IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 3, 2018, accepted December 31, 2018, date of publication January 7, 2019, date of current version January 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891123

Insulator Detection in Aerial Images for Transmission Line Inspection Using Single Shot Multibox Detector

XIREN MIAO, XINYU LIU, JING CHEN, (Member, IEEE), SHENGBIN ZHUANG, JIANWEI FAN, AND HAO JIANG, (Member, IEEE)

College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

Corresponding author: Hao Jiang (jiangh@fzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703105 and Grant 61703106, in part by the Natural Science Foundation of Fujian Province of China under Grant 2017J01500, in part by the Foundation of Fujian Educational Committee under Grant JZ160415 and Grant JAT170107, in part by the Research Foundation of Fuzhou University under Grant XRC-1623 and Grant XRC-17011, and in part by the Student Research Training Program of Fuzhou University under Grant 201710386043.

ABSTRACT The detection of insulators with cluttered backgrounds in aerial images is a challenging task for an automatic transmission line inspection system. In this paper, we propose an effective and reliable insulator detection method based on a deep learning technique for aerial images. In the proposed deep detection approach, the single shot multibox detector (SSD), a powerful deep meta-architecture, is incorporated with a strategy of two-stage fine-tuning. The SSD-based model can realize automatic multi-level feature extractor from aerial images instead of manually extracting features. Inspired by transfer learning, a two-stage fine-tuning strategy is implemented using separate training sets. In the first stage, the basic insulator model is obtained by fine-tuning the COCO model with aerial images, including different types of insulators and various backgrounds. In the second stage, the basic model is fine-tuned by the training sets of the specific insulator types and specific situations to be detected. After the two-stage fine-tuning, the well-trained SSD model can directly and accurately identify the insulator by feeding the aerial images. The results show that both the porcelain insulator and composite insulator can be quickly and accurately identified in the aerial images with complex background. The proposed approach can enhance the accuracy, efficiency, and robustness significantly.

INDEX TERMS Insulator detection, deep learning, single shot multibox detector (SSD), fine-tuning.

I. INTRODUCTION

Transmission line inspection plays an essential role in a power transmission system to ensure the safety and uninterrupted reliable operation of power service. In transmission line, the insulator is a widely used equipment with the dual function of electrical insulation and mechanical support. The condition monitoring of the insulators becomes one of the most important and difficult task during the inspection of the transmission line [1]. The earlier detection of the defects of insulators can reduce power cuts and prevent the huge economic losses and bad customer care in power transmission system. It is necessary to inspect the insulators of transmission line and confirm that their performance meets the specification requirements.

Instead of traditional manual patrol, transmission line inspection by unmanned aerial vehicles (UAVs) is employed to realize more automatic and efficient inspection [2], [3]. During the UAV inspection, a large number of aerial images and videos of insulators as the photographic records will be generated and can be used to prepare for the operation and maintenance (OM) thereby reducing operational cost. In order to achieve automatic inspection and intelligent diagnosis, the key prerequisite is accurately detecting the insulators from the aerial images for further inspection tasks including fault diagnosis, camera tracking, data management, etc. However, the aerial images from UAVs contain the cluttered backgrounds and various types of insulators. The external disturbing factors, such as the changing visual angle,

different lighting, and partial occlusion, make it more difficult to detect the insulators. Owing to the wide view of UAVs, the original non-processing aerial images are high-resolution and larger than the nature scene images. Hence, the insulator detection in aerial images is a challenging task.

To identify the insulators in aerial images, the existing detection methods mainly extract the features of aerial images by means of image processing. The color, shape and texture features are commonly used to distinguish the insulators from the complex background. Zhang *et al.* [4] adopted a connected component analysis method in hue-saturation-intensity color space for tempered glass insulator identification. Reddy *et al.* [5], [6] applied discrete orthogonal S-transform (DOST) with adaptive neuron-fuzzy inference system (ANFIS) to find the locations of the insulators and ascertain the condition of the insulators using extracted color features. The DOST-SVM method is developed to distinguish proper bounding boxes containing the insulators by applying support vector machine and k-means clustering. The color feature based methods are sensitive to the complex background and need a well-adjusted threshold parameter.

Several insulator detection methods based on the utilization of the shape and texture feature information have been developed. Based on the statistical properties of the insulator shape feature, Li *et al.* [7] presented the profile projection method and trained a support vector machine classifier for insulator location. Oberweger *et al.* [8] introduced a circular descriptor based method by extracting the difference of Gaussians (DoG) key-points and clustering them through k-nearest neighbor (KNN). Liao and An [9] proposed the multiscale-multifeature (MSMF) descriptor for shape feature extraction and find several spatial orders features to improve the robustness. Zhao *et al.* [10] proposed an insulator detection approach based on orientation angle detection and binary shape prior knowledge which can handle different orientation angles in complex aerial image. Wu *et al.* [11], [12] used semi-local operator to extract the semi-local texture distribution and performed image segmentation by global minimum active contour to extract the insulator from aerial images. Wang *et al.* [13] presented a method merged the color, shape and texture of insulator, local binary pattern and color component analysis are utilized to filter the candidate regions which obtained by parallel line features extracting. However, the detection method based on the image feature extraction generally require hand-crafted feature extractors, and the performance greatly depends on the complexity of background and result of image processing. In addition, few studies consider the trade-off between the detection accuracy and speed in the insulator detection process. When processing the high-resolution aerial images, it is time-consuming to find the location of insulators.

Deep learning is emerging as a leading machine learning technique in the imaging recognition and computer vision domains. Unlike other machine learning methods with the shallow architectures, deep learning is composed of multiple layers of neural network that provides a different level of abstraction to enhance the learning ability on large and complex data [14]. Among various deep learning techniques, the convolutional neural network (CNN) is the most popular deep learning network and proven to be a powerful tool for image processing. More recently, many CNN-based methods have achieved great success in object detection, such as Faster R-CNN [15], R-FCN [16], SSD [17]. But little work focuses on applying the CNN to detect the insulators in aerial images.

In this paper, we attempt to explore deep learning based method to detect insulators in the aerial images for power transmission line inspection system. A deep insulator detection architecture is developed based on the single shot multibox detector (SSD) and fine-tuning strategy. The single shot multibox detector (SSD), proposed by Liu *et al.* [17], is one of popular deep CNN framework for object detection. Unlike traditional guideless hand-crafted feature extractors, SSD based method can automatically and efficiently extract multi-level features from aerial image.

Due to the regional differences and data confidentiality of power transmission line inspection, each inspection region is relatively independent and few training data are available. Each electricity company of the inspection region would favor training a single-scene model that utilizes its local data for high accuracy detection rather than producing a multi-scenes model with feature dispersion problem. For training a single-scene model, the fine-tuning is usually applied based on the COCO dataset. However, there is no images of power system domain in the COCO dataset, thus, the knowledge transfer between COCO dataset and specific insulator dataset meets with the challenges of domain gap and data size gap. In order to solve the issues of insufficient training data and domain gap, a two-stage fine-tuning process is presented. In the first stage, the basic insulator model is obtained by fine-tuning the COCO model with the aerial images including insulators with extensive characteristics. Then in the second stage, the basic model is fine-tuned by the specific small-size training data related to the insulator detection of target scene. The two-stage fine-tuning strategy can enhance the detection accuracy and robustness of the detection model. Additionally, the basic insulator model can be applied as a common base pre-trained model for insulator detection in power transmission line inspection. For other scenarios of inspection region, the electricity company only needs to perform the second fine-tuning stage without repeating the first stage. In other words, there is no extra workload for electricity company once the basic insulator model is well-trained.

The rest of this paper is organized as follows. We describe the proposed framework of insulator detection in Section II. The experimental results on the different tasks are presented in Section III. In section IV, we draw conclusions from the results and also discuss about our future work plans.

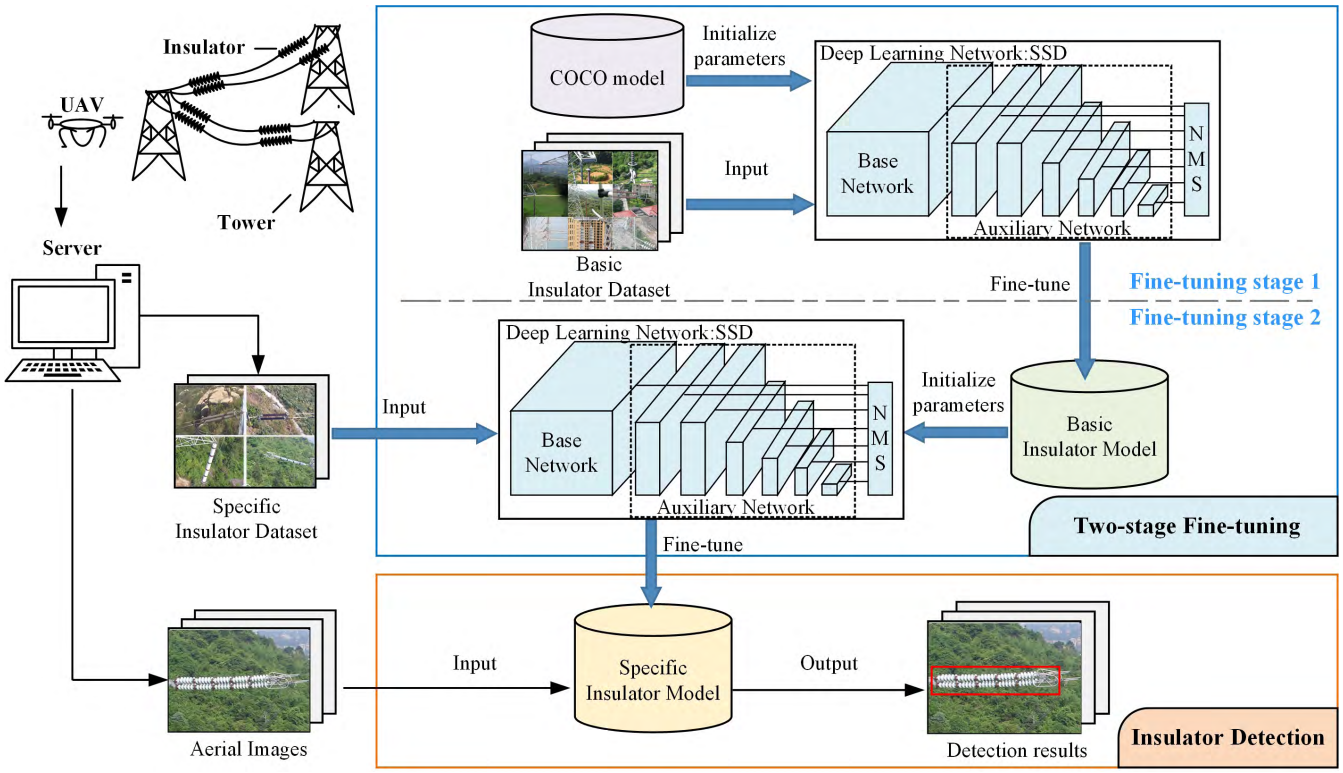

FIGURE 1. Schematic diagram of the deep learning based insulator detection approach. SSD: single shot multibox detector, COCO: a large-scale image dataset from Microsoft.

II. METHODOLOGY

A. FRAMEWORK OF INSULATOR DETECTION

The schematic diagram of the proposed insulator detection method in aerial images for transmission line inspection is illustrated in Fig. 1. The deep learning based architecture is consist of three parts: aerial image preprocessing, SSD model training process, and real-time insulator detection. First, the aerial images are collected from the UAVs for transmission line inspection. Two different datasets are defined and will be used in the fine-tuning stages. The original aerial images are transformed to the standard data format for model training after preprocessing such as resizing, cropping, labeling, etc. In training process, SSD is utilized to train the detection model by feeding the images. Different from other deep learning algorithms, we apply a two-stage fine-tuning procedure to enhance the detection accuracy and robustness. Once the specific insulator model is well-trained, it can be directly used to realize the insulator detection from the aerial images taken by UAVs for transmission line inspection system.

B. IMAGE PREPROCESSING

The aerial images are captured by UAVs. In the proposed method, there are two kinds of training sets: basic insulator dataset and specific insulator dataset. The training data for the first fine-tuning stage is defined as basic dataset. The basic dataset is composed of the aerial images collected

FIGURE 2. Samples of basic insulator dataset including the aerial images of various types of insulators with complex backgrounds.

from different transmission line inspection regions. Owing to the different geographic environments and different voltage levels of transmission line inspection, the aerial images in basic dataset contain various types of insulators with various complex backgrounds. The backgrounds not only have unique environments in the electricity domain, but also have diverse transmission line components besides insulators. As shown in Fig. 2, the types of insulators include the porcelain insulator, composite insulator, and glass insulator. And the scenes of backgrounds are main factors to distinguish different inspection regions in aerial images that involve

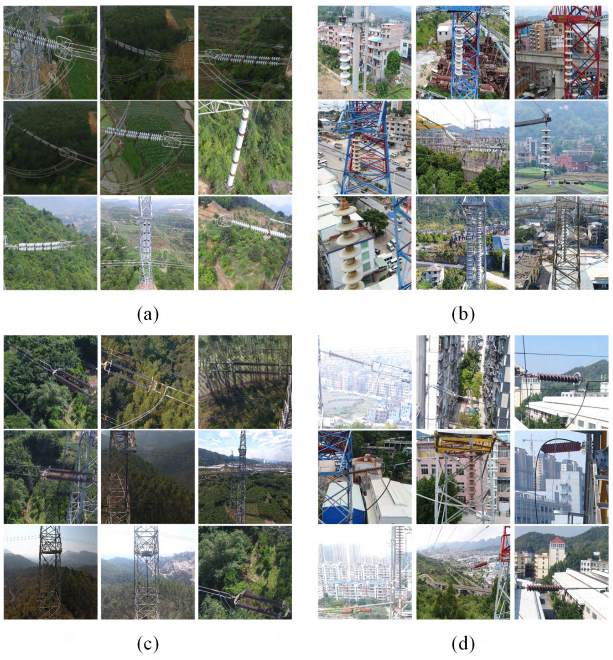

FIGURE 3. Samples of specific insulator dataset including the aerial images of (a) porcelain insulator with forest backgrounds (b) porcelain insulator with building backgrounds (c) composite insulators with building backgrounds (d) composite insulator with building backgrounds.

forest, building, desert, etc. The specific dataset is a relatively small-scale training set that is used to fine-tune the basic insulator model. The aerial images in specific dataset only contain the particular types of scenes and the specific types of insulators to be detected, as shown in Fig. 3. As a way of data augmentation, the images in insulator specific dataset need to be flipped (horizontally and vertically) and cropped to fight with the over-fitting problem. The graphical image annotation tool ''LabelImg'' is applied to label the aerial images. The annotations of images are imageID, class of object, path of image, coordinates of boxes. All the images are converted into TFRecord file format, which is a simple record-oriented binary format commonly used in TensorFlow.

C. TWO-STAGE FINE-TUNING

In order to solve the issues of data insufficiency and domain gap in the traditional one-stage fine-tuning, the proposed approach presents a two-stage fine-tuning procedure to train the insulator detection model. In the application of transmission line inspection, the electricity company is only responsible for the inspection task in the local area and prefers the single-scene model to the multi-scenes model, due to the feature dispersion problem in multi-scenes model.

Generally, the deep learning model requires a large quantity of labeled training samples to train a deep convolutional architecture. However, when using the local data to train a single-scene model, it is difficult to guarantee that the available training data is sufficient, owing to the organizational independence and data confidentiality of each regional power grid. But by simply using the data augmentation method such as flipping and cropping, the over-fitting problem caused by insufficient data cannot be resolved very well and the detection performance fails to meet the high accuracy requirement. In common, the transfer learning technique is utilized to combat with the data insufficiency by fine-tuning the deep learning model with COCO dataset. Transfer learning tries to transfer the knowledge learned from the old task called ''source task'' to the new task called ''target task''. In the field of insulator detection, the traditional one-stage fine-tuning transfers the knowledge from COCO dataset to insulator dataset. However, the COCO dataset does not have images of electricity domain and the quantity of insulator dataset is minimal compared to the COCO dataset. Therefore, the one-stage fine-tuning for insulator detection exists huge domain gap and data size gap that affect the model performance.

For addressing the issues mentioned above, this paper introduces a two-stage fine-tuning strategy that inserts a transitional task with basic insulator dataset between the source task with COCO dataset and the target task with specific insulator dataset. The implementation of the two-stage fine-tuning strategy is a generic-to-specific, coarse-to-fine approach, as shown in Fig. 1.

In the first fine-tuning stage, the COCO model and the basic insulator dataset are applied. The COCO model is trained using the MS COCO dataset, containing more than 200000 images and 80 object categories [18]. We use the COCO model to initialize the model parameters in order to improve the contextual richness of the basic insulator model [19], [20]. But in the COCO dataset, there are no relevant images in the electric power field. Hence, the basic insulator dataset, which contains images depicting diverse types of insulators and diverse backgrounds, is gathered to provide a widespread general features for insulator detection. Combing the COCO model and basic insulator dataset, the basic insulator model is built as a coarse model for subsequent training.

The second fine-tuning stage aims to train the specific model for insulator detection in specific inspection region that has its distinctive insulator types and scenes. In this procedure, the specific insulator dataset is utilized. Different from the basic insulator dataset, the specific insulator dataset only contains the aerial images relating to the specific transmission line. Through inputting the specific insulator dataset, the parameters of basic insulator model are fine-tuned and more meaningful features of insulators are learned. Compared with the traditional procedure of insulator detection, the two-stage fine-tuning will not increase the workload to the electricity company. As shown in Fig. 4, the proposed method can be easily extended to apply to additional scenarios. If the inspection task is changed to a different region, the training procedure only needs to replace the specific insulator dataset according to the specific scene, while the basic insulator model remains the same. This means that the basic model can be repeatedly used without redoing the first fine-tuning stage. This way can greatly reduce the training cost and meanwhile improve the flexibility of the training process for

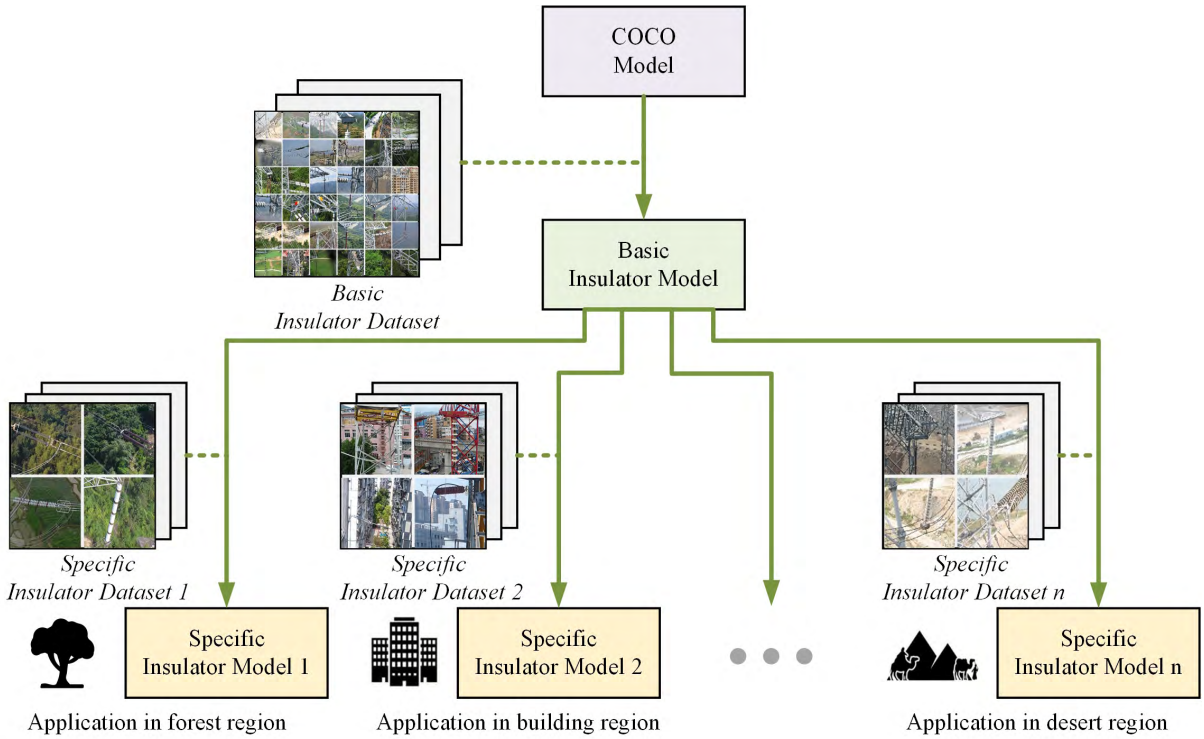

FIGURE 4. Two-stage fine-tuning for different application regions.

insulator detection model. Additionally, through two-stage fine-tuning strategy, the accuracy and robustness of model can be enhanced when training data is insufficient. Once the final specific insulator model is well-trained, it can be used to identify the insulator in aerial images for transmission line inspection system.

D. SINGLE SHOT MULTIBOX DETECTOR

The SSD handles the problem of object detection in real-time application with high accuracy. This approach is based on a feed-forward convolutional network that generates a fixed-size collection of bounding boxes and scores for the presence of an object class in each box by using the small convolutional filter with the size of 3×3, and then utilizes Non Maximum Suppression (NMS) to produce the final detections. The detection speed of SSD is faster than the previous state-of-the-art for single shot detectors. Furthermore, SSD has a contribution that it can deal with the multiple scales of input images which is useful in insulator detection. The SSD network architecture is shown in Fig. 5. The framework mainly consists of two parts: base network, auxiliary structure.

1) BASE NETWORK

The Base Network is a standard architecture used for high quality image classification which is truncated before the classification layers. The VGG-16 network is chosen as the base network in [17]. Depending on the application scenarios, the base network can be replaced by appropriate CNN architecture for classification. In this paper, the base network of SSD is MobileNet. MobileNet is an efficient CNN which only has 4.2 million parameters compared with VGG16 (138 million parameters) that can be used at the mobile end [21]. When testing on the ImageNet benchmark [22], MobileNet(70.6%) is nearly as accurate as VGG16(71.5%) while being 32 times smaller. MobileNet is easy to transplant in embedded device which means the proposed method has the potential for real-time detection in UAV. [23].

2) AUXILIARY NETWORK

The SSD adds auxiliary structure after the base network to produce detections with the following key features:

A set of convolutional feature layers is appended to the end of the truncated base network. These layers decrease in size progressively in order to predict the object at multiple scales. The receptive field of lower convolutional layer is small, and it is bigger in higher layer. The network has the ability to detect the object of multiple scales and improve the accuracy of recognition.

Each feature map corresponds to a fixed set of default bounding boxes. The default boxes can be seen as the initial detections of the object in input image and then firstly filtered by confidence which computed by CNN. After that, NMS is utilized for final filtration and then output the detection results. Assume that there are M (M=6 in this paper) feature maps for prediction. The scale of the default boxes for each feature map is computed as:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{M - 1}(k - 1), \quad k \in [1, M] \qquad (1)$$

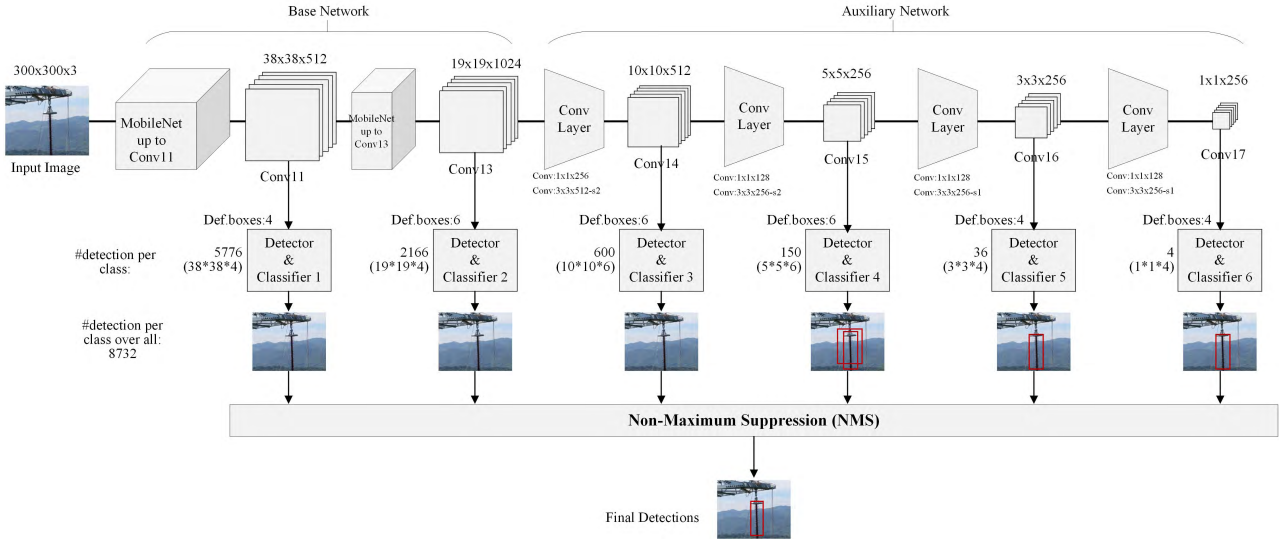

FIGURE 5. Network architecture of SSD. The model adds several feature layers to the end of a base network based on CNN. These feature layers predict the offsets to default boxes (Def.boxes) of different scales and aspect ratios and their associated confidences. A $m \times n$ feature map after convolution can generate $m \times n \times b$ detections per class where b is the number of Def.boxes for each feature map cell (A $m \times n$ feature map has $m \times n$ cells).

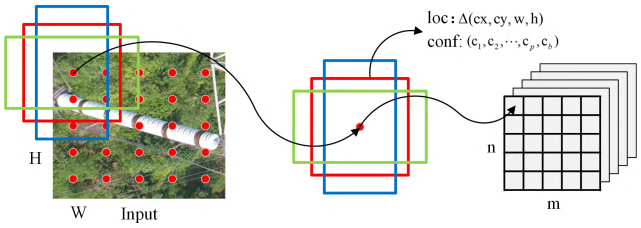

FIGURE 6. Example of default box generation for one cell from a feature map.

where $s_{\min} = 0.2$ and $s_{\max} = 0.95$, meaning the lowest layer has a scale of 0.2 and the highest layer has a scale of 0.95. The width of default box is $s_k W\sqrt{a_r}$ and the height is $s_k H/\sqrt{a_r}$ where a_r is the aspect ratio and (W, H) is the size of input. In practice, one can design a special distribution of default boxes by changing the size and number of the box for each feature map. An example of default box generation is shown in Fig. 6, the size of input image is $W \times H$, the feature map has $m \times n$ cells. If $a_r \in \{1, 2, 1/2\}$, then there are three default boxes(width, height) generated for each cell. For each default box, it predicts both the shape offsets(cx, cy, w, h) of location and the confidences for C object categories $(c_1, c_2, \ldots, c_p, c_{background})$. Specifically, each cell corresponds to b default boxes, every default box computes C class scores and the 4 offsets relative to the original default box shape. This result in a total of $(C+4)bmn$ outputs for a $m \times n$ feature map. The class number C is 3 in this paper: porcelain insulator, composite insulator and background.

Once the default boxes are determined, the loss function and back propagation can be applied end to end. The definition of the loss function is as follows. Suppose $x_{ij}^p = \{0, 1\}$ is an indicator for matching the i-th default box to the j-th ground truth box of category p. When $x_{ij}^p = 1$ meaning the Intersection-over-Union (IoU) described in (2) between default box and ground truth box is higher than a threshold(0.5).

$$IoU = \frac{DefaultBox \cap GroundTruthBox}{DefaultBox \cup GroundTruthBox} \tag{2}$$

The overall objective loss function is a weighted sum of the localization loss (loc) and the confidence loss (conf):

$$L(x, c, l, g) = \frac{1}{N}\left(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)\right) \tag{3}$$

where N is the number of matched default boxes.

$$L_{conf}(x, c) = -\sum_{i \in Pos}^{N} x_{ij}^p \log\left(\hat{c}_i^p\right) - \sum_{i \in Neg} \log\left(\hat{c}_i^0\right) \tag{4}$$

$$\hat{c}_i^p = \frac{\exp\left(c_i^p\right)}{\sum_p \exp\left(c_i^p\right)} \tag{5}$$

The confidence loss is the Softmax loss over multiple classes confidences (c). The c_i^0 refers to the background class corresponding to negative default boxes that do not have object. And the c_i^p refers to the object class corresponding to positive default boxes that have object with category p.

$$L_{loc}(x, l, g) = \sum_{i \in Pos}^{N} \sum_{m \in Box} x_{ij}^k smooth_{L1}\left(l_i^m - \hat{g}_j^m\right) \tag{6}$$

$$\hat{g}_j^{cx} = \left(g_j^{cx} - d_i^{cx}\right)/d_i^w \quad \hat{g}_j^{cy} = \left(g_j^{cy} - d_i^{cy}\right)/d_i^h \tag{7}$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right) \quad \hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right) \tag{8}$$

TABLE 1. Description of training datasets.

	Basic insulator dataset	Specific insulator dataset I	Specific insulator dataset II
Insulator type	various	porcelain & composite	porcelain & composite
Image scene	various	forest	building
Data size	6700	450	455

The localization loss is the Smooth L1 [24] loss between the predicted box (l) and the ground truth box (g) parameters when the default bounding box defined as d. The 4 shape offsets $m \in \{cx, cy, w, h\}$ are defined as the center (cx, cy) of the bounding box and its width (w) and height (h). Once the gradient is computed with backpropagation, an optimizer which called RMSProp is utilized to update the model parameters. The RMSProp [25] utilizes a moving average of squared gradients to normalize the gradients, and modulates the learning rate of each weight based on the magnitudes of its gradients.

III. EXPERIMENTAL RESULTS

In this section, in order to evaluate the performance of the proposed approach, two insulator detection models are trained and used to identify both the porcelain insulator (PI) and the composite insulator (CI) in aerial images. The complex backgrounds are considered, including forest and building scenes. Then, we present the advantages of the two-stage fine-tuning strategy by carrying out the training process under different conditions. Finally, a comparison is made between the results of the proposed method and the other existing methods.

A. DATA PREPARATION

In this paper, all the aerial images are provided by China Power Grid from the UAVs for power transmission line inspection. The DJI Phantom 4 Professional is used in transmission line inspection. DJI Phantom 4 is equipped with a 4K video camera that has a 1/2.3" CMOS sensor, 94-degree field of view, 12.4 megapixel (4000×3000 pixels). Therefore, all the original aerial images are 4000×3000 pixels. For training, three datasets are used in the experiments, as shown in Table 1. Basic insulator dataset contains 6700 images that collected from different transmission line inspection regions. The basic insulator dataset is used to train the basic insulator model in the first fine-tuning stage. The sizes of specific insulator dataset I and specific insulator dataset II are 450 and 455, respectively. The specific insulator dataset I contains images of porcelain and composite insulator in forest scenes. In the application of transmission line inspection, the forest scenes are corresponding to mountainous regions. Different from specific insulator dataset I, specific insulator dataset II is used to train the model for insulator detection in building scenes corresponding to urban regions. For each model, 200 aerial images are used as a testing set. Among them, 100 aerial images are PI images, and the others are CI images. After pre-processing, a single ground-truth box is labeled in each aerial image.

B. IMPLEMENTATION

The proposed approach is implemented in TensorFlow [26] with Object Detection API [27]. It is running on a computer equipped with Intel Core i7-6850K CPU, a NVIDIA GeForce GTX 1080 Ti GPU, and 32GB of RAM memory under Ubuntu 16.04 LTS. In order to verify the embedded potential, we also test the proposed method on an embedded device so-called JETSON TX2 with ARM A57/2 and HMP Denver 2/2 CPU, a NVIDIA Pascal GPU, and 8GB of RAM memory under JetPack 3.1 system. The TX2 can be loaded on a UAV. To realize real-time inspection at the present, the aerial images are transmitted to the computer from UAVs in real time by the OcuSync system of DJI.

The parameters of the proposed method are set as follows: aspect ratios in default box generator $a_r \in \{1, 2, 3, 1/2, 1/3\}$, $s_{min} = 0.2$, and $s_{max} = 0.95$. In the NMS phase, each class can retain 100 detections. The score threshold is 0.01, the IoU threshold is 0.6 and confidence threshold is 0.4. Here, the Root Mean Square Prop (RMSProp) [25] is used to train the model. The initial learning rate is 0.004, momentum is 0.9, and batch size is 24. To prevent overfitting and save the time-consuming in some hyper-parameter tuning, Batch Normalization (BN) is applied after convolutional layers, and before non-linearities layers [28]. Max training step is set to be 20000. In first and second fine-tuning stages, all layers of the deep learning model are fine-tuned. To evaluate the detection performance, four widely used metrics are applied, including precision, recall, precision-recall curve (PRC), and average precision (AP).

C. DETECTION RESULTS

1) INSULATOR DETECTION IN FOREST SCENES

First, we use the basic insulator dataset and the specific insulator dataset I to train the first SSD model using the proposed method. The well-trained model can detect both PI and CI from a forest scene in aerial images. The testing dataset is compose of 100 aerial images of CI and 100 aerial images of PI in forest scenes. Fig. 7 shows the APs of PI detection and CI detection at different training steps. During the entire 20000 training steps, the tests were conducted at every 2000 steps. It can be seen that for PI and CI detection, the APs reach 92.12% and 82.06% after 10 thousand steps, respectively. At the 20000th step, the AP for PI detection is 94.12% and the AP for CI detection is 86.70%. In addition, the average running time per step is 0.4 s.

The testing performance criteria are listed in Table 2. For the detection of PI, it achieves an excellent performance with the precision of 93.75% and the recall of 90%. For the detection of CI, the precision and recall are 85.29% and 87%, respectively. The precision and recall of CI detection are slightly worse than that of PI detection. Because the task of the CI detection is more difficult and complicated than PI detection owing to the low-contrast color of the caps and the narrow shapes of composite insulators in the images. The corresponding precision-recall curves are shown in Fig. 8.

TABLE 2. Performance results of insulator detection in forest scenes.

Insulator Type	TP	FP	FN	Precision(%)	Recall(%)	AP	PC Time (ms)	TX2 Time (ms)
porcelain	90	6	10	93.75	90	94.12	22.96	62.87
composite	87	15	13	85.29	87	86.7	22.89	63.54

Note: TP is true positive, FP is false positive, FN is false negative.

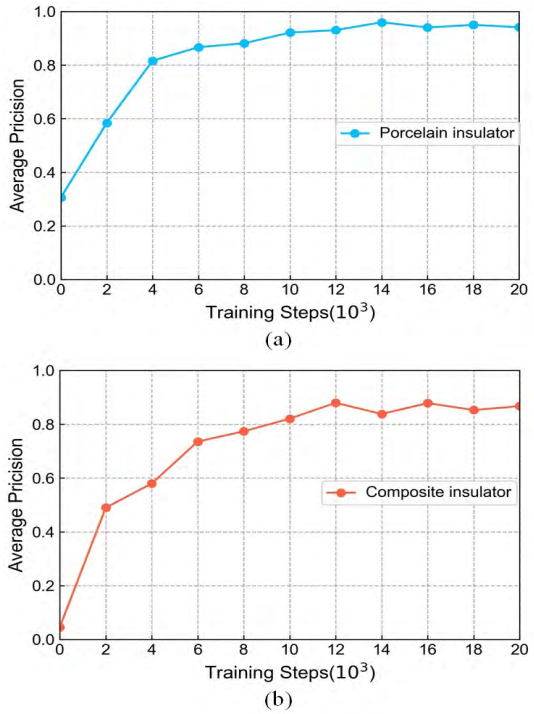

FIGURE 7. Average precision vs. the number of training steps. (a) Porcelain insulator detection. (b) Composite insulator detection.

From the precision-recall curves, it can be seen that our detector remains high precision of 95% when the recall varies from 0 to 85% in Fig. 8 (a), and 0 to 73% in Fig. 8 (b). It indicates that the proposed method yields a good trade-off between precision and recall with high AP. The average computation time for a single image is $\sim$ 23 ms on PC and $\sim$ 63 ms on TX2.

Fig. 9 and Fig. 10 illustrate some testing results of aerial images in the forest scenes. In the images, the proposed method can automatically generate a green box to denote the localization of PI and a blue box to denote the localization of CI. From Fig. 9, the porcelain insulators have a distinct color from the backgrounds of forests and can be easily identified in case (a) and case (b), but in case (c) and case (d) the porcelain insulators also can be correctly located in the cluttered backgrounds with a mix of forests, power line towers and the roads. In Fig. 10, the composite insulators are quite difficult to be distinguished from the backgrounds. Especially in case (d), the color of composite insulators is similar with the trees and the illumination contrast is low. It can be seen that the proposed method can successfully detect the composite insulator in the cluttered and complicated background.

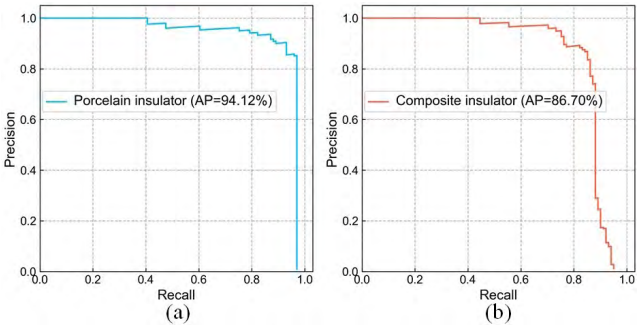

FIGURE 8. Precision-recall curve. (a) Porcelain insulator detection. (b) Composite insulator detection.

2) INSULATOR DETECTION IN BUILDING SCENES

To further verify the proposed method, we train the second specific insulator model to detect PI and CI in building scenes. In this training process, we reuse the same basic insulator model mentioned above, without repeating the first fine-tuning stage. In second fine-tuning stage, the basic insulator model is fine-tuned with the specific insulator dataset II.

The testing results are listed in Table 3 and the precision-recall curves are shown in Fig. 11. For the detection of PI, the precision is 91.67%, the recall is 88%, and the corresponding AP is 90.51%. For the detection of CI, the precision is 85%, the recall is 85%, and the corresponding AP is 87.29%.

The detection images are illustrated in Fig. 12 and Fig. 13. Compared with the forest scenes, the difference is not obvious between the characteristics of the two types of insulators and the backgrounds. The results show that the proposed model can yield good performance and accurately detect both the porcelain insulators and composite insulators from the complex building backgrounds. In addition, it demonstrates that the proposed method is applicable to detect different types of insulators for different scenes in a flexible way. The proposed method can easily extend to other types of insulators or other scenes by reusing the obtained basic insulator model. We only need to replace the specific insulator dataset and perform the second fine-tuning stage, which can save considerable computation time and improve the training efficiency.

D. ADVANTAGES OF TWO-STAGE FINE-TUNING STRATEGY

To demonstrate the importance of the two-stage fine-tuning strategy, a comparison was carried out by training the SSD model under four cases. In case 1 so-called one-stage fine-tuning with basic dataset, the model is trained with the COCO model and the basic insulator dataset. Actually, the model of case 1 is the basic insulator model. The training

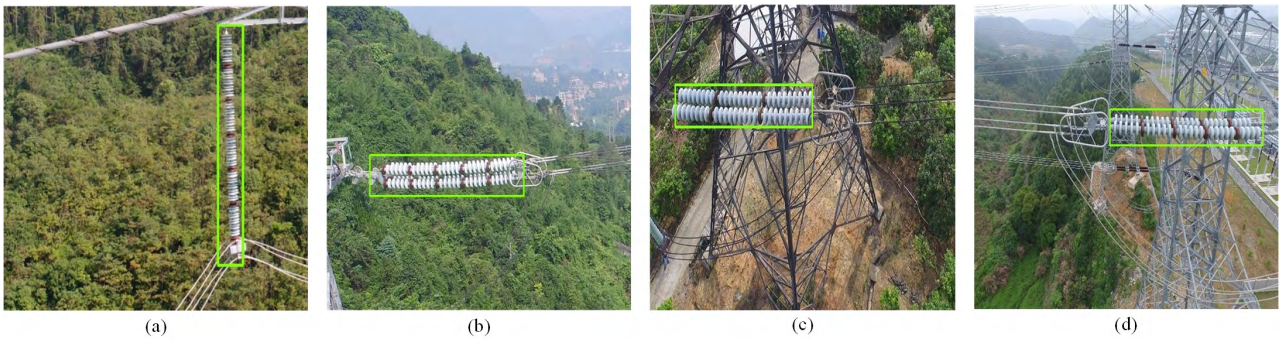

FIGURE 9. Detection examples for porcelain insulator in forest scenes.

FIGURE 10. Detection examples for composite insulator in forest scenes.

TABLE 3. Performance results of insulator detection in building scenes.

Insulator Type	TP	FP	FN	Precision(%)	Recall(%)	AP(%)	PC Time (ms)	TX2 Time (ms)
porcelain	88	8	12	91.67	88	90.51	22.92	62.75
composite	85	15	15	85	85	87.29	23.01	63.11

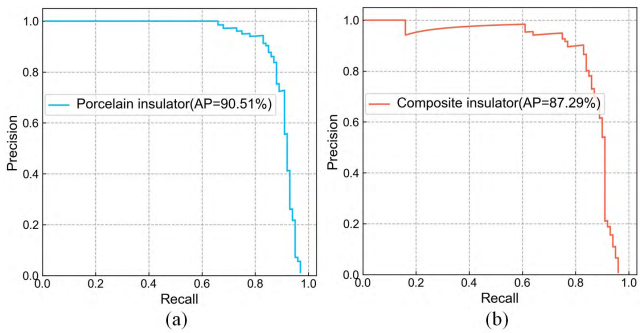

FIGURE 11. Precision-recall curve. (a) Porcelain insulator detection. (b) Composite insulator detection.

set of case 2 so-called one-stage fine-tuning with joint dataset consists of both the basic insulator dataset and the specific insulator dataset I. The model of case 2 is directly trained with the COCO model and the joint dataset. In case 3 so-called one-stage fine-tuning with specific dataset, we only use the COCO model and specific insulator dataset I to train the model. Case 4 is the proposed model using the proposed two-stage fine-tuning strategy. The same testing set was used. The obtained results are given in Table 4 and Fig. 14. As can be observed, the result of case 1 is an invalid model with low AP due to the big difference between the basic dataset and the test images of specific scenes. Case 2 still results in poor detection performance. The images in basic insulator dataset are diverse and may not represent the specific features, while the specific insulator dataset in the joint dataset has a low ratio owing to its small size. The results of case 3 have made some progress but still cannot meet the requirements of engineering application. The undesirable performance of case 3 is owing to the domain gap between the COCO dataset and the specific insulator dataset. Compared the case 3 with the proposed model, the proposed model gets higher precisions and recalls than those of case 3. Due to the great complexity of the detection of CI, AP is 64.25% when the transitional stage is not inserted. The AP of the proposed model exceeds that of case 3 by 22.45%. From the precision-recall curves, it can be observed that the precision of model without two-stage fine-tuning leads to high fluctuation when the recall increases. The results clearly reveal that the two-stage fine-tuning strategy can greatly enhance the accuracy and the robustness for insulator detection.

TABLE 4. Performance comparison of four training cases.

Model	Type	TP	FP	FN	Precision(%)	Recall(%)	AP(%)
Case 1	porcelain	38	35	62	52.05	38	30.62
	composite	18	55	82	24.66	18	4.5
Case 2	porcelain	60	31	40	65.93	60	67.09
	composite	40	102	60	28.17	40	23.15
Case 3	porcelain	79	12	21	86.81	79	81.28
	composite	75	20	25	78.95	75	64.25
Case 4	porcelain	90	6	10	93.75	90	94.12
	composite	87	15	13	85.29	87	86.7

FIGURE 12. Detection examples for porcelain insulator in building scenes.

FIGURE 13. Detection examples for composite insulator in building scenes.

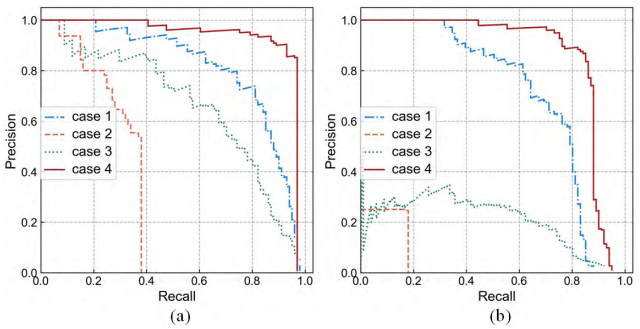

FIGURE 14. Precision-recall curves of four different training cases. (a) Porcelain insulator detection. (b) Composite insulator detection.

To test the effect of the size of the training data, the models were trained with the specific insulator dataset I by varying the sample size from 100 to 450. The precision-recall curves of training sets with different sample sizes are shown in Fig. 15. It can be seen that only 100 training images still can yield the AP of 89.16% and 74.21% for PI and CI, respectively. Although training set is insufficient, the total performance of the proposed method is robust and stable at a relatively high accuracy level. With two-stage fine-tuning strategy, the proposed method can effectively solve the small sample size problem of insulator detection.

E. COMPARISON WITH OTHER METHODS

The proposed SSD based approach is compared with the previously reported detection approaches in [8], [9], and [12] in terms of the precision and recall. We also compare the performance with other two object detection architectures: Faster-RCNN [15], [29] and R-FCN [16], [30]. The Faster-RCNN and R-FCN network were trained in a usual way with the same specific dataset. Table 5 summarizes the comparison results. It shows that our SSD based approach achieves a competitive precision and recall compared to the

TABLE 5. Comparison with other methods.

Method	Type	Feature	Precision(%)	Recall(%)	AP(%)
KNN [8]	porcelain	shape	33	98	-
MSMF Descriptor [9]	glass	shape	87	91	-
Semi-Local Operator [12]	glass	texture	85	86	-
Faster-RCNN+InceptionV2	porcelain	multi-level	71.09	91	79.44
	composite		61.68	60	44.47
R-FCN+ResNet101	porcelain	multi-level	81.48	88	86.98
	composite		80.56	87	82.88
The proposed method	porcelain	multi-level	93.75	90	94.12
	composite		85.29	87	86.70

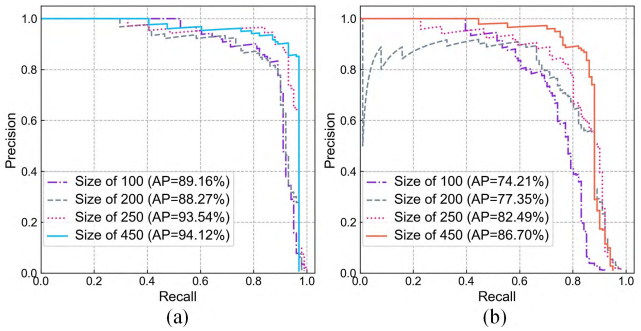

FIGURE 15. Precision-recall curves of different training data size. (a) Porcelain insulator detection. (b) Composite insulator detection.

previous work. Most previous methods focused on detecting the porcelain insulator and glass insulator, but not considered the composite insulator owing to the difficulty of locating this type of insulator in complicated backgrounds. Another advantage of our approach is that the multi-level features can be automatically extracted, and it can liberate human labor from manually extracting different features. The benefit is greatly attractive to the applications of UAV inspection of power transmission line. Compared with Faster-RCNN and R-FCN, the proposed model outperforms Faster-RCNN and R-FCN for both porcelain insulator and composite insulator. The proposed method gives a good trade-off between precision and recall and improves the robustness of the detector.

Once the model is well-trained, we can apply the model to the automatic autonomous UAV transmission line inspection system. The insulator detection method is one of main components and necessary precondition for smart insulator inspection. According to the application scenarios, the SSD model is utilized as a detection module embedded into the inspection system. In addition, applications of the detection model also can be processed on mobile and embedded devices such as TX2. Combined with the embedded device of UAV, the insulator can automatically be segmented and local enlarged to obtain more details of the insulator in the aerial images captured from the UAV in real-time. Based on the results of detection model, further inspection tasks such as status classifications of insulator will be performed.

IV. CONCLUSION

In this paper, we explore a novel insulator detection approach based on the deep convolutional neural networks for aerial images of UAVs in the power transmission line inspection. The SSD is utilized to implement the automatic feature learning process on the aerial image set. Instead of the inefficient and guideless hand-crafted feature extractors, the model trained by SSD can extract high-level features and improve the detection speed. To address the issues of insufficient training aerial images and domain gap, we introduce the two-stage fine-tuning strategy in the SSD training procedure. We first use the COCO model and the basic insulator dataset with general features of electricity domain to obtain the basic insulator model. Then, the second fine-tuning stage is employed to fine-tune the parameters of basic insulator model using the specific insulator dataset with more specific features of insulators. Once the final detection model is trained, it can be directly used to identify porcelain and composite insulators for aerial image. The results show that the porcelain and composite insulators can be accurately and quickly detected with the AP of 94.12% and 86.70% in the aerial images with forest backgrounds, respectively. Even in the complicated building backgrounds, AP of 90.51% for porcelain insulator and 87.29% for composite insulator can be achieved. The computation time is 23ms on PC and 63 ms on embedded device for a single image. The two-stage fine-tuning strategy can greatly enhance the generalization and robustness of the detection model. Compared with previously reported results, the SSD based method can realize automatic multi-level feature extraction and meanwhile give a competitive precision and recall. The proposed technique shows the potential in the real-time insulator detection for aerial images. Future work is needed to extend the technique to detect the fault of broken insulators for the further application of UAVs inspection of power transmission line.

REFERENCES

[1] Y. Song, H. Wang, and J. Zhang, "A vision-based broken strand detection method for a power-line maintenance robot," *IEEE Trans. Power Del.*, vol. 29, no. 5, pp. 2154–2161, Oct. 2014.

[2] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning," *Int. J. Elect. Power Energy Syst.*, vol. 99, pp. 107–120, Jul. 2018.

[3] 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.

[4] X. Zhang, J. An, and F. Chen, "A simple method of tempered glass insulator recognition from airborne image," in *Proc. Int. Conf. Optoelectron. Image Process. (ICOIP)*, vol. 1, Nov. 2010, pp. 127–130.

[5] M. J. B. Reddy, B. K. Chandra, and D. K. Mohanta, "A dost based approach for the condition monitoring of 11 kv distribution line insulators," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 18, no. 2, pp. 588–595, Apr. 2011.

[6] M. J. B. Reddy, B. K. Chandra, 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] Q. Wu, J. An, and B. Lin, "A texture segmentation algorithm based on PCA and global minimization active contour model for aerial insulator images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1509–1518, Oct. 2012.

[12] 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.

[13] W. Wang, Y. Wang, J. Han, and Y. Liu, "Recognition and drop-off detection of insulator based on aerial image," in *Proc. 9th Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2016, vol. 1, pp. 162–167.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

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

[17] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[18] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[19] L.-C. Chen, J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille, "Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4545–4554.

[20] M. Mostajabi, M. Maire, and G. Shakhnarovich, "Regularizing deep networks by modeling and predicting label structure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5629–5638.

[21] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[23] A. G. Howard *et al.* (2017). "MobileNets: Efficient convolutional neural networks for mobile vision applications." [Online]. Available: https://arxiv.org/abs/1704.04861

[24] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.

[25] G. Hinton, N. Srivastava, and K. Swersky, "Rmsprop: Divide the gradient by a running average of its recent magnitude," *Neural Netw. Mach. Learn.*, vol. 4, pp. 26–31, 2012.

[26] M. Abadi *et al.* (2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." [Online]. Available: https://arxiv.org/abs/1603.04467

[27] J. Huang *et al.* (2016). "Speed/accuracy trade-offs for modern convolutional object detectors." [Online]. Available: https://arxiv.org/abs/1611.10012

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

XIREN MIAO received the B.S. degree from Beihang University, Beijing, China, in 1986, and the M.S. and Ph.D. degrees from Fuzhou University, Fuzhou, China, in 1989 and 2000, respectively, where he is currently a Professor with the College of Electrical Engineering and Automation. His research interests include electrical and its system intelligent technology, and online monitoring and diagnosis of electrical equipment.

XINYU LIU received the B.S. degree from Fuzhou University, Fujian, China, in 2016, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include deep learning and condition monitoring of overhead transmission line.

JING CHEN received the B.S., M.S., and Ph.D. degrees from Xiamen University, Fujian, China, in 2010, 2013, and 2016, respectively. She is currently a Lecturer with the College of Electrical Engineering and Automation, Fuzhou University. Her research interests include intelligent fault diagnosis and artificial intelligence.

SHENGBIN ZHUANG received the B.S. degree from Fuzhou University, Fujian, China, in 2016, where he is currently pursuing the Ph.D. degree in power system and automation. His research interests include smart grids, online monitoring and diagnosis of electrical equipment, and the protection and control of ac/dc hybrid systems.

JIANWEI FAN received the B.S. degree from Fuzhou University, Fujian, China, in 2018, where he is currently pursuing the M.S. degree in electrical engineering. His research areas are mainly in the control systems of UAV and indoor positioning.

HAO JIANG received the B.S. and Ph.D. degrees from Xiamen University, Fujian, China, in 2008 and 2013, respectively. He is currently an Associate Professor with the College of Electrical Engineering and Automation, Fuzhou University. His research interests include artificial intelligence and machine learning.