IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR ELECTRIC POWER SYSTEMS

Received September 19, 2019, accepted October 20, 2019, date of publication October 29, 2019,
date of current version November 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950053

Research on Recognition Method of Electrical Components Based on YOLO V3

HAIPENG CHEN[1,2], ZHENTAO HE[1,2], BOWEN SHI[2], AND TIE ZHONG[1,2]

[1]Key Laboratory of Modern Power System Simulation and Control and Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China
[2]Department of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China

Corresponding author: Tie Zhong (519647817@qq.com)

ABSTRACT The reliability of electrical components affects the stable operation of the power system. Electrical components inspection has long been important issues in the intelligent power system. The main problems of traditional recognition methods of electrical components are low detection accuracy and poor real-time performance, which are challenging to extract necessary features from the inspection images. This paper proposes a way to detect the electrical components in the Unmanned Aerial Vehicle (UAV) inspection image based on You Only Look Once (YOLO) V3 algorithm. Due to some of the inspection images are not clear, which result in the reduction of the available dataset. On this basis, we adopt Super-Resolution Convolutional Neural Network (SRCNN) to realize super-resolution reconstruction on the blurred image, which achieves the expansion of the dataset. We compare the performance of the proposed method with other popular recognition methods. The results of experiment verify the effectiveness of the proposed method, and the technique reaches high recognition accuracy, good robustness, and strong real-time performance for UAV power inspection system.

INDEX TERMS Deep Learning, SRCNN, YOLO V3, electrical components, object detection.

I. INTRODUCTION

The traditional manual inspection has been used for electrical components detection for several decades. However, this method is significantly constrained by the real inspection environment, such as steep terrain and harsh climate. Also, this traditional solution appears to be slow, expensive, and potentially dangerous. In recent years, the Unmanned Aerial Vehicle (UAV) inspection technology [1]–[4] has been widely applied in the detection of electrical components by worldwide power grid companies. The UAV inspection images enable the engineer to detect the electrical components effectively. Meanwhile, the increasing industrial needs have attracted more scholars for further research in the image recognition technology. For example, Wang *et al.* in [5] propose an insulator defect-recognition method which integrates the shape, colour, and texture of insulators to reduce the influence of the background texture and illumination effectively. Lin *et al.* in [6] conduct a faulty insulator diagnosis method for insulator detection based on repetitiveness feature from the UAV video sequence. Jin *et al.* in [7] realize shockproof hammer recognition algorithm based on Haar-like features and cascade AdaBoost classifier. A. Cerón *et al.* in [8] present a novel method named Circle Based Search (CBS) for power line detection based on the search of lines between two opposite points. Juan I. Larrauri *et al.* in [9] present a new automatic system in almost real-time for overhead power line inspection by UAV. Carlos Sampedro *et al.* in [10] put forward a supervised learning approach for solving the tower detection and classification problem. The UAV inspection combined with various image processing skills can significantly improve the speed and efficiency of power maintenance and overhaul, reduce labour intensity and cost, and significantly improve the safety of inspection operators and other advantages. However, the images obtained by UAV inspection are also restricted by the environment, which brings higher requirements for UAV image recognition technology. In the past few years, the traditional electrical component recognition algorithms are mainly based on local features of images. Yan *et al.* in [11] present a fault diagnosis method for transmission line based on Scale Invariant Feature Transform (SIFT). Martinez *et al.* in [12] propose an approach based

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

on HOG features and MLP neural networks for detecting and classifying electric towers for power line inspection. Han *et al.* in [13] conduct a method based on DPM to detect and locate the rod-insulators in the image taken from the high-speed railway catenary system. The above methods are lack of scalability and poor real-time performance for some reason. It is difficult to fully extract features from inspection images to identify electrical components, especially when dealing with occlusion problems, which makes these algorithms hard to achieve the desired accuracy.

Due to the higher demand for image recognition of UAV inspection, deep learning [14] has received extensive attention since 2012. Some scholars have applied improved deep learning algorithms [15]–[17] to conduct related object detection. A great number of object detection frameworks are proposed and improved successively. Two branches of them are two-stage methods and one-stage methods. As for two-stage methods, Girshick *et al.* in [18] combine the CNN [19] with Region Proposal (RP), and propose the R-CNN framework. However, the R-CNN network needs to fix the size of the input image. Besides, the algorithm is computationally complex and memory-consuming. In response to this problem, Girshick [20] propose the Fast R-CNN framework to realize the image input at any scale in the network by adding the SPP layer. But there are still bottlenecks of R-CNN in the selection of candidate regions. To solve this problem, Ren *et al.* [21] propose the framework of Faster R-CNN which replaces Selective Search (SS) with RPN and shares the convolutional layer between RPN and Fast R-CNN. Furthermore, candidate boxes are generated by convolutional feature map, which significantly improves the detection speed and performance. It needs to be emphasized that the training process of this algorithm is still complex, and there is a significant optimization space in the calculation process. Joseph Redmon *et al.* propose the YOLO V1 [22] and its enhanced versions which not only overcome the bottleneck of Faster R-CNN in recognition speed, but also receive extensive attention for extreme real-time performance and excellent accuracy. For the one-stage algorithms, Single Shot Multibox Detector (SSD) [23] and YOLO V1 have poor performance in small object detection. YOLO V2 [24] and YOLO V3 [25] are the typical methods which improve the speed and accuracy of object detection while changing the structure constantly. Therefore, for the purpose of improving the recognition speed and accuracy of electrical components, object detection algorithms based on deep learning have been widely used in power inspection systems in recent years. Tao *et al.* in [26] propose a novel deep CNN cascading architecture for both localization and detection of insulators. Zhao *et al.* in [27] propose a CNN model with a multi-patch feature extraction method to inspect and represent the status of insulators. Lei *et al.* in [28] put forward a method based on Faster R-CNN to locate the broken insulators, and Zhang *et al.* in [29] also adopt Faster R-CNN to identify the grading ring. In [30], Wang *et al.* adopt YOLO to realize the identification of the insulators and the location of the discharge position based on ultraviolet imaging. It is worth mentioning that the accuracy and detection speed of the above methods fail to meet the actual engineering requirements, which lack of reliability. When the object is multiple electrical components, the identification efficiency is not good. It is of considerable significance to improve the efficiency and speed of inspection.

In this paper, a framework of electrical components recognition algorithm based on SRCNN and YOLO V3 is proposed. The SRCNN network is used to achieve the super-resolution reconstruction for blurred images. The reconstructed clear images are combined with other training images to enlarge the datasets. Then the YOLO V3 accomplishes the recognition of electrical components. In addition, we have conducted further analysis on the production of dataset, the selection of image classification algorithm, and the determination of network parameters. And the experimental results show that the proposed algorithm can realize recognition effect in UAV inspection images compared with Faster R-CNN and SSD.

The other parts of the paper are organized as follows: Section II describes the detection procedure of electrical components and the role of the various parts of the framework. In addition, this section introduces the theoretical background of the proposed SRCNN algorithm and introduces the detection principle of the proposed YOLO V3 in detail. Dataset, experimental results, evaluation of our algorithms and comparison with other methods are discussed in Section III. Finally, Section IV concludes this paper.

II. PROPOSED ELECTRICAL COMPONENTS DETECTION FRAMEWORK

In this section, the application of the detection framework for electrical components is presented. The procedure for electrical components detection includes the reconstruction process of SRCNN and detection process of YOLO V3. Firstly, the network structure of SRCNN used in super-resolution reconstruction is given, and then the network training of SRCNN is introduced briefly. Secondly, the detection model, network structure and network training process of YOLO V3 are described in detail. Meanwhile, we optimize the gradient descent algorithm in the training process of YOLO V3.

A. PRINCIPLE OF ELECTRICAL COMPONENTS DETECTION

1) DETECTION PROCEDURE OF ELECTRICAL COMPONENTS

The main electrical components detection procedure is shown in Figure 1. The UAV inspection image preprocessing can be achieved via the following steps. Firstly, the original UAV inspection image set can be classified into two types. One is the qualified image set which can be used as the image training set, the other type is kind of blur image set with low resolution. Secondly, the blur image set is processed via SRCNN for super-resolution reconstruction, which means the blur image set can be switched from the lower-resolution to higher resolution. The processed images combined with the original images can be applied as the suitable inspection image sets.

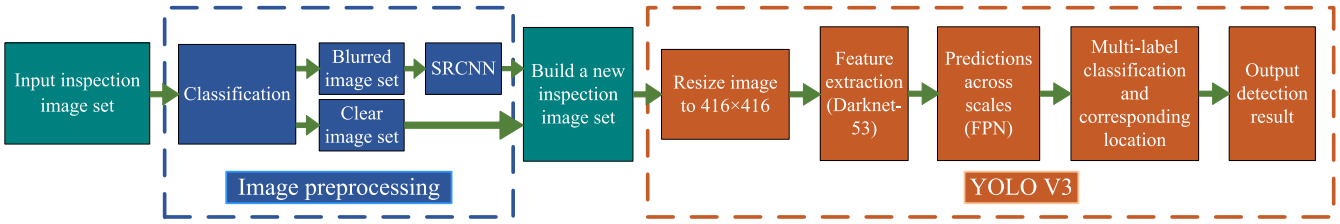


FIGURE 1. The main electrical components detection procedure.

In addition, the new inspection image set is adjusted from the original resolution to resolution 416×416 in the YOLO V3 structure. The resized image set is input into the Darknet-53 to extract electrical component features. Then the feature pyramid networks (FPN) makes the predictions across 3 different scales via feature outputs of Darknet-53. Overall, the prediction results of YOLO V3 contain the parameters of bounding box, objectness score and prediction over classes. Next, YOLO V3 filters out the anchors that overlap the ground truth object by more than a chosen threshold and then outputs the classification and corresponding positioning of each bounding box. Finally, the network outputs the detection results.

2) IMAGE CLASSIFICATION

In this paper, we apply the Laplace variance algorithm to classify the blurred image and the non-blurred image. Laplace operator is adopted to calculate the second-order differential of the image, which increases the difference between neighborhood pixels and makes the mutation part of the image more obvious. This paper uses Laplace operator to convolute the input image firstly, and then calculate the variance. In normal pictures, the boundary appears to be much clearer, so the variance is larger. Instead, the boundary information of blurred pictures is relatively less, so the variance is smaller. Therefore, if the variance is lower than the predefined threshold, the image is marked as ''blurred''. Conversely, if the variance is higher than the predefined threshold, the image is marked as ''clear''. After the above operation, the original inspection image set is divided into a blurred image set and a clear image set.

B. RECONSTRUCTION PROCESS OF SRCNN

1) SRCNN

Due to UAV fuselage shaking and problems of imaging exposures, some inspection images show blurred quality, this will severely hinder the effective recognition of electrical components of UAV inspection. Therefore, the image preprocessing should be a necessary step for forming the deep learning training set. Chao Dong *et al.* in [31] propose the SRCNN algorithm which is a network learning algorithm about data preprocessing. This algorithm combines the super-resolution reconstruction algorithm of deep learning [32]–[36] for the first time, which effectively solves the above problems by reconstructing low-resolution images to corresponding high-resolution photos. The SRCNN mainly contains three steps: image block extraction and representation, non-linear mapping and high-resolution image block reconstruction.

a: IMAGE BLOCK EXTRACTION AND REPRESENTATION

The typical operation is to extract the image blocks from the original inspection image, and then use a series of filters to convolute with the image. Each filter can be regarded as a base, and the optimization of the bases can be incorporated into the optimization of the network. Based on the above discussion, the processing of the first layer could be denoted as:

$$F_1(Y) = \max(0, W_1 * Y + B_1) \tag{1}$$

where W_1 corresponds to n_1 filters of size $c \times f_1 \times f_1$, c is the number of channels of the input image, f_1 is the spatial scale of the filter, B_1 is an n_1-dimensional vector, and $*$ represents the convolution operation. The output of the convolution operation consists of n_1 feature maps, and the output image of the first convolution layer is obtained through the activation function ReLU(max(0, x)) [37].

b: NON-LINEAR MAPPING

In the first convolutional layer, an n_1-dimensional feature vector is extracted from each image block. In the second convolution layer, the output n_1-dimensional feature vector is nonlinearly mapped to the n_2-dimensional feature vector. The processing of the second layer can be described as:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \tag{2}$$

where W_2 corresponds to n_2 filters of size $n_1 \times f_2 \times f_2$, B_2 is an n_2-dimensional vector. The output of the convolution operation consists of n_2 feature maps.

c: HIGH-RESOLUTION IMAGE BLOCK RECONSTRUCTION

The third convolutional layer combines all the high-resolution image blocks generated by the second layer to create a high-resolution image, i.e. the final output image of the SRCNN network. On this basis, the processing of the third layer could be indicated as:

$$F(Y) = W_3 * F_2(Y) + B_3 \tag{3}$$

where W_3 corresponds to c filters of size $n_2 \times f_3 \times f_3$, B_3 is a c-dimensional vector.

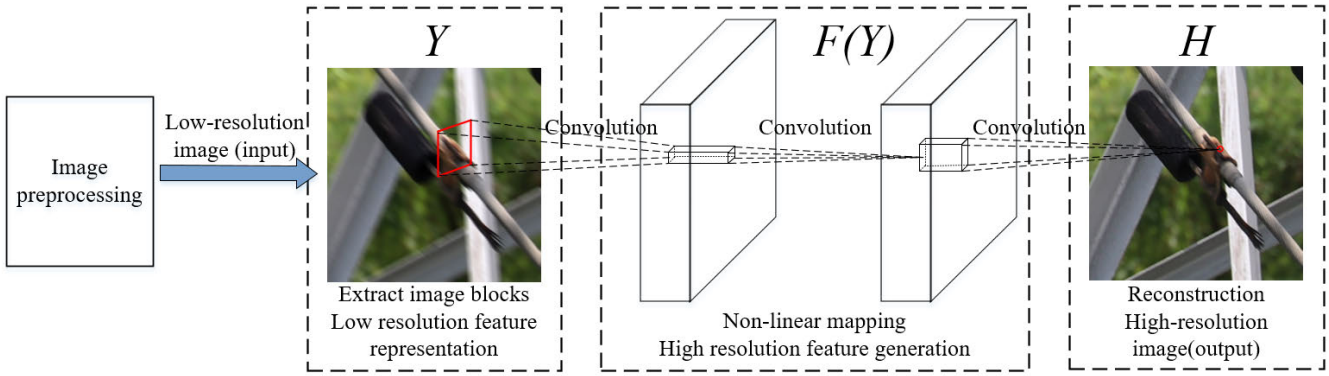


FIGURE 2. SRCNN network structure.

2) DATASET PREPARATION FOR TRAINING

Due to most blurred images in UAV inspection are caused by jitter. Therefore, further processing of inspection images should be considered before the training of SRCNN network. This paper follows the training method of reference paper [38] and [39] when preparing the training dataset for SRCNN. We collect 100 clear images from the clear image set. The blurred images can be obtained through using a 2×2 convolution kernel to dawn-sample the clear inspection images. Meanwhile, the kinematics blur generating formula (4) is used to process the clear images to get blurred images. Finally, two parts of blurred images and corresponding clear images are merged into blurred and clear image pairs to train the SRCNN network.

$$g(x, y) = \int_0^T f\left[\left(x - x_0(t)\right), \left(y - y_0(t)\right)\right] dt \tag{4}$$

3) NETWORK STRUCTURE OF SRCNN

The SRCNN network first enlarges the extracted blurred inspection image to the target size using the Bicubic interpolation algorithm and records the interpolated image as Y. The goal of super-resolution reconstruction is to restore Y to the high-resolution image H, which is similar to the original resolution image X. And obtain the corresponding "end-to-end" mapping function $F(Y)$ through training. The basic structure of SRCNN network is shown in Figure 2. It can be seen from Figure 2 that the whole structure consists of a three-layer convolution neural network, which can be divided into three levels, corresponding to the three steps of image super-resolution reconstruction.

1) The first convolutional layer extracts the image blocks from Y, and then represent these low-resolution features.
2) The second convolutional layer generates high-resolution features through non-linear mapping.
3) The third convolutional layer accomplishes the reconstruction of high-resolution images, which is equivalent to create images that are close to the original high-resolution images.

4) NETWORK TRAINING OF SRCNN

In the training process of the network, it is necessary to learn the network parameter $\theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$ to obtain the mapping function F between high-resolution images. The learning of these parameters needs to minimize the error between the reconstructed image $F(Y; \theta)$ and the original high-resolution image X. The mean square error (MSE) is adopted as the loss function, and MSE is shown below:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \|F(Y_i; \theta) - X_i\|^2 \tag{5}$$

where n represents the number of training samples, X_i is the original high-resolution image, Y_i is the input low-resolution image, and $F(Y_i; \theta)$ is the high-resolution image reconstructed by SRCNN model.

C. DETECTION PROCESS OF YOLO V3

1) YOLO V3

YOLO V3 predicts boxes at different scales using the concept of FPN. It mainly utilizes a certain amount of convolutional layers and residual layers to complete the detection process, and uses the features of the entire image to predict each bounding box. At the same time, it predicts all classes of all bounding boxes to realize end-to-end training, which maintains high average accuracy and strong real-time performance.

YOLO V3 starts with dividing the input image into N×N grids, and assigns one bounding box anchor for each ground truth object. As shown in Figure 3, the network predicts 4 parameters (t_x, t_y, t_w, t_h) for each bounding box, and then applies a function to predict 4 corresponding coordinates: the center point coordinates (b_x, b_y) of the bounding box, the width b_w and the height b_h. The prediction of the bounding box and the Intersection-over-Union (IOU) are shown as the following equations:

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$

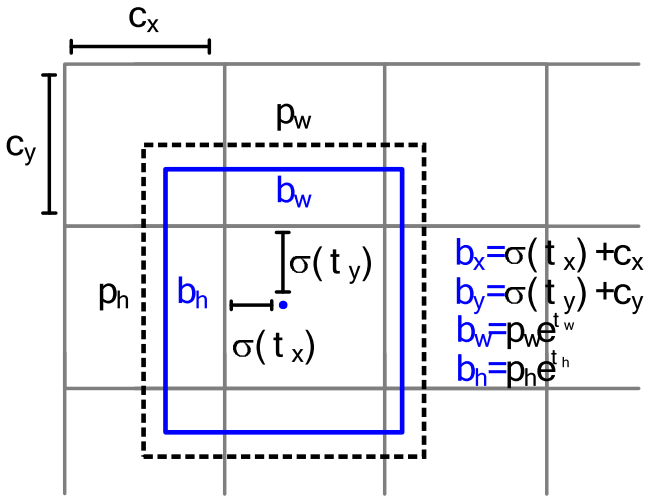


FIGURE 3. Bounding boxes with dimension priors and location prediction.

$$b_w = p_w e^{t_w}$$
$$b_h = p_h e^{t_h} \quad (6)$$
$$IOU = \frac{area\left(BB_{dt} \cap BB_{gt}\right)}{area\left(BB_{dt} \cup BB_{gt}\right)} \quad (7)$$

where the IOU in Equation (7) is the overlap rate between the bounding box detected by the system and the ground truth box. BB_{gt} is the ground truth box based on training label, BB_{dt} is the detection bounding box, and area(·) indicates the region.

2) DETECTION MODEL BASED ON YOLO V3

In the actual inspection images taken by UAV, due to the complex shooting background and the small proportion of the electrical components especially the shockproof hammer of the entire image, it is necessary to optimize the network structure to detect and recognize the state accurately. Figure 4 shows the detection model map of YOLO V3. Before training, Figure 4(a) divides each input inspection image into 10×10 grids. Each grid in Figure 4(b) and Figure 4(c) predicts bounding boxes and categories of probability. Figure 4(d) outputs the final prediction with a 10×10×(3×(4+1+2)) tensor. As the detection results show that this model detects an insulator and four shockproof hammers.

3) NETWORK STRUCTURE OF YOLO V3

Figure 5 shows the main structure of the YOLO V3 network, which adopts the Darknet-53 structure. This structure is a combination of YOLO V2, Darknet-19 and ResNet. And YOLO V3 mainly uses 1×1 and 3×3 convolutional kernels, and some shortcut connection structures. In Figure 5, the input inspection image is first pre-processed, then its resolution is adjusted to 416×416, and next the image is operated through YOLO V3 subsequently.

1) The first part consists of 2 convolutional layers. The input image size is 416×416×3, and the convolution kernel size is 3×3×32 and 3×3×64. The size of the output feature map is compressed to 208×208×64 after the convolution operation.
2) The second part is composed of 3 convolutional layers and a residual layer. The size of the convolution kernels is 1×1×32, 3×3×64 and 3×3×128, and the output feature map is compressed to 104×104×128 after the convolution operation.
3) The third part is composed of 5 convolutional layers and 2 residual layers. The size of the convolution kernels is 1×1×64, 3×3×128 and 3×3×256, and the output feature map is compressed to 52×52×256 after the convolution operation.
4) The fourth part is composed of 17 convolutional layers and 8 residual layers. The size of the convolution kernels is 1×1×128, 3×3×256 and 3×3×512, and the output feature map is compressed to 26×26×512 after the convolution operation.
5) The fifth part is composed of 17 convolutional layers and 8 residual layers. The size of the convolution kernels is 1×1×256, 3×3×512 and 3×3×1024, and the output feature map is compressed to 13×13×1024 after the convolution operation.
6) The sixth part is composed of 8 convolutional layers and 4 residual layers. The size of the convolution kernels is 1×1×512 and 3×3×1024, and the size of the output feature map remains the same after the convolution operation.
7) The seventh part is composed of 3 prediction networks. YOLO V3 predicts boxes at 3 different scales and then extracts features form those scales. The prediction result of the network is a 10×10×(3×(4+1+2)) tensor for the 4 bounding box offsets, 1 objectness prediction, and 2 class predictions.

4) NETWORK TRAINING OF YOLO V3

The network training of YOLO V3 is mainly divided into three steps:

Step 1: Since the resolution of the original inspection image taken by UAV is 3840×2160, which is too large to be the input of the network. The original image is adjusted to a size of 416×416 to speed up the network training process.

Step 2: VOC2007 dataset format is adopted to mark the outer frame of insulators or shockproof hammers appearing in each image.

Step 3: Initialize the network parameters of the YOLO V3 model, and then train the YOLO V3 model to obtain the training parameters for the detection of electrical components. This paper makes a further analysis on the selection of three important parameters.

1) *Batch size:* In theory, the bigger the value is, the better the training will be. However, we can not increase the value indefinitely due to the limitation of hardware conditions, so we tried four batch sizes of 128, 64, 16 and 8, respectively. When the batch size is selected as 64, 16, 8 during training, there is no memory shortage, so we choose 64 as the batch size according to the above discussion.

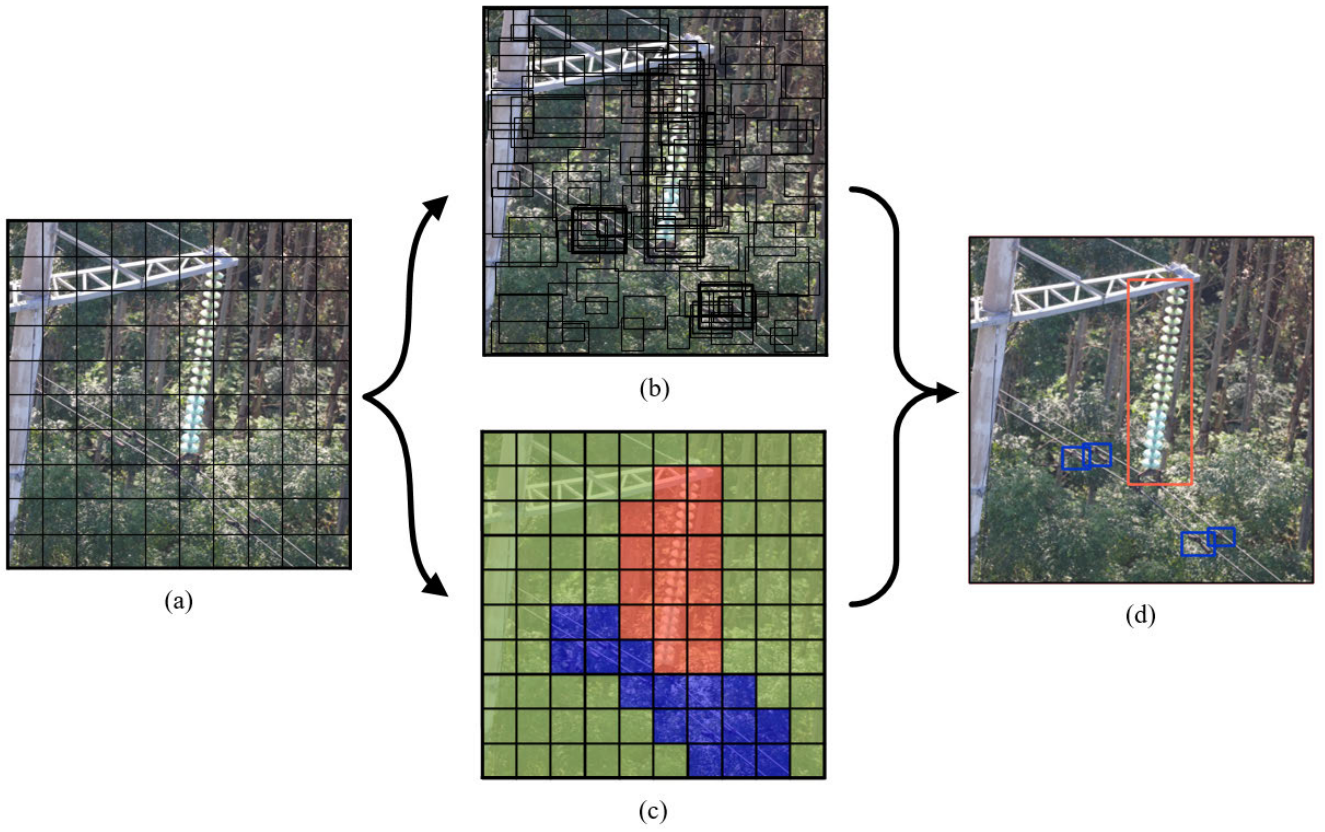


FIGURE 4. Detection model of YOLO V3. (a) Divide the inspection image into 10×10 grids. (b) The predicted bounding boxes with confidence. (c) Class probability map. (d) Final detection of YOLO V3.

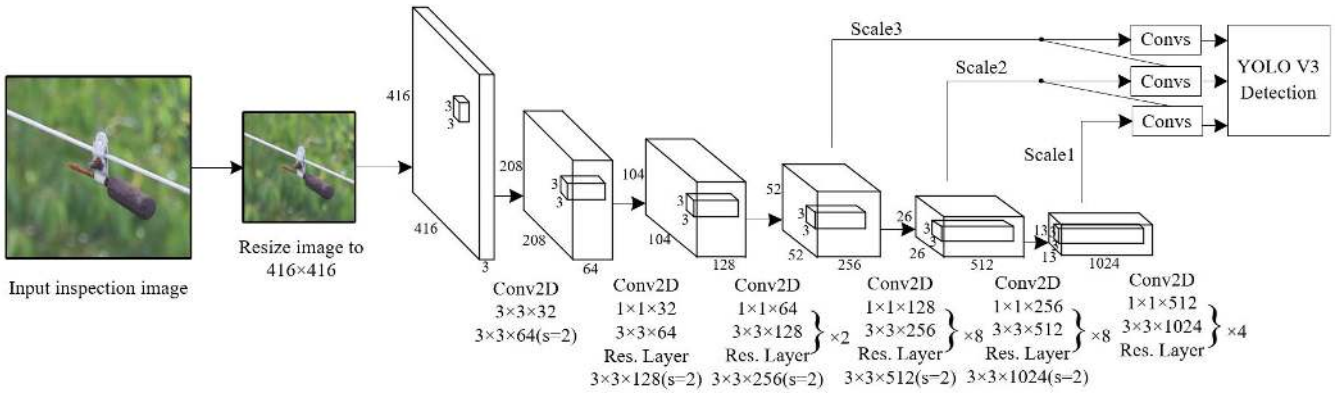


FIGURE 5. YOLO V3 Network structure.

2) *Weight decay:* In order to prevent overfitting, we fix a suitable learning rate firstly, and then adjust the decay value from a fixed value (0.01), and finally the value is 0.0005.
3) *Ignore_thresh:* It refers to the size of the IOU threshold, which determines the number of IOU involved in the calculation of loss. If the threshold is too small, the number will be too large, which will lead to underfitting. Conversely, if the threshold is too large, it is easy to cause overfitting. Therefore, the value of ignore_thresh is 0.65 based on the above discussion and the actual situation.

The selection of the above three parameters will affect the recognition accuracy. So it is necessary to adjust these parameters to effectively avoid the underfitting and overfitting when training the model, so as to improve the average recognition accuracy.

YOLO V3 adopts the multi-label classification, which is different from the previous versions that use the mutually exclusive label. And it uses a logistic classifier to calculate the objectness score for each bounding box. For the loss of the classification during training, YOLO V3 adopts the binary cross-entropy loss for each label, which takes the place of the mean square error used generally in the previous versions.

The loss function used in parameter training of YOLO V3 is indicated as follows:

$$L(s_n) = \begin{cases} -\log(s_n) & g_n = 1 \\ -\log(1 - s_n) & g_n = 0 \end{cases} \quad (8)$$

where n indicates the sample index, $s_n \in [0,1]$ represents the objectness score predicted by the network, which measures the predicted probability that the n-th sample is an electrical component. And g_n indicates the ground truth. It should be noted that $g_n \in \{0,1\}$ means if the n-th sample belongs to the class of objects. The network parameters are trained through minimizing the loss for all samples, i.e., $\Sigma_n L(s_n)$.

5) GRADIENT DESCENT OPTIMIZATION ALGORITHM

In this paper, we adopt Adaptive Moment Estimation(Adam) [40] optimization algorithm to update network parameters. Adam is a first-order optimization algorithm, which can replace the traditional SGD [41] process and update the network weight iteratively based on the training data. The algorithm designs the corresponding adaptive learning rate for different parameters by calculating the first and second-moment estimation of the gradient. It combines two optimization algorithms, including the advantages of adaptive gradient algorithm [42] and root-mean-square propagation [43], which is beneficial to improve the performance of sparse gradients and the training efficiency.

III. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL ENVIRONMENT

The basic configuration of PC used in this experiment is as follows: Intel(R) Xeon(R) CPU E5-8276, primary frequency 4.0 GHz, using NVIDIA Tesla K80 as graphics card, dual GPU accelerator, 24 GB GDDR4 memory, 128 GB memory, and supporting 480 GB/s bandwidth. The operating system is Ubuntu 16.04, and the deep learning framework of algorithm programming is Tensorflow.

B. DATASET

Since there is no publicly available dataset of the insulator and shockproof hammer, the dataset used in this experiment is all derived from the UAV inspection images. The dataset used in this experiment is derived from the power inspection images taken by UAV in a specific area of China, clear images and blurred images can be obtained by classifying the dataset. Since training requires enough dataset, it is necessary to restore the blurred images through super-resolution reconstruction to supplement the dataset.

The dataset is divided into two categories, a total of 8128 pieces. Due to multiple electrical components may appear in the same image, this dataset includes 4416 insulator images, and 4352 shockproof hammer images. This paper applies the hold-out method to divide the dataset containing 8128 samples into a training set containing 6432 samples and a test set containing 1696 samples. Since the size, shape, light intensity and other physical factors of insulators and shockproof hammers in the dataset are different from each other, in order to ensure the generalization performance of the YOLO V3 model, more training samples are generated by adjusting the rotation angle, saturation, exposure and tone of some samples in the dataset during the training process. The sample images of partial insulator and shockproof hammer are shown in Figure 6 and Figure 7, respectively.

FIGURE 6. Images of insulators.

FIGURE 7. Images of shockproof hammers.

C. EXPERIMENTS AND RESULTS ANALYSIS

The SRCNN network and YOLO V3 network have experimented in this section, and the experimental results are analyzed in detail.

FIGURE 8. Reconstruction performance of three algorithms.

1) ANALYSIS OF EXPERIMENTAL RESULTS OF SRCNN NETWORK

The blurred image set obtained by classification is taken as the experimental object for super-resolution reconstruction. The performance analysis of SRCNN network shows that the reconstruction effect is affected by the number and size of filters and the number of network layers. This paper refers to the setting result of the SRCNN model parameters in [38], and sets the parameters used in this experiment to $f_1 = 9$, $f_2 = 1$, $f_3 = 5$, $n_1 = 64$, $n_2 = 32$, and the upscaling factor is 3. At the same time, traditional algorithms such as Bicubic interpolation and Sparse Coding [44] are used for experimental comparison. The experimental results in Figure 8 show that the PSNR value of SRCNN achieves 32.35 dB, which reaches the ideal reconstruction quality level. To demonstrate the performance superiority of the SRCNN network more obviously, Figure 9 shows the average PSNR curves of the three algorithms. The following conclusions can be drawn intuitively and clearly from Figure 9:

1) SRCNN exceeds Bicubic interpolation and Sparse Coding only through a few training iterations.
2) The SRCNN algorithm further enhances the performance of super-resolution reconstruction through more training iterations.
3) The SRCNN algorithm can be effectively applied to the super-resolution reconstruction of images. Therefore, for the blurred power inspection images obtained by classification, SRCNN algorithm can be used to train a certain number of iterations to generate the corresponding high-resolution images. The process is shown in Figure 10.

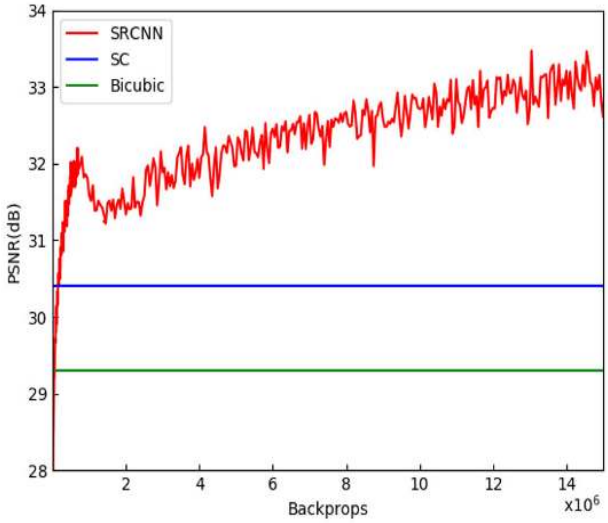


FIGURE 9. The test convergence curve of SRCNN and results of other methods.

2) ANALYSIS OF EXPERIMENTAL RESULTS OF YOLO V3 NETWORK

This experiment adopts the optimized algorithm of the YOLO V3 model in the previous chapter and initializes the network on this basis. Then set some key parameters, in which the kernel function of the convolutional layer is set to 1×1 or 3×3, the stride size is 1, the batch size is 64, the parameters are updated once for each batch of training samples, and the regularization coefficient of weight decay is 0.0005, the adaptive learning rate is changed dynamically with an initial value of 0.01, and the maximum number of iterations is 1600.

In the training process, to check the detection performance of YOLO V3 compared with Faster R-CNN and SSD, the accuracy of the three algorithm models in identifying the training samples is counted when the number of iterations is 400, 800, 1200 and 1600 respectively. The results are shown in Table 1. From the data in Table 1, it can be seen that with the increase of iteration times, the average recognition accuracy of the three algorithms also increases gradually, among which the recognition accuracy of YOLO V3 algorithm is the highest, reaching 95.84%. It also can be concluded from Table 1 that in the case of the same number of training samples, the YOLO V3 algorithm takes the least training time to convergence, that is, the algorithm has the fastest training speed. The recognition results of the insulators and shockproof hammers in the inspection image are shown in Figure 11. It can be seen clearly from Figure 11 that the algorithm proposed in this paper can accurately locate the electrical component. In the meanwhile, it also verifies the effectiveness of the algorithm adopted in this paper in recognition of electrical components.

After the network training is completed, the test set is used to test and verify the detection performance of electrical components in test samples by three algorithm models through test accuracy and Mean Average Precision (mAP). The test results are shown in Table 2. Both the test accuracy and mAP

TABLE 1. The recognition accuracy of the three methods.

Algorithm	Training samples	Average Accuracy/% (400 iterations)	Average Accuracy/% (800 iterations)	Average Accuracy/% (1200 iterations)	Average Accuracy/% (1600 iterations)	Training time/h
Faster R-CNN	6432	66.16	77.26	86.35	92.14	25.21
SSD	6432	68.34	80.46	89.57	94.21	22.78
YOLO V3	6432	72.48	83.72	91.45	95.84	21.52

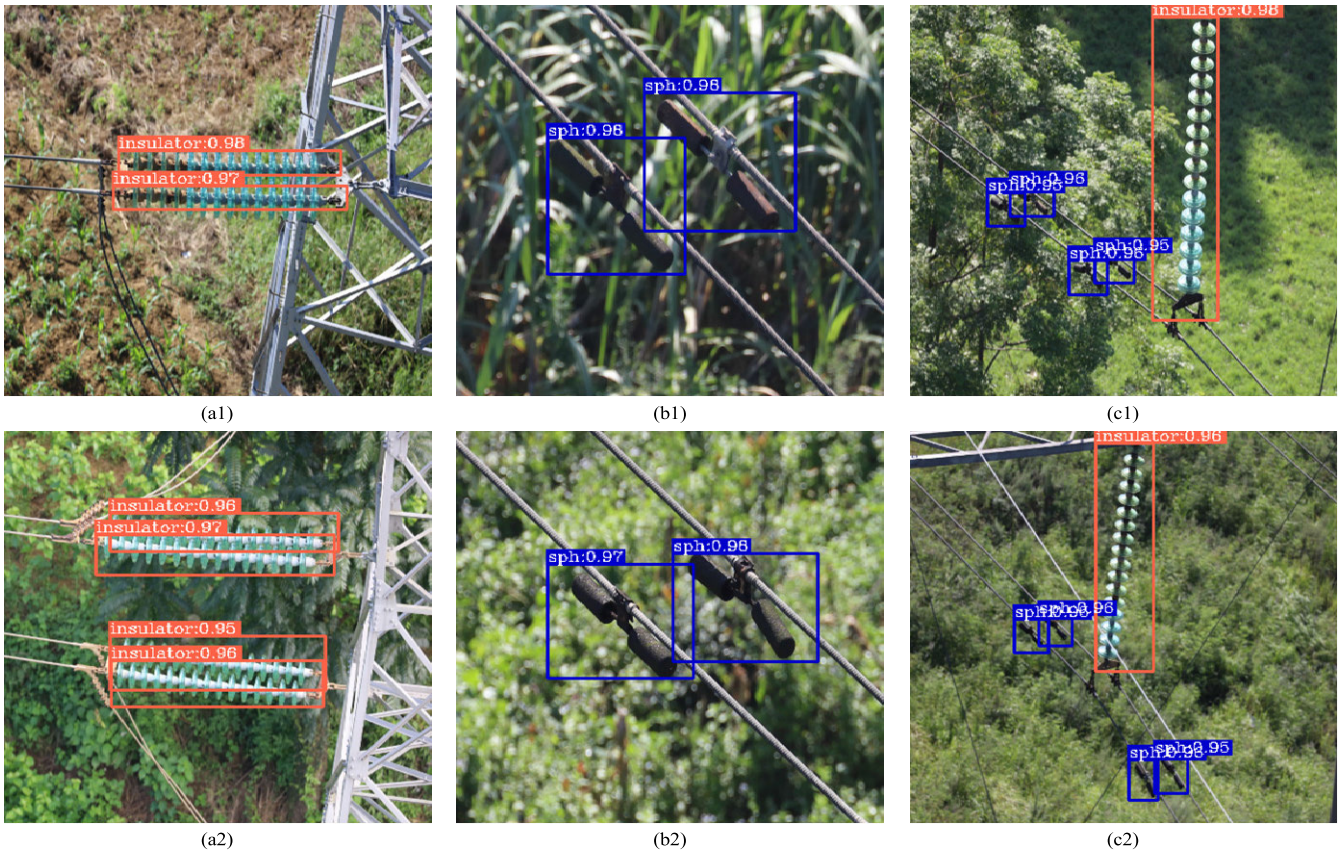
TABLE 2. Performance comparison of three algorithms.

Algorithm	Test samples	Test Accuracy/%	Test time/min	mAP
Faster R-CNN	1696	93.26	2.83	0.861
SSD	1696	95.33	1.57	0.913
YOLO V3	1696	96.45	1.13	0.936

FIGURE 10. Super-resolution reconstruction process using SRCNN (From left to right: (a) 15000 training iterations with SRCNN. (b) 35000 training iterations with SRCNN. (c) 50000 training iterations with SRCNN).

value of YOLO V3 are the maximum, which means that YOLO V3 achieves stronger generalization ability. Meanwhile, the testing process of YOLO V3 requires less time, which further verifies that the speed of YOLO V3 algorithm is the fastest.

In the process of training with the algorithm proposed in this paper, to intuitively reflect the detection accuracy of corresponding objects in the dataset, the IOU is used to measure the correlation between the real and the prediction. The higher the association, the higher the IOU value. As the training



(a1) (b1) (c1)

(a2) (b2) (c2)

FIGURE 11. Some recognition results on the test set. (a1) and (a2): Recognition result of insulators. (b1) and (b2): Recognition result of shockproof hammers. (c1) and (c2): Recognition result of insulators and shockproof hammers.

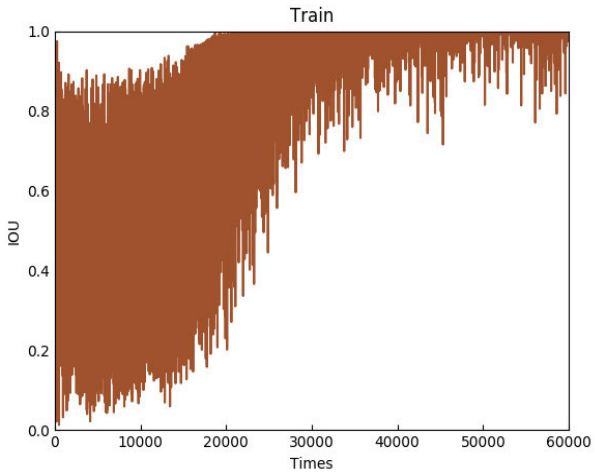


FIGURE 12. The Region Average IOU curves.

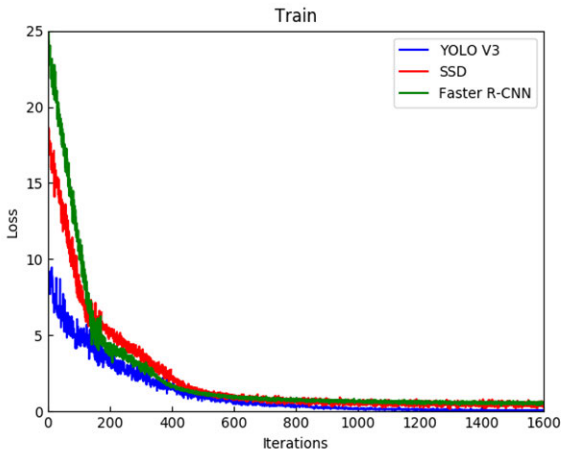


FIGURE 13. Loss function comparison.

batch increases during the training process, the IOU value also increases, the IOU trend graph is shown in Figure 12.

The loss function is a way to measure the gap between the predicted value and the actual value of the network output, which is used to measure the quality of the model prediction. The visualization function image is shown in Figure 13. It can be concluded obviously from Figure 13 that the loss of the three algorithms decreases gradually with the increase of iteration times. When the number of iterations is constant, the loss of YOLO V3 algorithm is the minimum, and the convergence speed is faster than the other two algorithms.

In conclusion, compared with Faster R-CNN and SSD, the proposed method has more advantages in the detection performance of electrical components.

IV. CONCLUSION

Detection and recognition of electrical components is an essential part of power inspection. To solve the problem that the traditional algorithm cannot extract the image features adequately under the interference of complex background and other factors, this paper proposes a YOLO V3-based method for electrical component detection from UAV inspection images. The YOLO V3 model, combined with SRCNN, can accurately identify the position and state of electrical components under different angles, backgrounds and illumination intensities. The experimental results show that the recognition accuracy of YOLO V3 is 1% to 3% higher than Faster R-CNN and SSD. Meanwhile, YOLO V3 outperforms the other two methods in recognition speed, which can achieve almost real-time performance. The method proposed in this paper has application value for the intelligentization of power system inspection and has particular practical significance for improving power inspection technology.

REFERENCES

[1] G. Zhou, J. Yuan, I.-L. Yen, and F. Bastani, "Robust real-time UAV based power line detection and tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 744–748.

[2] S. Jordan, J. Moore, S. Hovet, J. Box, J. Perry, K. Kirsche, D. Lewis, and Z. T. H. Tse, "State-of-the-art technologies for UAV inspections," *IET Radar, Sonar Navigat.*, vol. 12, no. 2, pp. 151–164, Feb. 2018.

[3] X. Hui, J. Bian, X. Zhao, and M. Tan, "Deep-learning-based autonomous navigation approach for UAV transmission line inspection," in *Proc. 10th Int. Conf. Adv. Comput. Intell. (ICACI)*, Xiamen, China, Mar. 2018, pp. 455–460.

[4] F. Tian, Y. Wang, and L. Zhu, "Power line recognition and tracking method for UAVs inspection," in *Proc. IEEE Int. Conf. Inf. Automat.*, Aug. 2015, pp. 2136–2141.

[5] 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)*, Hangzhou, China, Dec. 2016, pp. 162–167.

[6] L. Lin, B. F. Li, L. Wang, Y. Cong, and Y. D. Tang, "Faulty insulator diagnosis for UAV videos based on repetitiveness feature," *Appl. Mech. Mater.*, vol. 423, pp. 2535–2542, Sep. 2013.

[7] L.-J. Jin, S.-J. Yan, and Y. Liu, "Vibration damper recognition based on Haar-like features and cascade AdaBoost classifier," *J. Syst. Simul.*, vol. 24, pp. 1806–1809, Sep. 2012.

[8] A. Cerón, I. F. M. B, and F. Prieto, "Power line detection using a circle based search with UAV images," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Orlando, FL, USA, May 2014, pp. 632–639.

[9] J. I. Larrauri, G. Sorrosal, and M. González, "Automatic system for overhead power line inspection using an unmanned aerial vehicle—RELIFO project," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Atlanta, GA, USA, May 2013, pp. 244–252.

[10] 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)*, Beijing, China, Jul. 2014, pp. 1970–1977.

[11] S. Yan, L. Jin, Z. Zhang, and W. Zhang, "Research on fault diagnosis of transmission line based on SIFT feature," in *Proc. Int. Symp. Neural Netw.* Berlin, Germany: Springer, 2013, pp. 569–577.

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

[13] Y. Han, Z. Liu, D.-J. Lee, G. Zhang, and M. Deng, "High-speed railway rod-insulator detection using segment clustering and deformable part models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 3852–3856.

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

[15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[16] G. Tao, C. Fengxiang, W. Wei, S. Ping, S. Lei, and C. Tianzhu, "Electric insulator detection of UAV images based on depth learning," in *Proc. 2nd Int. Conf. Power Renew. Energy (ICPRE)*, Chengdu, China, Sep. 2017, pp. 37–41.

[17] H. Tayara, K. G. Soo, and K. T. Chong, "Vehicle detection and counting in high-resolution aerial images using convolutional regression neural network," *IEEE Access*, vol. 6, pp. 2220–2230, 2018.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.

[19] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Columbus, OH, USA, Jun. 2014, pp. 512–519.

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

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

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[24] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 7263–7271.

[25] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Mar. 2018, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767

[26] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, to be published.

[27] 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)*, Vancouver, BC, Canada, Jul. 2016, pp. 3187–3194.

[28] X. Lei and Z. Sui, "Intelligent fault detection of high voltage line based on the Faster R-CNN," *Measurement*, vol. 138, pp. 379–385, May 2019.

[29] J. Wang, X. Zhang, L. Zheng, and S. Masanori, "A study on the grading ring recognition method of power line based on deep learning," in *Proc. Int. Conf. Inf. Commun. Technol. Robot. (ICT-ROBOT)*, Busan, South Korea, Sep. 2018, pp. 1–4.

[30] S. Wang, L. Niu, and N. Li, "Research on image recognition of insulators based on YOLO algorithm," in *Proc. Int. Conf. Power Syst. Technol. (POWERCON)*, Guangzhou, China, Nov. 2018, pp. 3871–3874.

[31] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[32] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.

[33] R. Timofte, R. Rothe, and L. Van Gool, "Seven ways to improve example-based single image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1865–1873.

[34] W.-T. Lu, C.-W. Lin, C.-H. Kuo, and Y.-C. Tung, "Image super-resolution based on error compensation with convolutional neural network," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Kuala Lumpur, Malaysia, Dec. 2017, pp. 1160–1163.

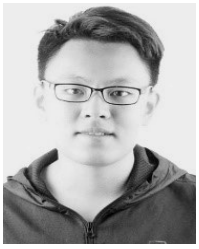
[35] Y. Kato, S. Ohtani, N. Kuroki, T. Hirose, and M. Numa, "Image super-resolution with multi-channel convolutional neural networks," in *Proc. 14th IEEE Int. New Circuits Syst. Conf. (NEWCAS)*, Vancouver, BC, Canada, Jun. 2016, pp. 1–4.

[36] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883.

[37] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 807–814.
[38] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[39] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 391–407.
[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[41] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.
[42] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016, *arXiv:1609.04747*. [Online]. Available: https://arxiv.org/abs/1609.04747
[43] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[44] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

HAIPENG CHEN received the B.S. and M.S. degrees from the Changchun University of Science and Technology, in 2006 and 2010, respectively, and the Ph.D. degree from Jilin University, in 2015. He is currently an Associate Professor with Northeast Electric Power University. His main research interests include deep learning and high-performance computing in power system.

ZHENTAO HE received the B.S. degree in communication engineering from Northeast Electric Power University, in 2018, where he is currently pursuing the M.S. degree in information and communication engineering. His main research interests include deep learning, image processing and computer vision.



BOWEN SHI is currently pursuing the B.S. degree with Northeast Electric Power University. His main research interests include deep learning and computer vision.

TIE ZHONG received the bachelor's, master's and Ph.D. degrees from Jilin University, in 2006, 2008, and 2016, respectively. He is currently an Associate Professor with Northeast Electric Power University. His main research interests include weak signal detection and intelligent information processing.