection frameworks are existing frameworks that will miss some small objects.

As shown in Fig. 12, an intuitive comparison result is presented among the five methods. The green box denotes the normal insulator, while the defective insulators and the defects are bounded by the red box. Each box is associated with its category and a confidence score. Fig. 12(a) shows that Faster R-CNN misjudges the end insulator region as a missing cap. Fig. 12(b) shows that ILN+CNN successfully locates all insulators, but misclassifies all of the insulators as defect-free insulators. Fig. 12(c) shows that the ACF detector locates the missing cap successfully, but it seems like that more extraneous false results are obtained. Fig. 12(d) shows that the Cascaded DNN detector [27] successfully detected the missing cap, but it missed the insulator in the lower left-hand corner of the image and inaccurately positioned the insulator in the lower right-hand corner of the image. Fig. 12(e) shows that the proposed method, which combines ILN and DDN, has a good performance. As can be seen from Figs. 11 and 12, it is necessary to locate the insulator defects in stages, and the cascading structure can accurately detect the missing cap of the insulator from the aerial image under a complex scene. The defect detection of the insulator seems to be a classification task, but it can be better performed by detection in our application. This is because if the classification method is used, the entire insulator region must be used as a complete pattern for classification. In this case, a large number of real defective samples are needed, the defects of which locate at different positions with different sizes. In the absence of a sufficient

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10                                                                                                    IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS
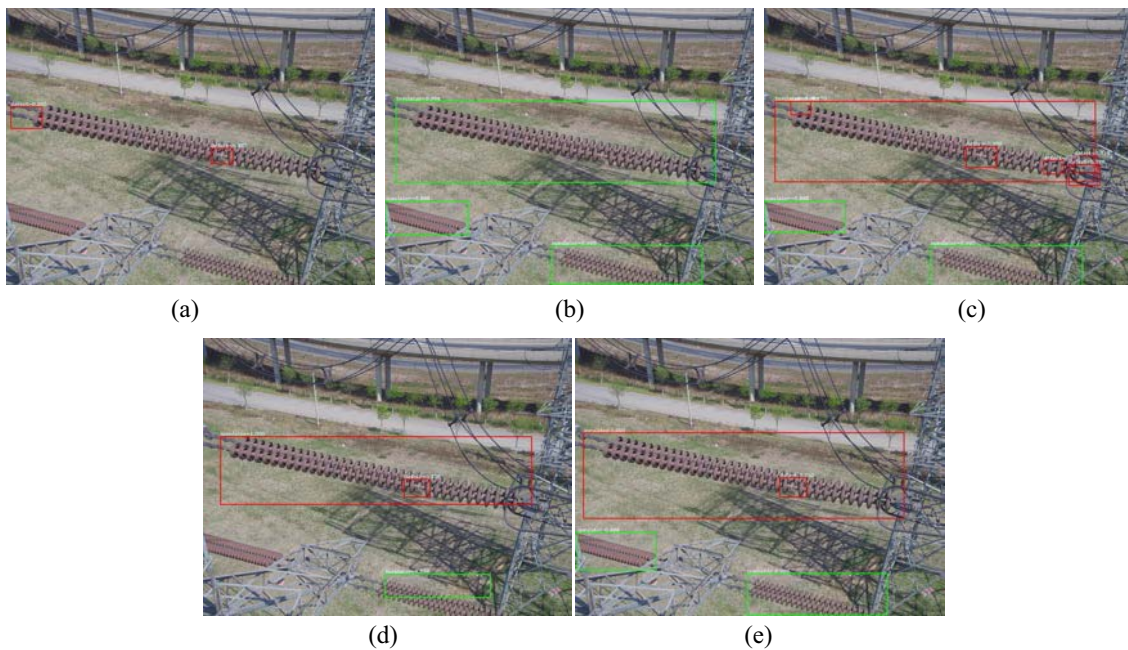


Fig. 12.   Five insulator defect detection results. (a) Faster R-CNN. (b) ILN+CNN. (c) ILN+ACF. (d) Cascaded DNN. (e) Proposed method.

dataset, DDN is designed for defect detection. The pattern of a defective area with the dumbbell structure is small, so that fewer samples are required. Moreover, the features are automatically learned by the network, rather than the features being hand-crafted as is in the ACF detector, so its effect is better than that of both CNN and ACF detector.

### C. Effect of Data Augmentation

To evaluate the contribution of the data augmentation used in this paper, we conduct experiments with different parts of the dataset and compare the defect detection results on these datasets. The methods in this paper and [28] are both evaluated. The experiments follow the same parameter settings as before, with the only difference being the data used to train the DDN. Six databases were built with the same number of augmentation methods for training data. The testing data of the six databases were the same. Note that since 25% or 50% of the image selection was random, some variation in image backgrounds was expected, since Gaussian blur and brightness changes were applied at the same time.

From Table VI, it can be seen that all the data augmentation methods increased the precision. When using changed background images as additional images to train the network, our method was far better than [28], indicating the effectiveness of the background fusion for aerial images. It is more useful than using Gaussian blur and modifying the brightness, with the combination of the four types of data producing the best result. This is because the diversity of defect data affects the effect of defect detection. In theory, the more actual defect data, the better the detection effect will be. The U-Net segmentation network is used to obtain defect contours. With precise contour and background fusion, a variety of new synthetic defects close to the real scene can be obtained.

TABLE VI
PERFORMANCE OF TRAINING DATASET USING
DIFFERENT AUGMENTATION

| Index | Training | Testing | Training Data Augmentation Method | Precision |
|---|---|---|---|---|
| 1 | 284 | 310 | Affine Transform [28] | 62.58% |
| 2 | 284 | 310 | Affine Transform +100% Background Fusion | 85.84% |
| 3 | 284 | 310 | Affine Transform +50% Background Fusion | 88.71% |
| 4 | 284 | 310 | Affine Transform +50% Background Fusion +50% Gaussian Blur | 89.35% |
| 5 | 284 | 310 | Affine Transform +50% Background Fusion +50% Brightness Change | 83.54% |
| 6 | 284 | 310 | Affine Transform +50% Background Fusion +50% Gaussian Blur +50% Brightness Fusion | 94.19% |

TABLE VII
PERFORMANCE OF DATA AUGMENTATION APPLIED TO OTHER METHODS

| Model Name | Precision | | Increased Precision |
|---|---|---|---|
| | Affine Transform | Our method | |
| ILN+ACF [25] | 0.578 | 0.638 | 6% |
| ILN+CNN [16] | 0.72 | 0.74 | 2% |
| Faster R-CNN [17] | 0.767 | 0.795 | 2.83% |

We also applied our data augmentation method to test other three other methods. The comparative results are summarized in Table VII. Obviously, compared with affine transformation, the precision of defect detection has a very high improvement with our method, especially for the ACF detector. This is because there are deficiencies in the classification method and Faster R-CNN method. In the classification method, the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TAO *et al.*: DETECTION OF POWER LINE INSULATOR DEFECTS USING AERIAL IMAGES ANALYZED WITH CNNs 11



Fig. 13. Insulator defects detection results. The green box denotes the normal insulator, the defective insulators and the defects are bounded by the red box. Each box is associated with its category and a confidence score.

insulators in different cropped images do not have unified scale and shape. The missing cap defects are too small compared to the original image in the Faster-RCNN. Therefore, simply increasing the amount of data is not enough to improve the detection accuracy of the above two models.

### D. Results on the CPLID

To demonstrate the effectiveness of our method, we evaluated it in the CPLID as well. This portion of insulator defects detection results is shown in Fig. 13. The green box denotes the normal insulator and the red box represents defective insulators or defects. The confidence is shown on the top left of each box. As can be seen from Fig. 13, our method can successfully detect most of the insulator defects in cluttered backgrounds. Moreover, there are some defect-free insulators and high-voltage power tower in Fig. 13(a), (j), and (t), which have a similar feature with the insulator defects. No insulator defect was found in these images, proving that the proposed method has a strong ability to exclude similar interference. Overall, the proposed approach has significant potential for a wide range of applications. However, detection of defective

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS

insulators through aerial images is still not sufficiently based on data-driven methods. In a future study, we will utilize the GAN to generate defective insulator images.

## VI. CONCLUSION

In this paper, a novel deep CNN cascading architecture is presented for both localization and defect detection of insulators. Defect inspection is converted to a two-level object detection problem based on the cascading CNN. To remedy the small number of defect images in real inspection environments, a data augmentation methodology is proposed, and we describe how to use it to improve the insulator defect detection performance of CNNs. The proposed methodology can significantly contribute to a new horizon of deep learning for the detection of defects using aerial images. The defect detection precision and recall of the proposed method are 0.91 and 0.96 using our standard insulator dataset. Experimental results have shown that our method can meet the robustness and accuracy requirements of insulator defect detection.

## REFERENCES

[1] K.-C. Park, Y. Motai, and J. R. Yoon, "Acoustic fault detection technique for high-power insulators," *IEEE Trans. Ind. Electron.*, vol. 64, no. 12, pp. 9699–9708, Dec. 2017.

[2] Y. Zhai, D. Wang, M. Zhang, J. Wang, and F. Guo, "Fault detection of insulator based on saliency and adaptive morphology," *Multimedia Tools Appl.*, vol. 76, no. 9, pp. 12051–12064, May 2017.

[3] C. Martinez, C. Sampedro, A. Chauhan, and P. Campoy, "Towards autonomous detection and tracking of electric towers for aerial power line inspection," in *Proc. Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, May 2014, pp. 284–295.

[4] C. Sampedro, C. Martinez, A. Chauhan, and P. Campoy, "A supervised approach to electric tower detection and classification for power line inspection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 1970–1977.

[5] H. Zhao, R. Dai, and C. Xiao, "A machine vision system for stacked substrates counting with a robust stripe detection algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.

[6] Y.-P. Huang, L. Sithole, and T.-T. Lee, "Structure from motion technique for scene detection using autonomous drone navigation," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.

[7] A. Çağlar, Y. Rezaeitabar, S. Dogru, and L. Ulusoy, "Railway fastener inspection by real-time machine vision," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 7, pp. 1101–1107, Jul. 2015.

[8] Z. Zhao, N. Liu, and L. Wang, "Localization of multiple insulators by orientation angle detection and binary shape prior knowledge," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 6, pp. 3421–3428, Dec. 2015.

[9] B. Li, D. Wu, Y. Cong, Y. Xia, and Y. Tang, "A method of insulator detection from video sequence," in *Proc. 4th Int. Symp. Inf. Sci. Eng. (ISISE)*, Dec. 2012, pp. 386–389.

[10] S. Liao and J. An, "A robust insulator detection algorithm based on local features and spatial orders for aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 963–967, May 2015.

[11] Z. Zhao, G. Xu, and Y. Qi, "Representation of binary feature pooling for detection of insulator strings in infrared images," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 5, pp. 2858–2866, Oct. 2016.

[12] M. Oberweger, A. Wendel, and H. Bischof, "Visual recognition and fault detection for power line insulators," in *Proc. 19th Comput. Vis. Win. Workshop*, Feb. 2014, pp. 1–8.

[13] Q. Wu and J. An, "An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3613–3626, Jun. 2014.

[14] V. S. Murthy, K. Tarakanath, D. K. Mohanta, and S. Gupta, "Insulator condition analysis for overhead distribution lines using combined wavelet support vector machine (SVM)," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 1, pp. 89–99, Feb. 2010.

[15] M. J. B. Reddy, C. B. Karthik, and D. K. Mohanta, "Condition monitoring of 11 kV distribution system insulators incorporating complex imagery using combined DOST-SVM approach," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 2, pp. 664–674, Apr. 2013.

[16] Z. Zhao, G. Xu, Y. Qi, N. Liu, and T. Zhang, "Multi-patch deep features for power line insulator status classification from aerial images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3187–3194.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[18] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2012, pp. 746–760.

[19] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, Dec. 2016, pp. 379–387.

[20] M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof, "TVSeg-interactive total variation based image segmentation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2008, pp. 44–46.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, Oct. 2015, pp. 234–241.

[22] T. Chen *et al.*, "MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems," *arXiv preprint arXiv: 1512.01274*, 2015.

[23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Isserman. *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*. Accessed: May 20, 2017. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html

[24] I. E. Nordeng, A. Hasan, D. Olsen, and J. Neubert, "DEBC detection with deep learning," in *Proc. Scandinavian Conf. Image Anal. (SCIA)*, May 2017, pp. 248–259.

[25] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, Aug. 2014.

[26] P. Dollar. (2006). *Piotr's Computer Vision MATLAB Toolbox (PMT)*. [Online]. Available: http://pdollar.github.io/toolbox/

[27] J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 257–269, Feb. 2018.

[28] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework," *GISci. Remote Sens.*, vol. 54, no. 5, pp. 741–758, May 2017.

**Xian Tao** received the M.Sc. degree in mechanical and automotive engineering from the China University of Mining and Technology, Beijing, China, in 2013 and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, in 2016.

He is currently an Assistant Professor with the Research Center of Precision Sensing and Control, IACAS. His current research interests include deep learning and automated surface inspection for industry.

**Dapeng Zhang** received the B.Sc. and M.Sc. degrees in mechatronic engineering from the Hebei University of Technology, Tianjin, China, in 2003 and 2006, respectively, and the Ph.D. degree in mechatronic engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2011.

He is currently an Associate Professor with the Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing, and the University of the Chinese Academy of Sciences, Beijing. His current research interests include robotics and automation, in particular, medical robot and virtual robotic surgery.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TAO *et al.*: DETECTION OF POWER LINE INSULATOR DEFECTS USING AERIAL IMAGES ANALYZED WITH CNNs 13

**Zihao Wang** received the B.Sc. degree in control science and engineering from the Civil Aviation University of China, Tianjin, China, in 2015, where he is currently pursuing the M.Sc. degree in control science and engineering with the Sino-European Institute of Aviation Engineering.

His current research interests include computer vision and deep learning.

**Hongyan Zhang** received the B.S. and M.S. degrees in applied physics and telecommunication engineering from Xidian University, Xi'an, China, in 2000 and 2003, respectively, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Science, Beijing, China, in 2011.

He is currently with the School of Information Science and Technology, Hainan Normal University, Haikou, China. His current research interests include applied mathematics, computer vision, and nondestructive testing.

**Xilong Liu** received the B.Sc. degree in control science and engineering from Beijing Jiaotong University, Beijing, China, in 2009 and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, in 2014.

He is currently an Associate Professor with the Research Center of Precision Sensing and Control, IACAS. His current research interests include image processing, pattern recognition, visual measurement, and visual scene cognition.

**De Xu** (M'05–SM'09) received the B.Sc. and M.Sc. degrees in control science and engineering from the Shandong University of Technology, Jinan, China, in 1985 and 1990, respectively, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2001.

He has been with Institute of Automation, Chinese Academy of Sciences (IACAS), since 2001, where he is currently a Professor with the Research Center of Precision Sensing and Control. His current research interests include robotics and automation, such as visual measurement, visual control, intelligent control, welding seam tracking, visual positioning, microscopic vision, and micro-assembly.