Insulator Fault Detection in Aerial Images Based on Ensemble Learning With Multi-Level Perception

HAO JIANG, (Member, IEEE), **XIAOJIE QIU, JING CHEN, XINYU LIU,
XIREN MIAO, AND SHENGBIN ZHUANG**
College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

Corresponding author: Xinyu Liu (xinyu3307@163.com)

ABSTRACT Insulator fault in the transmission lines is the main factor of power transmission accident. The images captured from the aerial inspection can be utilized to detect the fault of insulators for further maintenance. For automatic transmission lines inspection system, the insulator fault detection is an interesting and challenging task due to the complex background and diversified insulators. In this paper, we propose a novel insulator fault detection method based on multi-level perception for aerial images. The multi-level perception is implemented by an ensemble architecture which combines three single-level perceptions. These single-level perceptions include the low level, middle level, and high level that are named by the attention to the insulator fault. They detect the insulator fault in the entire image, multi-insulator image, and single-insulator image, respectively. To address the filtering problem in the combination of three single-level perceptions, an ensemble method is proposed for generating the final results. For training the detection models employed in the multi-level perception, a powerful deep meta-architecture so-called single shot multibox detector (SSD) is utilized. The well-trained SSD models can automatically extract high quality features from aerial images instead of manually extracting features. By using the multi-level perception, the advantages of global and local information can achieve a favorable balance. Moreover, limited inspection images are fully utilized by the proposed method. Fault detection recall and precision of the proposed method are 93.69% and 91.23% testing in the practical inspection data, and insulator fault under various conditions can be correctly detected in the aerial images. The experimental results show that the proposed method can enhance the accuracy and robustness significantly.

INDEX TERMS Transmission lines inspection, insulator fault detection, ensemble learning, multi-level perception, deep learning.

I. INTRODUCTION

In transmission lines, the insulator is a widely used equipment with the dual function of electrical insulation and mechanical support. Due to long-term exposure to the natural environment, the insulator will occur various faults such as missing-cap, flashover and icing. The fault of insulators can seriously threaten the reliable operation of transmission lines and imperil the security of power system. The early fault

The associate editor coordinating the review of this manuscript and approving it for publication was Javier Medina.

detection of the insulator can reduce power cuts and prevent the huge economic losses and bad customer care in power transmission system. However, the transmission line covers a wide range of areas and the geographical conditions along the path are complex and diverse that causes difficulty on insulator inspection. Therefore, the fault detection of the insulator is one of the most important and difficult task during the inspection of the transmission line [1].

In recent years, instead of traditional manual patrol, UAV inspection develops into a common way of transmission lines inspection for intelligent requirements [2], [3]. The aerial
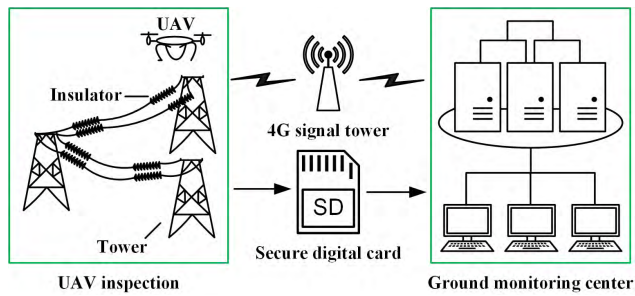
FIGURE 1. The aerial inspection system of power transmission lines.

inspection system consists of two parts: images acquisition by UAV and data analysis in the ground monitoring center, as shown in Fig. 1. The onboard camera captures the images of insulators according to the inspection planning. And the collected data for analysis should be sent to the ground monitoring center by 4G communication network or storage card. During the UAV inspection, a massive photographic record will be generated and can be used to detect the insulator fault that requires much manual labor. There is an urgent need for automatic detection of insulator fault in aerial images. However, the aerial images captured from UAVs contain the cluttered background and various types of insulators. There are two factors in the background: the scenes such as forest and building, and the electrical components including towers, fittings, wires etc. The scenes and electrical components are various for different inspection regions that leads to regional differences and data confidentiality. Therefore, each inspection region is relatively independent and few training data are available. In the available data, the images of insulator fault is much less than the normal data that makes against the model training. The external disturbing factors, such as the changing visual angle, different lighting, and partial occlusion, make it more difficult to detect the insulators and its fault. Hence, the fault detection of insulators in aerial images is a challenging task [2].

In the existing literatures, the works of insulator fault detection are relatively more difficult than insulator detection due to the small size of the faulty area. Oberweger *et al.* [4] introduced a circular descriptor and a noise-tolerant voting scheme for insulator detection, then the elliptical descriptor based on Local Outlier Factor (LOF) is utilized to analyze the fault of each insulator cap. Zhao *et al.* [5] applied OAD-BSPK algorithm [6] to detect the insulator and divided the insulator into multiple caps based on histogram valley detection. The 4096-dimensional features of each cap will be computed by deep convolutional neuron network for status classification. Wang *et al.* [7] introduced a multi-features method for insulator recognition that utilized shape, color and texture features. The region of insulator was rotated to horizontal configuration and then separated into 23 parts. GLCM was calculated in each part for abnormal analysis. These approaches divided an insulator into several caps [4], [5] or parts [7], but they did not consider the differently sized and partially visible insulator. Zhai *et al.* [8] proposed a saliency and adaptive

morphological algorithm which can fill the gaps between normal insulator caps. Therefore, the vacancy caused by missing-cap will be highlight, then the missing-cap can be located according to the distance between the caps. Such a method can only applied to glass insulator. In [9], the author improved the morphology based algorithm while considering both glass and ceramic insulators. However, these morphology based method [8], [9] will fail when there are multiple insulators in one image.

The insulator fault detection methods mentioned above detect the insulator first, and then analyze the condition in the extracted insulator region. They focus on the insulator region, and perform well on the extraction of fault features that can be defined as the local information. However, this ability is strongly affected by the results of insulator detection. Once the insulators are missed or incorrectly detected, it will greatly reduce the performance of insulator fault detection. Liu *et al.* [10] applied a detection network of deep learning for insulator fault detection. The Faster R-CNN was utilized to detect the fault directly on the entire image. Nordeng *et al.* [11] introduced a deep-learning-based method for detecting dead end body of insulator It also removed the procedure of insulator detection. These deep-learning-based methods [10], [11] unrestricted by the results of insulator detection. The models trained from entire images can learn rich global information that reduced the possibility of false detection. However, they show poor performance when dealing with the aerial image at a distant view.

All of the existing methods for insulator fault detection either focus on local information or global information only. The methods focusing on local information detect the fault in the image region of insulator. The methods concerning about global information detect the fault in the entire aerial image. When exploiting only the local information, the method will reach high recall but is powerfully affected by the performance of insulator detection. When pouring attention into the global information, the method will achieve high precision while losing performance to the detection of small objects. Until now, no studies that concentrate on taking advantage of both local and global information. Therefore, our goal is to proposed a well-balance method for insulator fault detection.

This paper focus on the missing-cap which is the most frequent fault of insulators. We attempt to achieve accuracy and fast fault detection in the aerial images for power transmission lines inspection system. Under the situation of complicated background and insufficient training data, we try to solve the issue of unbalance between the local and global information in the previously researches. A multi-level perception method based on ensemble learning is presented. This ensemble architecture consists of three single-level perceptions: low level perception, middle level perception and high level perception. They are named according to the attention of local information, and they detect the insulator fault in the entire image, multi-insulator image and single-insulator image respectively. To address the filtering problem in the combination of three single-level perceptions, an ensemble
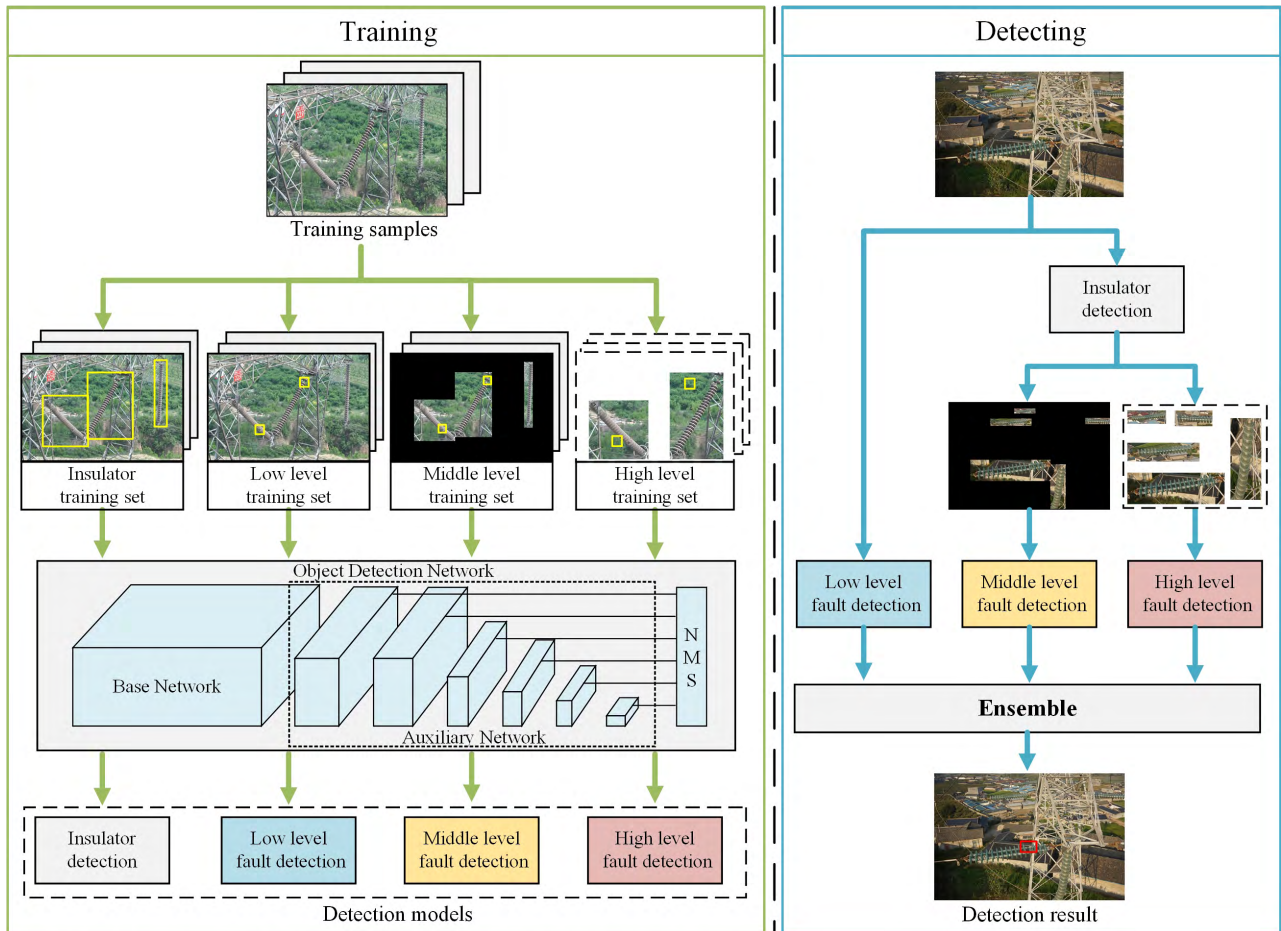
FIGURE 2. Schematic diagram of the proposed method for insulator fault detection.

method is proposed for generating the final results. By means of multi-level perception, the global and local information can be balanced processed together that improves the accuracy. It also makes full use of the limited inspection data. Meanwhile, the deep-learning based detection network using in the proposed method that guarantees the preferable detection accuracy and speed. Due to the big capacity of deep learning network, the proposed method also solves the fault detection problem of diversified insulators including different colors, shapes and sizes.

The following parts of this paper are organized as follows. We describe the proposed framework of insulator fault detection in Section II. Section III provides the experimental results and analysis. In section IV, we draw conclusions from the results and also discuss about our future work plans.

II. METHODOLOGY
A. FRAMEWORK OF INSULATOR FAULT DETECTION
The schematic diagram of the proposed insulator fault detection method in aerial images for transmission lines inspection is illustrated in Fig. 2. The proposed ensemble architecture of multi-level perception consists of three single-level

perceptions: low level, middle level and high level. They are named according to their attention of the insulator fault which is defined as local information in this paper. The low level perception looks at the entire image that can provide the strong global information. The middle level perception focus on the multi-insulator image, thereby most of the background can be filtered out. The high level perception looks at the image region of single-insulator, the attention of the model is concentrated on the local information. Theses single-level perceptions are increasingly concerned about the local information. To achieve single-level perception, the deep convolutional neural network for object detection is utilized. Four deep learning models should be trained individually: an insulator detection model and three insulator fault detection models. Thereby, four training sets need to be prepared: insulator training set, low level training set, middle level training set and high level training set. For accomplishing multi-level perception, three single-level perceptions should be combined based on ensemble learning. Once the results of the single-level perceptions are obtained, an ensemble approach is utilized to produce the final results of insulator fault. By using the multi-level perception, the global and local

information can be synthetically utilized for high quality fault detection.

B. MULTI-LEVEL PERCEPTION BASED ON ENSEMBLE LEARNING

In the application of transmission lines inspection, the electricity companies are only in charge of their local area. On the one hand, the background of aerial images, the photographic condition of the aerial inspection and the insulator characteristics are various according to different inspection regions. On the other hand, owing to the organizational independence and data confidentiality of each regional power grid, it is difficult to guarantee that the available training data is sufficient. Moreover, the fault data is much less than normal data in the available training data.

In order to detect the insulator fault in complex conditions and solve the issue of unbalance between the local and global information, the proposed approach presents a multi-level perception method using ensemble learning. Further, the multi-level perception is achieved by combining several single-level perceptions. Different single-level perceptions look at different receptive fields of the aerial image, as shown in Fig. 2. This paper introduces three single-level perceptions. They are defined as low level, middle level and high level perception according to the attention of insulator fault in aerial images. The ensemble of single-level perceptions merges the global and local information in a specific way, and makes full use of the limited inspection data. The local information here refers to the image features of insulator fault, and the image features other than this are defined as global information such as background.

1) IMPLEMENTATION OF LOW LEVEL PERCEPTION

The perception in low level is an end-to-end procedure, the input image is fed into the deep learning model and the insulator fault is directly located in the original image. It can be seen in Fig. 3, the region of insulator fault is quite small relative to the whole image. The image features of the insulator fault as the local information are difficult to learn in the detection model, while the global information such as the unique environments of transmission corridor is abundant. Therefore, the low level perception adapts in discriminating the background according to the whole image. It provides widespread general features of the electricity domain which ensure the completeness of the global information. Meanwhile, the low level perception pays less attention to the insulator fault in the image.

To achieve low level perception, a low level fault detection model should be trained. The low level training set contains unprocessed aerial images and the annotation is the insulator fault. The annotation refers to the coordinates and category of the bounding box which has the object.

2) IMPLEMENTATION OF MIDDLE LEVEL PERCEPTION

For middle level perception, a two-stage procedure is introduced as shown in Fig. 4. In stage 1, the insulators will be
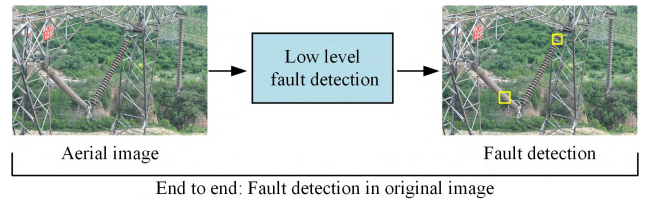

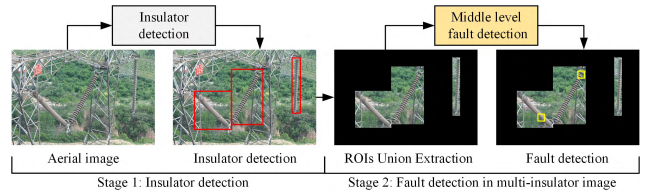**FIGURE 3.** The procedure of low level perception.

FIGURE 4. The procedure of middle level perception.

detected in the aerial image by using a deep learning detection model. After the detection of insulators, an operation of image processing named ROIs Union Extraction (RUE) is presented. We conduct the rectangular area of a single insulator as the Region of Interest (ROI) and extract the ROIs by covering the background region with zero matrix in 3 color channels. After RUE, only the rectangular regions of insulators will exist in the aerial image. Then in the stage 2, the multi-insulator image is fed into the fault detection model for locating the missing-cap of the insulator. In the middle level perception, most of the background is filtered out while the pixel size is maintained. Compared to low level perception, the middle level perception increases the attention to the insulator fault. And it simultaneously provides part of the global semantic information such as the shapes, sizes, amount and positional relation of insulators in the original aerial image. It plays a role of auxiliary for voting the final results of multi-level perception.

For realizing middle level perception, an insulator detection model and a middle level fault detection model should be trained. Two training sets are prepared for model training: insulator training set and middle level training set. The insulator training set contains aerial images without processing and the annotation is insulator. The middle level training set comprises images that are processed by RUE and the annotation is insulator fault.

3) IMPLEMENTATION OF HIGH LEVEL PERCEPTION

The high level perception also applies the two-stage procedure as shown in Fig. 5. The stage 1 of the high level perception is similar to middle level. After detecting the insulator in the aerial image, an operation of image processing is utilized to generate the input for next stage. Different from middle level perception which utilizes RUE, we adopt ROIs Extraction (RE) instead. The definition of ROI is the same as in middle level detection, we extract the ROIs by cropping the insulator region from the original aerial images. After
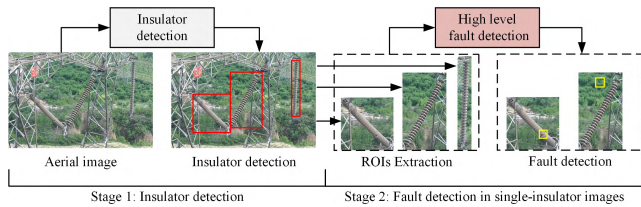
FIGURE 5. The procedure of high level perception.

FIGURE 6. Samples of collected aerial images including varied insulators with complex backgrounds.

RE, a set of images will produced according to the existing insulators in the original aerial image. The generated image is the rectangular area of the insulator with rare background. In general, there is only one insulator in one generated image. In the stage 2 of insulator fault detection, the single-insulator images are sequential fed into the fault detection model for detecting the insulator fault. The high level perception completely pour its attention into the insulator fault that discards the influence of background and the global semantic information. It provides complete and highly centralized features of the insulator fault.

To reach high level perception, two models should be trained: an insulator detection model and a high level fault detection model. Each model corresponds to a training set: insulator training set or high level training set. The insulator training set is the same as middle level perception. The high level training set includes images processed by RE and the annotation is insulator fault.

4) ENSEMBLE METHOD

To achieve multi-level perception, an ensemble learning approach is utilized to mix the results of single-level perceptions and generate the final results. An ensemble method called bounding box voting is introduced in [12]. It applied a standard non-max suppression (NMS) first, and the final bounding box was further refined by having each box to vote. But for two reasons, the bounding box voting cannot be directly applied in this paper. Firstly, few results will be preserved after applying NMS that makes against voting. Secondly and most importantly, this ensemble method retains the original confidence that makes difficulty for the determination of independent box. Therefore, we make some changes to the bounding box voting. Each single-level perception outputs a set of bounding box locations $B_p = \{B_{i,p}\}$ with confidence scores $s_p = \{s_{i,p}\}$. The $p \in \{L, M, H\}$ refers to the index of single-level perception. For i-th result of low level perception, the final bounding box coordinates $B'_{i,L}$ are further refined by:

$$B'_{i,L} = \frac{w_L B_{i,L} + w_M B_{\max,M} + w_H B_{\max,H}}{\sum w_p} \quad (1)$$

where the w_p refers to the weight of each single-level perception which is a constant. The $B_{\max,p}$ denotes the box in p level that has the maximal IoU with $B_{i,L}$. If the IoU between $B_{\max,p}$ and $B_{i,L}$ is less than 0.5, setting the corresponding w_p to zero.

Similarly, the confidence score $s'_{i,L}$ can be computed by:

$$s'_{i,L} = \frac{w_L s_{i,L} + w_M s_{\max,M} + w_H s_{\max,H}}{\sum w_p} \quad (2)$$

If there is no matched box with IoU over than 0.5, setting the $s_{\max,p}$ to zero while keeping the w_p. The results processing of other two levels are similar to the low level perception. Notice that we remove the NMS in the proposed ensemble method and each box only participates in the calculation once.

C. IMAGE PREPROCESSING

The aerial images are captured by UAVs in the inspection task from different regions. The collected images contain variety insulators with various complex backgrounds. The backgrounds not only have unique environments in the electricity domain, but also have diverse components of transmission lines that besides insulators. Each image from these collected images contains at least one insulator and one insulator fault. The samples of collected images are shown in Fig. 6, the types of insulators include the porcelain and glass insulator with devised colors and shapes. It can be seen that the backgrounds and the photographic conditions are varied.

In the proposed method, there are four kinds of training sets: insulator training set, low level training set, middle level training set and high level training set. And two kinds of labeled classes: insulator and insulator fault. Fig. 7 shows two samples of image preprocessing for the generation of training sets. It can be seen in Fig. 7 (a), the annotation of insulator training set is the category and location of insulator in the original image. For low level training set Fig. 7 (b), the annotation is the insulator fault in the original image. The training data for the middle level perception is defined as middle level training set. The images of middle level training set are produced from collected images by extracting the ROIs using zero covering. As shown in Fig. 7 (c), the background is dyed black and the rectangular area of insulators are highlighted. Notice that the pixel size of middle level training image is the same with the original image. For middle level training set, the annotation is the insulator fault in the image generated
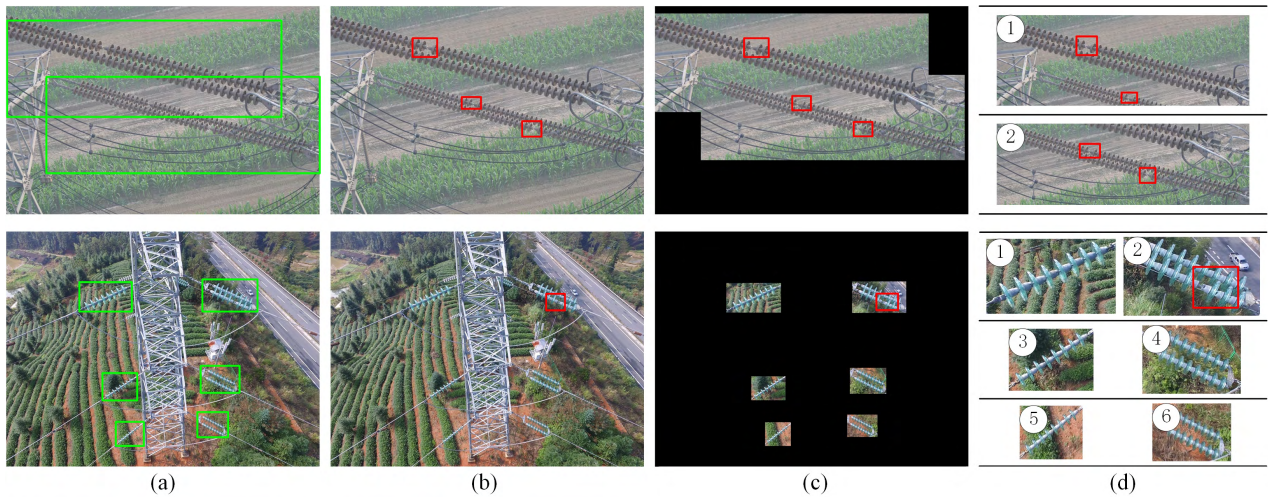
FIGURE 7. Image preprocessing for model training. The green box denotes the insulator and the red box refers to the insulator fault. (a) Insulator training set. (b) Low level training set. (c) Middle level training set generated by ROIs Union Extraction (RUE). (d) High level training set generated by ROIs Extraction (RE).

by RUE. The training data for the high level perception is defined as high level training set that is also generated from collected images. Each image of high level training set contains an insulator with the insulator fault which is cropped from the original image. The samples can be seen in Fig. 7 (d), the background near insulator is tight and the image size is various according to the insulator shape. In special cases, as shown in the first sample, an insulator fault will be reserved in two generated images. The ROIs extraction generated six images in the second sample, but only the one with insulator fault will be added to the training set.

D. CONVOLUTIONAL NEURAL NETWORK FOR OBJECT DETECTION

Convolutional Neural network achieves great success in many image applications including object detection [13]. In this paper, we choose Single Shot multibox Detector [14] as the meta-architecture of object detection that based on the convolutional neural network. The SSD composed by base network and auxiliary network as shown in Fig. 8. In training process, the input data are fed into the network to compute the loss and then updated the parameters using back-propagation. When detecting, only the forward operation is performed to infer the detection result.

1) BASE NETWORK

The base network is utilized to extract high quality features of the images. There are a host of standard convolutional neural networks can be served as base network that achieve promising results in the ImageNet benchmark [15] such as AlexNet [16], VGG [17], GoogleNet [18] and MobileNet [19]. With the view to the efficiency of the transmission lines inspection and the possibility of embedded application in UAV, this paper chooses the MobileNet as the base network. The MobileNet is a lightweight convolutional neural network which only has 4.2 million parameters compared with VGG16 (138 million parameters). In the ImageNet benchmark, the MobileNet achieves 70.6% mAP that nearly as accurate as VGG16(71.5%).

2) AUXILIARY NETWORK

After base network, the SSD appends an auxiliary network in order to produce detections of the object. The auxiliary network includes a set of convolutional layers that decrease in size progressively for predicting multi-scale objects. As shown in Fig. 8, the feature maps generated by the auxiliary network descend in height and width by utilizing the 2 stride of convolutional operation. Cells of each feature map corresponding to a window in the input image. Small feature map in deep layer has lager receptive field than big feature map in shallow layer. Therefore, the auxiliary network makes the network has the ability to detect the object of multiple scales and improves the accuracy of detection.

A fixed set of default bounding boxes are generated by feature maps. The size of the default bounding boxes depends on the input size (W, H), scale s_k and aspect ratio a_r where the (W, H) is width and height of the input image, s_k meaning the k-th layer has the scale of s_k and $a_r \in \{1, 2, 3, 1/2, 1/3\}$. Then the size of default boxes (W_d, H_d) can be computed as:

$$W_d = W s_k \sqrt{a_r}, \quad H_d = H s_k \sqrt{a_r} \quad (3)$$

Each default box predicts the shape offsets (cx, cy, w, h) of location and the class confidences for p object categories $(c_1, c_2, \ldots, c_p, c_{background})$. The object categories in this paper is insulator missing-cap fault and background. Fig. 9 shows an example of default box generation for one cell in a feature map.

In the training process, two loss functions are utilized to compute the loss between predicted label and ground-truth label, they are Softmax loss function L_{conf} which computes
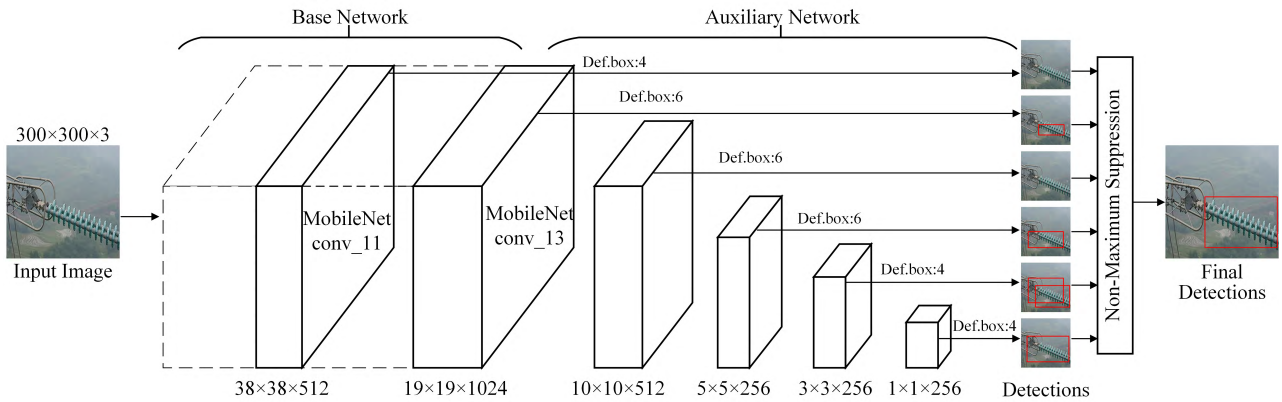
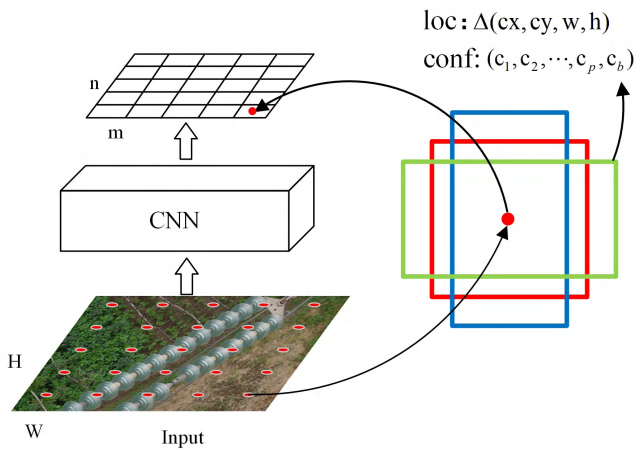**FIGURE 8.** The network architecture of SSD. Def. boxes: Default boxes.

FIGURE 9. Generation of the default boxes for one cell from the feature map.

the class confidence and SmoothL1 loss function L_{loc} for location. The predicting label has positive label and negative label according to the ground-truth box. Suppose predicted label $x_{ij}^p = \{0, 1\}$ is an indicator for matching the i-th default box to the j-th ground truth box of category p. When the Intersection-over-Union (IoU) between default box and ground truth box is higher than a threshold(0.5), the $x_{ij}^p = 1$ which means the positive label.

Once the predicted label is defined, the Softmax loss function L_{conf} can be described as:

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

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

where c refers to confidence and N is the number of predicted boxes, the c_i^0 meaning 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. The SmoothL1 loss

function for location is:

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

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

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

where l refers to the predicted box, g meaning the ground-truth box and d refers to the default box, 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). Note that the predicted box and the default box are corresponding one by one. The $SmoothL1$ is denoted as:

$$Smooth_{L1}(X) = \begin{cases} 0.5(X)^2 & if \ |X| < 1 \\ |X| - 0.5 & otherwise \end{cases} \quad (9)$$

$$where \quad X = l_i^m - \hat{g}_j^m \quad (10)$$

The total loss L of the object detection network can be compute as:

$$L(x, c, l, g) = \frac{1}{N}\left(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)\right) \quad (11)$$

After predicting the bounding boxes that possible contain the target object, Non-Maximum Suppression [20] is utilized for preliminary screening. Then, the predicting boxes are sorted according to class confidence. Finally, detection results are selected from the predicting boxes on the basis of confidence threshold, this paper sets the threshold to 0.5.

III. EXPERIMENTAL RESULTS
In this section, the proposed method is evaluated on the aerial image from UAV inspection of power transmission lines. The characteristics of the experiment are illustrated first including dataset description, evaluation method and experiment setup. Then, we analyze the performance of single-level perception

TABLE 1. Description of datasets.

Dataset	Images	objects
Insulator training set	385	601
Low level training set	385	413
Middle level training set	385	413
High level training set	387	414
Insulator testing set	100	160
Insulator fault testing set	100	111

and multi-level perception. In addition, the effects of ensemble mode and ensemble method for multi-level perception are discussed.

A. EXPERIMENT DESCRIPTION

1) DATA PREPARATION

All the aerial images are provided by China Power Grid from the UAVs for power transmission lines inspection. Notice that the aerial images are obtained in real-word and we do not perform any data augmentation operations for data preparation. The data augmentation only exits in the training process. Due to the individual photographic condition of each inspection region, the images have various pixel sizes from 1920×1080 to 9800×6500. There are at least one insulator and one insulator fault in a collected image. In other words, some inspection images have multiple normal insulators and multiple insulator faults.

In this paper, four sets are used for training: insulator training set, low level training set, middle level training set and high level training set. All training sets are generated from the collected images. For testing, two testing sets are prepared: insulator testing set and insulator fault testing set. The images in testing set are randomly selected and distinct from training data. The description of training and testing sets are shown in Table. 1. Objects refer to labeled boxes of the whole training set or testing set. In order to train the detection models in multi-level perception, insulator, low level, middle level and high level training set are generated from 385 collected images. The insulator training set has 601 labeled insulators. Low level and middle level training sets both have 413 labeled insulator faults. Notice that the high level training set has two additional training images, it is because some collected images have more than one faulty insulator. Thereby, multiple images will be generated by RE for a collected image. In special cases, as shown in Fig. 7, an insulator fault will be reserved in two generated images. Therefore, the high level training set has one more object than low level dataset and middle level dataset. The testing sets for insulator and insulator fault are generated from 100 collected images. They have 160 insulators and 111 insulator faults respectively.

2) IMPLEMENTATION

The proposed method is implemented in TensorFlow [21] using Object Detection API [22]. All the detection models

are running on a server computer with an environment of Ubuntu 18.04.1 LTS, an Intel Core i7-6850K CPU, four NVIDIA GeForce GTX 1080 Ti GPUs under 9.0 CUDA with 7.0 cuDNN version, and 32GB of RAM memory. The parameters of the convolutional neural network for object detection 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 and the IoU threshold is 0.5. Here, the momentum optimizer is used to train the model [23]. The momentum is 0.9 and batch size is 64. We use the warm up learning rate that initialized with 0.013 and grow to 0.04 at 2000 training steps, then decay to zero at the end of training process. 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 [24]. Max training step is set to be 20000. Due to limited training data, the data augmentation technology is utilized to increase the diversity of samples. In training process, three data augmentation modes are randomly utilized at each training step: gray, crop and rotate. The confidence threshold is set to 0.2 for keeping more detection results. For ensemble method, the weight w_p of each single-level perception is set to 1. To evaluate the detection performance, four metrics are applied including recall, precision, average precision (AP) and validation value (VAL). The first three metrics are commonly used for object detection. The validation value is a comprehensive metric for inspection task in transmission lines that is computed by:

$$VAL = 0.5 * AP + 0.3 * Rcall + 0.2 * Precision \quad (12)$$

Different weights represent the different emphasis of the electrical company on recall, precision and average precision. Compared to the precision, the electrical company cares more about the recall. This is because the false positives can be easily excluded in the limited images while the false negatives need to look through all the inspection data for manual check.

B. DETECTION RESULTS

1) INSULATOR DETECTION

The insulator will be detected in middle level and high level perceptions. We use the insulator training set to train the insulator detection model. The well-trained model can detect various types of insulators in aerial images from a complex background. The testing set is composed of 100 aerial images with 160 labeled insulators. The performance results of insulator detection can be seen in Table. 2. The recall achieves 96.95%, it indicates that most of the insulators are correctly detected.

Fig. 10 illustrates some testing results of insulator detection. Case (a) to case (d) show the results of glass insulator in the sky, field, forest and building scenes respectively. The glass insulators in these samples have light green color that is similar to the background color of the forest and field. The photographic condition also is an influenced factor to the

TABLE 2. Performance results of insulator detection.

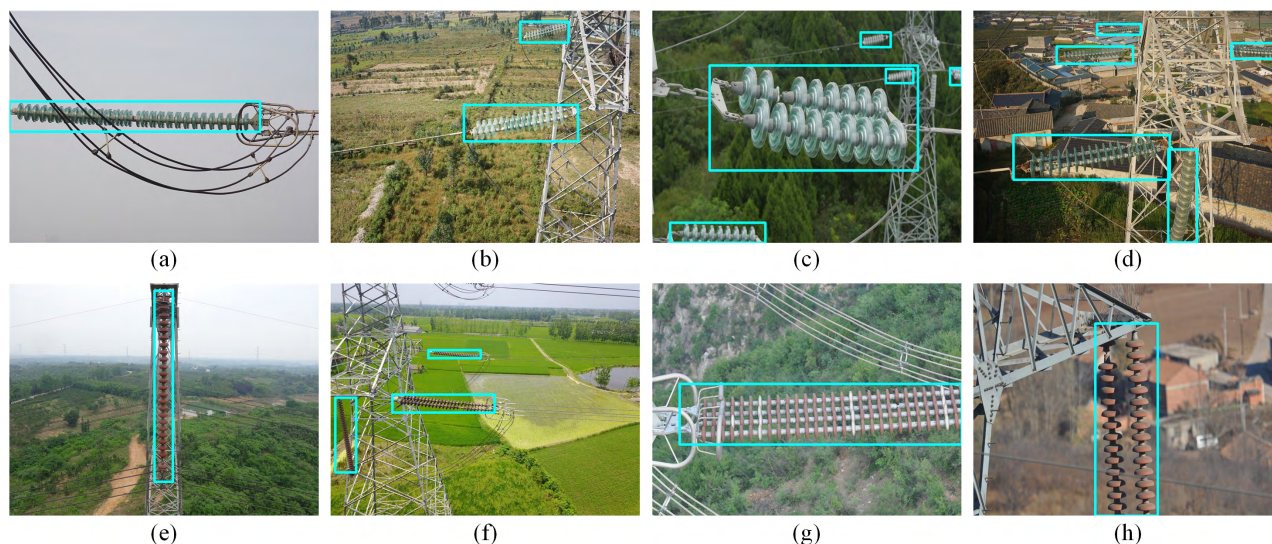| Model	Performance				
	R(%)	**P(%)**	**AP(%)**	**VAL(%)**	**Time (ms)**
Insulator detection	96.95	84.13	92.26	92.04	29.5

FIGURE 10. Detection samples of insulators in complex background. (a)–(d) Glass insulator. (e)–(h) Porcelain insulator.

detection performance especially in case (d). This image was taken at dusk. Hence, the weak light and complex background with tower and buildings that increase the difficulty of detection. The size of insulator is another factor that influences the detection performance. In case (c) and case (d), the insulators in the distance have small size and may be blurred due to focusing of camera. These samples show that the insulator detection model can handle the problems mentioned above and correctly locate the glass insulators. Similarly, case (e) to (h) demonstrate the detection results of porcelain insulator. Porcelain insulators have longer length and darker color in these samples. In case (g), the insulator has four strings and the length almost covers the width of the image. Inspection data with more than three insulator strings is relatively rare compared to one or two strings. The insulator in case (h) has similar color to the buildings and soil of the background. And the light and shadow also make challenges to the insulator detection. It can be seen that the insulator detection model can successfully detect the glass and porcelain insulators in the cluttered and complicated background under different scenes.

2) PERFORMANCE RESULTS OF SINGLE-LEVEL PERCEPTION
The obtained performances of three single-level perceptions are given in Table. 3. For low level perception which is an end to end procedure, the recall is 83.78% which is the lowest. However, it reaches highest precision with 94.90%. This illustrates that the low level perception pays more attention to

TABLE 3. Performance results of single-level perceptions.

Model	Performance				
	R(%)	**P(%)**	**AP(%)**	**VAL(%)**	**Time (ms)**
Low-level	83.78	94.90	83.55	85.89	29.5
Middle-level	85.59	92.23	85.02	86.63	60.4
High-level	90.99	84.87	89.05	88.80	86.1

the global information of background, while dilutes the local information of insulator fault. Thereby, it has the potential to treat the insulator fault as the background, resulting in a lower recall. The performance of middle level perception is a little better than low level perception. The recall is increased to 85.59% while the precision is kept above 90%. Finally, the validation value is improved by 0.74%. It demonstrates that the middle level perception preserves some of the global information while increasing the attention to the insulator fault. Thus, the middle level perception sacrifices a certain precision in exchange for the equivalent increase of the recall. For high level perception, the recall is increased to 90.99% and the precision is dropped to 84.87%. It indicates that the high level perception pays all of its attention to the insulator fault. It receives the highest recall and validation value, but this is exchanged with the lowest precision.

From low level to high level, the recall keeps on increasing and the precision keeps on descending. It demonstrates that the more attention the model pays to local information, the stronger the ability to identify the insulator fault.
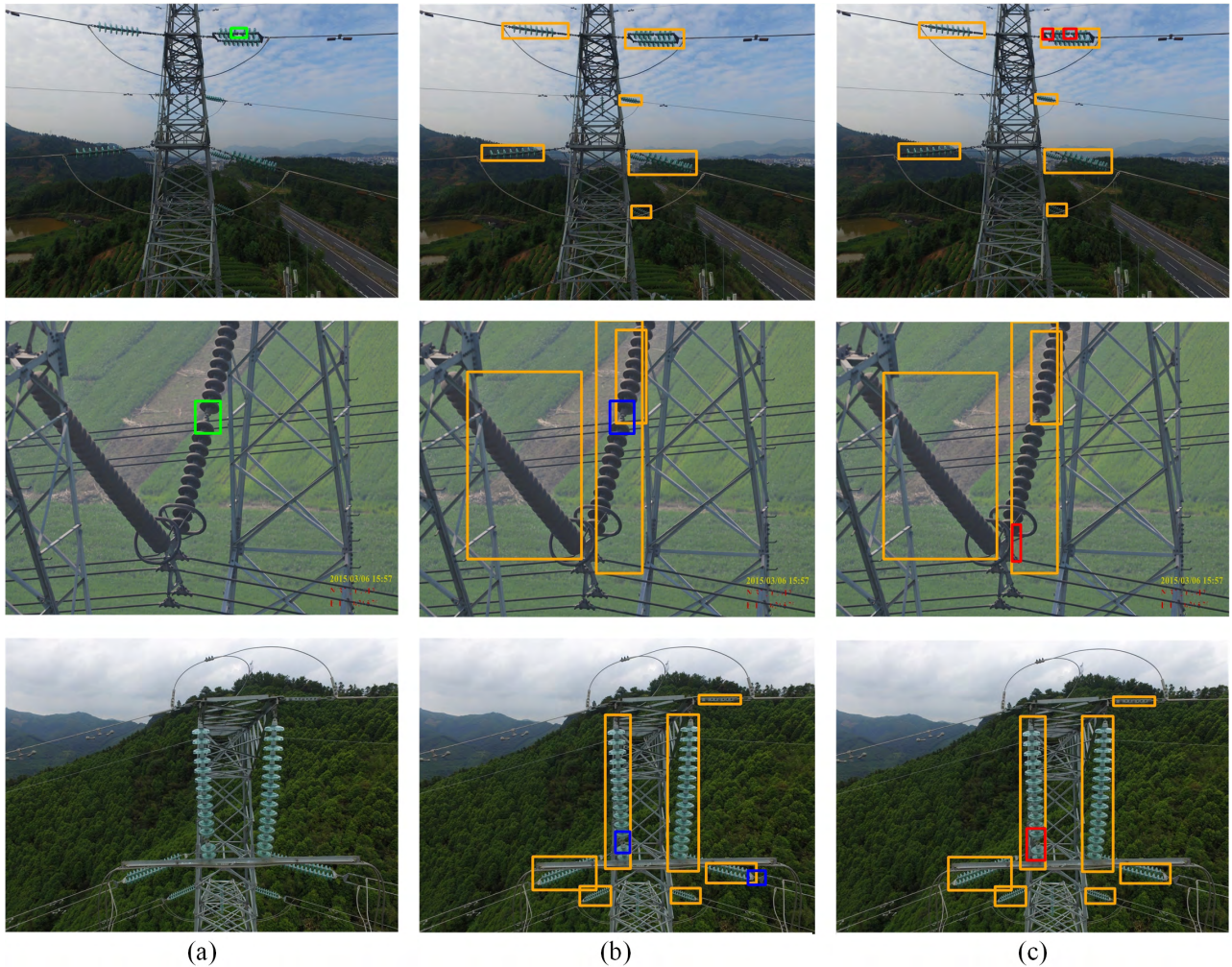
FIGURE 11. Detection samples of three single-level perceptions. (a) Low level. (b) Middle level. (c) High level. The insulator is bounded by the orange box. The green, blue, and red boxes denote the results of low level, middle level, and high level respectively.

Meanwhile, the reduction of global information leads to confusion between the background and the insulator fault. Hence, this transfer of attention results in more and more true positives and false positives of insulator fault.

The time refers to the computation time of single-level perception in one image. Notice that the inference time does not include the reading time of the input image. Due to the same detection network which is utilized in this paper, each well-trained model has the similar inference time with 30 ms. The low level perception applies one model on one image. Thereby, the computation time is 29.5 ms. The middle level perception employs two models on two images respectively. The detection time is 60.4 ms. The high level utilizes two models on an image and multiple ROIs. Therefore, the computation time is 86.1 ms that is more than twice the time of the single model.

Three detection samples of single-level perceptions are shown in Fig. 11. According to the implementation of the single-level perception, the results of low level perception

mark the insulator faults while others mark the insulators and their fault. In the first sample, the fault occurred on the insulator in the upper right corner. The low level perception (a) and high level perception (c) correctly detect the insulator fault while the middle level perception (b) failed. Notice that the high level perception also has a false identification. The second sample has two insulators and a fault in the right. As can be seen, the low level perception and middle level perception correctly detect the insulator fault. The high level perception fails to detect the fault and has a false detection. The low level perception cannot identify the fault in third sample while middle level perception and high level perception achieve successful detection. These detection samples illustrate that the single-level perceptions can be mutually verified and supplemented.

To further discuss, the insulator detection is a partial procedure of middle level perception or high level perception. The performance of insulator detection will affect the performances of these two single-level perceptions. In this situation,

TABLE 4. Performance results of multi-level perception.

Model	Performance				
	R(%)	P(%)	AP(%)	VAL(%)	Time (ms)
Insulator detection	93.69	91.23	92.26	92.48	145.7

the low level perception is a pleasurable supplement because it is unaffected by insulator detection.

3) PERFORMANCE RESULTS OF MULTI-LEVEL PERCEPTION

The multi-level perception of insulator fault is implemented by the ensemble learning method. Performance results of multi-level perception are illustrated in Table. 4. All accuracy metrics of multi-level perception are higher than the accuracy metrics of any single-level perception. The recall and precision are 93.69% and 91.23% respectively. The average precision with 92.26% demonstrates the reliability of the algorithm. Finally, the validation value achieves 92.48% which is at least 3.68% higher than the single-level perception. The computation time with 145.7 ms is the summation of three single-level perception. It indicates that the performance of the proposed method meets the requirements of the electrical company for accuracy and speed.

The comparison between single-level perceptions and multi-level perception can be seen in Fig. 12. Results of different perceptions are marked by different colors. The green, blue and red boxes denote the results of low level, middle level and high level respectively. For readers' convenience, a small tag with the same color is added above each box. Results of multi-level perception are bounded by the yellow box. The tag for intuitive display is added below the box.

The detection samples of multi-level perception can be seen in Fig. 12. The samples (a), (b) and (c) show the scenes of forest, town and desert respectively. Three single-level perceptions correctly detect the insulator fault and the multi-level perception well preserved these positive results. In samples (d), (e) and (f) that are also shown in Fig. 12, the negative results are filtered out by multi-level perception. Moreover, the samples (g), (h) and (j) demonstrate the filtration capacity of multi-level perception. The sample (j) is special due to the parallel faults of the insulator. The low level perception generates a false detection between the parallel faults. It is difficult to identify which box is accurate even for human. However, the multi-level perception successfully distinguishes the correct boxes. There are three insulator faults in sample (k), it can be seen that only high level perception can detect all the faults. The middle level perception merely detect the central one of these three faults while low level perception does not detect anything. Under this hard situation, the mutual verification between single-level perceptions is problematic. However, the proposed method preserves the correct boxes and eliminates the wrong detecting result. The sample (l) has six false identifications of single-level perceptions. The proposed ensemble method filters out half of them.

TABLE 5. Performance results of multi-level perception in different ensemble modes.

Ensemble mode	Performance			
	R(%)	P(%)	AP(%)	VAL(%)
Low+Middle	86.49	93.20	86.22	87.70
Low+High	93.69	88.14	92.31	91.89
Middle+High	93.69	85.95	92.19	91.39
Low+Middle+High	93.69	91.23	92.26	92.48

TABLE 6. Performance results of multi-level perception in different ensemble methods.

Ensemble method	Performance			
	R(%)	P(%)	AP(%)	VAL(%)
Concatenate (Baseline)	94.59	16.41	27.07	45.19
Bounding Box Voting [13]	94.59	80.15	93.48	91.15
This paper	93.69	91.23	92.26	92.48

4) EFFECT OF ENSEMBLE LEARNING FOR MULTI-LEVEL PERCEPTION

In order to discuss the effects of different ensemble modes for multi-level perception, we conduct experiments with four ensemble modes and compare the fault detection results on the same testing set. The ensemble of low level and middle level perceptions achieves highest precision with 93.20% and finally reaches 87.70% VAL. After replacing the middle level perception with high level perception in the ensemble mode, the precision is declined to 88.14% while the recall is increased by 7.2%. In addition, the comprehensive metric VAL is finally increased by 4.19%. The ensemble of high level and middle level perceptions decreases the precision to 85.95%, but it keeps on promoting with 91.39% VAL. Finally, the ensemble mode of three single-level perceptions achieved highest VAL with 92.48%.

Such promotion illustrates the contributions of different single-level perceptions. Low level perception provides rich global information for high precision and the high level perception supplies strong local information for high recall. However, these two single-level perceptions are excessively extreme. Thereby, they need an unbiased single-level perception to voting the final results for multi-level perception.

To evaluate the advantage of the proposed ensemble method quantitatively, we compare it with the ensemble method introduced in [12]. The ensemble method of concatenate is the baseline in this experiment. It refers to the merging of the results of three single-level perceptions without

FIGURE 12. Comparison between single-level perceptions and multi-level perception. The green, blue, and red boxes denote the results of low level, middle level, and high level respectively. Results of multi-level perception are bounded by the yellow box.

any processing. As can be seen in Table. 6, the concatenate shows poor performance in term of precision and average precision. This is not surprising since it retains all detection results unconditionally, and the evaluation metrics treat the results beside matching result as the false positives. Therefore, there will be a set of false positives when computing the evaluation metrics of concatenate. When applying the bounding box voting, the precision and average precision are greatly improved. But many false positives still preserved in the results of bounding box voting. There are two reasons behind this phenomenon. In the first place, after applying NMS, the number of detection results is small. It leads to the rare detection results for voting. Secondly, there is no processing of confidence in the bounding box voting. This is disadvantageous for the determination of the independent

box. The independent box means only one detector estimated there may exit the object in the box. The proposed ensemble method can solve the issues mentioned above. Hence, the precision is increased by 11.08% while merely a little recall is reduced.

C. COMPARISON WITH OTHER METHODS

The proposed ensemble architecture is compared with other four competitive methods: Faster R-CNN [25], SSD [14], R-FCN [26] and YOLOv3 [27]. The base networks of Faster R-CNN, SSD and R-FCN are ResNet-101 [28]. References [10] and [11] use deep-learning-based detection network for monitoring the condition of insulators in the entire image. Therefore, in the comparative experiment, only the missing-cap area is labeled and located. We directly detect

TABLE 7. Comparison with other methods.

Method	Performance			
	R(%)	P(%)	AP(%)	VAL(%)
Faster-RCNN [26]	85.59	90.48	85.45	86.50
SSD [15]	82.88	93.88	82.53	84.91
R-FCN [27]	80.18	93.68	80.08	82.83
YOLOv3 [28]	67.57	93.75	66.50	72.27
The proposed method	93.69	91.23	92.26	92.48

insulator faults on aerial images that this procedure is similar to the low level perception defined in this paper. For model training, the training sets are generated from the collected inspection images that are utilized in the proposed method. Table. 7 summarizes the comparison results. It can be seen that our proposed method achieves the best fault detection accuracy and shows a good trade-off between precision and recall. YOLOv3 reaches precision with 93.75% but shows poor performance in terms of other metrics. The recall and average precision of SSD and R-FCN are much better than YOLOv3 while maintaining high precision. The performance of Faster R-CNN is the best in the four competitive methods, but there is still a gap to the proposed method. Throughout these comparison methods, they all achieves high precision but performed poorly on the recall. This is not surprising since it is difficult to distinguish the 50×50-pixel objects in at least 1920×1080-pixel raw input image. Owing to the downsampling of the pooling layers in CNN, most of the semantic information of insulator fault will be replaced by the background. This will causes the fault area to be easily missed detected.

IV. CONCLUSION

In this paper, we explore a novel insulator fault detection approach based on ensemble learning with multi-level perception for aerial images of UAVs in the power transmission lines inspection. To address the issue of the trade-off problem between local and global information, we introduce the multi-level perception based on ensemble learning. The proposed method is implemented by the ensemble of three single-level perceptions. The low level perception looks for missing-cap fault through the entire image using the rich global information which is the guarantee of high precision. The middle level perception detects the fault on the multi-insulator image using the enhanced local information and partial global information which is the unbiased auxiliary for results voting. The high level perception individually detects the insulator fault on each single-insulator image using the strong local information which ensures the high recall. To remedy the filtration problem in the ensemble procedure, an ensemble method inspired by voting machine is proposed. For training the detection models in multi-level perception, the deep learning detection network SSD is utilized to implement the automatic feature learning process on

the aerial image set. The model trained by SSD can extract high quality features for classification and location with little consuming time. The proposed ensemble architecture can significantly contribute to a new horizon of machine learning for the insulator fault detection using aerial images. The results show that the missing-cap fault of insulators can be accurately and quickly detected with the recall of 93.69%, precision of 91.23%, AP of 92.26% and VAL of 92.48%. The computation time is 145.7ms on PC for a single image. Experimental results show that the proposed method meets the engineering requirements of off-line analysis for transmission lines inspection. Future works are needed to combining more single-level perceptions and reduce the computation cost.

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

[5] 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.

[6] 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.

[7] 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. Des. (ISCID)*, vol. 1, Dec. 2016, pp. 162–167.

[8] 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.

[9] Y. Zhai, R. Chen, Q. Yang, X. Li, and Z. Zhao, "Insulator fault detection based on spatial morphological features of aerial images," *IEEE Access*, vol. 6, pp. 35316–35326, 2018.

[10] X. Liu, H. Jiang, J. Chen, J. Chen, S. Zhuang, and X. Miao, "Insulator detection in aerial images based on faster regions with convolutional neural network," in *Proc. IEEE 14th Int. Conf. Control Automat. (ICCA)*, Jun. 2018, pp. 1082–1086.

[11] I. E. Nordeng, A. Hasan, D. Olsen, and J. Neubert, "DEBC detection with deep learning," in *Proc. Scand. Conf. Image Anal.* Cham, Switzerland: Springer, 2017, pp. 248–259.

[12] S. Gidaris and N. Komodakis, "Object detection via a multi-region and semantic segmentation-aware CNN model," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1134–1142.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

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

[15] 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.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

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

[18] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

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

[20] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3, Aug. 2006, pp. 850–855.

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

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

[23] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. Int. Conf. Mach. Learn.*, Feb. 2013, pp. 1139–1147.

[24] 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.

[25] 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.

[26] 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.

[27] J. Redmon and A. Farhadi. (2018). "YOLOv3: An incremental improvement." [Online]. Available: https://arxiv.org/abs/1804.02767

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

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.

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 the deep learning and condition monitoring of overhead transmission line.

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, on-line monitoring, and diagnosis of electrical equipment.

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.

XIAOJIE QIU received the B.S. degree from Fujian Agriculture and Forestry University, Fujian, China, in 2017. He is currently pursuing the M.S. degree in electrical engineering with Fuzhou University. His research interests include machine learning and computer vision.

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 grid, on-line monitoring and diagnosis of electrical equipment, and protection and control of AC/DC hybrid systems.

• • •