IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 28, 2019, accepted September 10, 2019, date of publication September 18, 2019, date of current version October 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2942079

An Efficient Method for High-Speed Railway Dropper Fault Detection Based on Depthwise Separable Convolution

SHIWANG LIU, LONG YU, AND DONGKAI ZHANG
School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China

Corresponding author: Long Yu (yulong.swjtu@163.com)

This work was supported in part by the Key Projects of National Natural Science Foundation of China under Grant U1734202, and in part by the National Key Research and Development Plan of China under Grant 2017YFB1200802-12.

ABSTRACT The overhead contact system (OCS) is an indispensable part of high-speed railway power supply system. The stable operation of the OCS depends on the stability of the contact between the pantograph and catenary. The dropper is an important guarantee for the stable contact. The three faults of the dropper, such as slack, breakage and disappearance can damage the quality of traction power supply and reduce the safety of the railway operation. Therefore, it is necessary to efficiently detect the dropper fault and guide the maintenance in time. Recently, automatic dropper fault detection methods based on monitor-video have been introduced to improve railway operation safety. However, the existing methods were still not stable enough in complex backgrounds. To improve the accuracy and real-time performance of the dropper fault detection, this paper proposed a method combining depthwise separable convolution with object detection network to detect the dropper fault. The proposed method consists of two stages. First, a dropper progressive location network (DPLN) was adopted to obtain the dropper. The DPLN was mainly composed of a pantograph location network (PLN) and a dropper location network (DLN). Then, the dropper fault recognition network (DFRN) was used to recognize the type of dropper fault. The experimental results demonstrated the accuracy and real-time performance of the proposed method.

INDEX TERMS OCS, dropper location, fault recognition, depthwise separable convolution.

I. INTRODUCTION

The overhead contact system (OCS) is a key part of the high-speed railway traction power supply system, and it undertakes the task of supplying electrical energy to the locomotives. Generally, the normal operation of the OCS depends on the good sliding contact between the catenary and the pantograph, and the dropper is an important guarantee for good sliding contact. The slack, breakage and disappearance of dropper will damage the quality of traction power supply and reduce the safety of the railway operation. Therefore, the status of the dropper is one of the greatest concerned issues of railway companies. It is necessary to efficiently detect the dropper fault and guide the maintenance in time.

In the OCS, the number of dropper is very large. Due to the exposure to the natural environment all year round, the dropper is not only fragile but also difficult to maintain. Railway workers have been required to periodically patrol railway lines for the detection of dropper faults. However, this method is inefficient and difficult to meet the large-scale maintenance requirements of high-speed railway. Therefore, the research on the intelligent detection of the OCS has made great progress. Many scholars have proposed automatic detection equipment and intelligent detection methods for the OCS. These methods are mainly divided into visual detection based on traditional image processing and visual detection based on deep learning.

The representative method of visual detection based on traditional image processing is as follows. The authors in [1] used feature-based image matching techniques to measure contact network dynamic interlacing. The Maximum Stable Extremity Area technique [2] was used to recognize and visualize the geometric features of the surface defect area of the rail head in the orbital image. Some scholars had

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

used structured light technology [3], [4] to detect geometric parameters of OCS. Reference [5] used mean shift tracking and GMM technology to perform robust state monitoring of the pantograph-contact system, which can effectively detect the outbreak of the arc and the irregular location of the contact line. An edge detection method was proposed to obtain the edge of the pantograph, which had achieved good results in a simple environment [6].

The representative method of visual inspection based on deep learning is as follows. Reference [7] used R-CNN to obtain the image area of the isoelectric line, and then segmented the image based on the MRF model to realize the diagnosis of the loose wire of high-speed railway. The deep convolutional neural network was used to automatically detect fastener defects in the OCS support device [8]. Scholars used Faster R-CNN for new insulator location and surface defect recognition through multi-task neural network, which achieved good results [9].

Over the last few years, the dropper fault detection has been performed primarily by manual methods, and some scholars have also adopted visual inspection methods. Reference [10] obtained the location of the dropper by installing the prior knowledge of the dropper, and used MLP to classify dropper faults. Some researchers [11], [12] proposed ridge filter based on Hessian matrix for edge detection, then expanded corrosion treatment, and finally used Hough Transform for dropper fracture detection. The author in [13] used a faster R-CNN to locate the dropper image and then used the Hough Transform to recognize the dropper fault. However, the above studies are applied to the ideal scene or experimental environment, and cannot cope well with the dropper fault detection task in the actual situation, and cannot effectively carry out the dropper location and recognition.

In recent years, due to the excellent performance of deep learning, it has been widely used in computer vision. The researchers proposed a R-CNN object detection method based on Region Proposal, and improved on this basis, and then proposed a better Faster R-CNN network [14]–[16], but the network runs at a slower speed. For this reason, researchers proposed an end-to-end object detection network YOLO, combined with the FPN [17] multi-scale predictions, and then proposed the excellent performance of YOLOV3 [18], [19]. The above method can better accomplish the object detection task in complex scenes. However, the dropper is very small in the image, and it is still difficult to detect the dropper directly in the entire image.

This paper proposed an efficient method for the dropper fault detection. The first contribution is that a dropper progressive location network (DPLN) was proposed to obtain an accurate and real-time dropper location. The second contribution is that a dropper fault recognition network (DFRN) was proposed to recognize the type of dropper fault. The third contribution is that this paper used data enhancement, deep convolution generative adversarial networks and Focal loss to solve the problem of scarcity of dropper faulty images, which improved DFRN performance.

This paper was organized as follows. Section II analyzed the OCS and gave an OCS online monitor device. The method of the dropper fault detection was described in Section III. Section IV presented the experimental results and conclusions were given in Section V.

II. THE OCS ANALYSIS AND OOM DEVICE

The OCS is the only way for high-speed train to obtain stable power. Therefore, the reliability of the OCS is very important to ensure the long-term stable operation of the electrified railways. The OCS is composed of contact line, messenger and dropper. The OCS is shown in Fig.1. The messenger and the dropper are used to fix the contact line. In each span, the dropper increases the suspension point to the contact line without increasing the pillar, so that the sag and elasticity of the contact line are improved, and the working quality of the contact line is improved. In addition, by adjusting the length of dropper, the height of the contact line to the rail surface is ensured to meet the technical requirements. Therefore, it is necessary to detect the dropper fault.

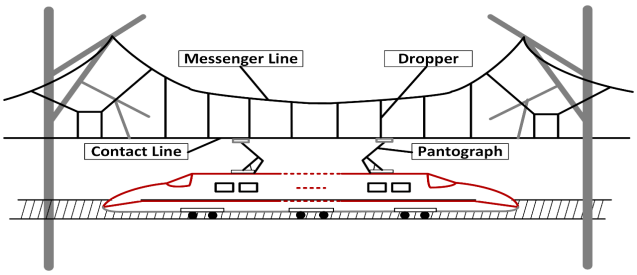


FIGURE 1. The overhead-contact system.

Due to the importance of the OCS, it is necessary to carry out scientifically monitoring and detection of its operating state. The automatic detection and online monitoring of the OCS operating state are realized by the OCS online monitor device, also known as OOM device. The OOM device can detect the state of the dropper through a set of global shutter cameras installed on the roof of the train, as shown in Fig.2.

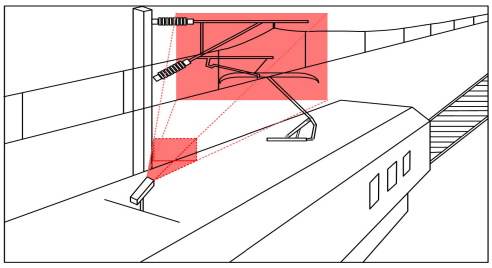


FIGURE 2. The OCS online monitor device.

The OOM device will capture the OCS images and the auxiliary information including mileage, station, etc. will also be stored in the index table of the corresponding OCS images. Due to the all-weather operation of the train, the background of the OCS image will dynamically change, such as

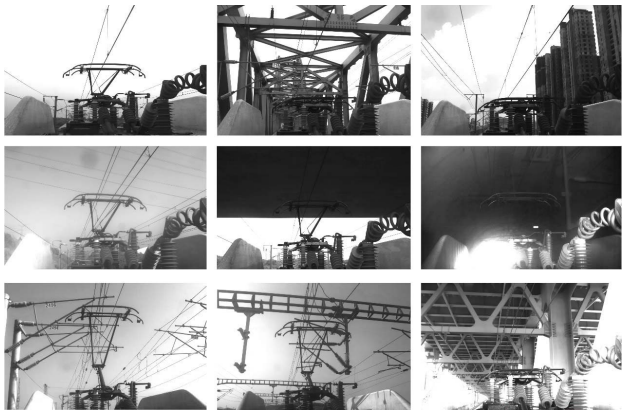


FIGURE 3. Typical images of OCS.

the tunnel environment, bridge background, building background, cloudy background, glare background and multiple complex backgrounds. The complex and variable background of OCS can reduce the accuracy of dropper fault detection. The typical images are shown in Fig.3.

III. THE DROPPER FAULT DETECTION METHOD

In order to improve the accuracy and real-time performance of the dropper fault detection in complex and variable background, this paper proposed an efficient dropper fault detection method based on depthwise separable convolution. The method is shown in Fig.4. The method consists of three stages: the pantograph location network (PLN), the dropper location network (DLN) and the dropper fault recognition network (DFRN). DPLN consists of PLN and DLN.

DPLN: Firstly, this paper used bilinear interpolation to modify the OCS image size to 320 × 320, and then the mean normalization process was performed. The image was input to the PLN to obtain the location of the pantograph, and the candidate location of the dropper was determined. This method input the candidate area image to DLN, and finally output the dropper image.

DFRN: This paper used a depthwise separable convolution module to construct a network to extract the characteristics of the dropper image, and then predicted the type of dropper fault through the fully connected layer and the Softmax layer, and finally obtained the result of the dropper fault detection. This paper used the deep convolution generative adversarial networks (DCGAN) and data enhancement to solve the problem of network over-fitting and low generalization performance caused by scarcity of dropper fault samples during network training.

A. DROPPER PROGRESSIVE LOCATION NETWORK

By analyzing the OCS image, it can be known that: First, the size of the OCS image captured by the camera was 1536 × 2048. The pixel area ratio of the dropper image was extremely small in the whole image, and the number of horizontal pixels was very small. Therefore, it was very difficult to extract the dropper image directly from the OCS image, and the processing speed will be low. Second, the installation rules of the OCS and the pantograph determined that the dropper image will appear in the rectangular area containing the image of the pantograph.

Object location based on deep learning is mainly divided into region proposal and end-to-end methods. For example, Faster R-CNN and YOLO, there is also a traditional object location method HOG. For the dropper location task, it is very difficult to extract the dropper directly from the whole image by using the above method. Therefore, based on the

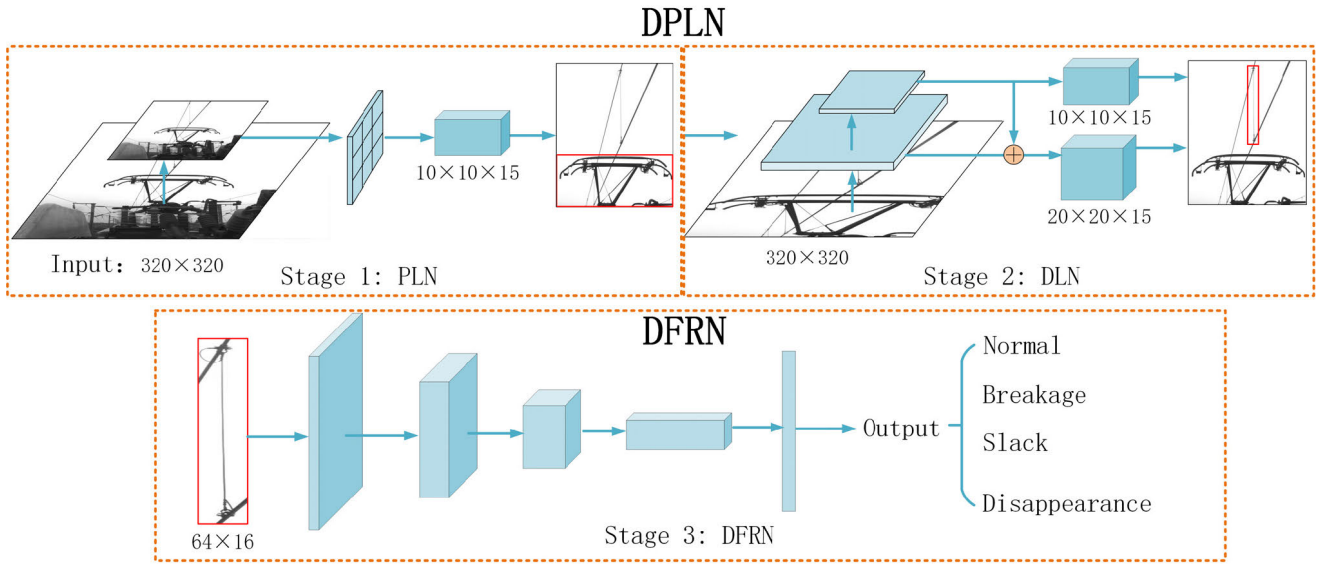


FIGURE 4. The flowchart of the dropper fault detection method.

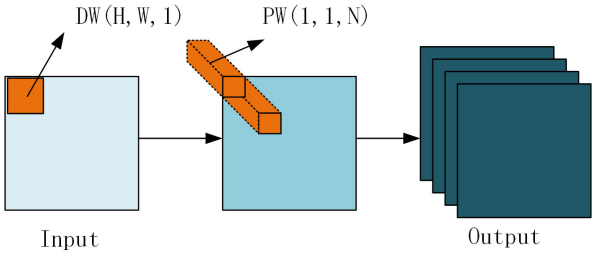


FIGURE 5. Depthwise separable convolution.

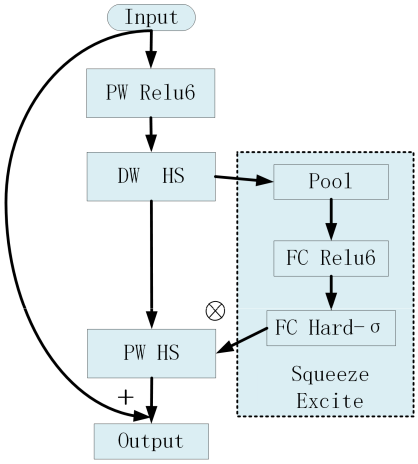


FIGURE 6. The OCS image feature extraction module (bneck block).

depthwise separable convolution [20] and the end-to-end object location structure, this paper proposed a dropper progressive location network (DPLN), which mainly included the pantograph location network (PLN) and the dropper location network (DLN). As shown in Fig.4.

Depthwise separable convolution (DWC) is made up of two parts: depthwise convolutions layer (DW) and pointwise convolutions layer (PW). As shown in Fig.5. The DW convolutes the input feature maps by channel-wise fashion and the PW convolution convolutes the feature maps of all channels with a 1×1 convolution kernel. Reference [21] verified the computational advantage of DWC by comparing the theoretical computation between traditional convolution and DWC.

The DPLN used MobileNet's excellent practice [22] to construct an OCS image feature extraction module (bneck block), as shown in the Fig.6.

The linear rectification function Relu6 (RE) or HS was used as the activation function of the network. The RE and HS function expressions are as follows.

$$RE = min\left(max\left(0, x\right), 6\right) \tag{1}$$

$$HS = x\frac{Relu6\left(x+3\right)}{6} \tag{2}$$

The residual connection was used to solve the problem of gradient dissipation and explosion of deep neural networks [23] and added a squeeze excitation module [24].

B. PANTOGRAPH LOCATION NETWORK

This paper used the PLN to locate the pantograph location. The network structure is shown in the Fig.7. The backbone network of the PLN was a pantograph image feature extraction network (PFEN). After inputting an image to PFEN, a matrix block with a length of 10, a width of 10, and a depth of 15 was obtained. In this paper, the end-to-end object location structure was used to map the 10×10 grid into the input image, and the image was divided into 10×10 grids. Each grid was responsible for predicting the pantograph center point to fall into the pantograph location within the grid. The pantograph location contained the center of the pantograph and the width and height of the pantograph image.

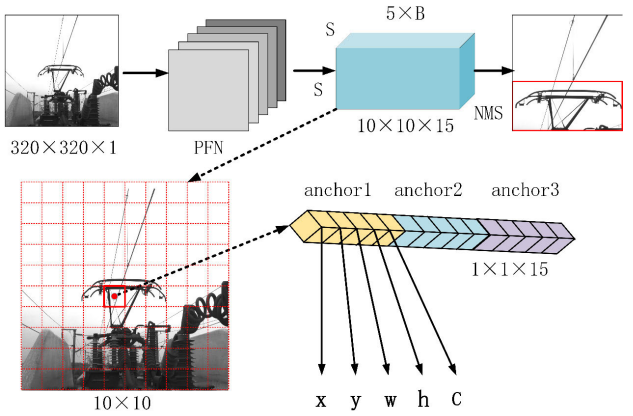


FIGURE 7. The pantograph location network (PLN).

For the three main forms of pantograph in the image sample, each grid was pre-set with three anchors: anchor1, anchor2, and anchor3, S×S represents the number of grids, B represents the number of anchors. Each anchor needs to return to the pantograph center point coordinates (x, y), width and height (w, h), confidence (C). Each anchor is responsible for locating objects with IOU greater than the threshold. C is equal to the probability (P) of the object contained in the anchor box multiplied by the IOU of the anchor box.

The FPEN was mainly composed of bneck block, and the detailed PFEN parameters are shown in the Table 1. Input represents the input of the upper layer, Operator represents the arithmetic module in the middle of the layer, Out represents the number of output channels, SE indicates whether the Squeeze Excitation module is used, NL indicates different activation modes, and S indicates the network step size.

After obtaining the confidence of each anchor, the threshold was set to filter out the anchor with low confidence, and the remaining anchor was subjected to non-maximum suppression processing to obtain the final locating result.

The PLN network loss function (L) mainly includes two parts. L1 is the location loss of the anchor, including the coordinate loss of the object center point and the loss of the object height and width; L2 is the loss of the confidence (C), including the presence or absence of the object in the grid. In order to make the model more stable, this paper added a

TABLE 1. Pantograph image feature extraction network (PFEN).

Input	Operator	Out	SE	NL	S
$320^2 \times 1$	conv2d 3×3	16	0	HS	2
$160^2 \times 16$	bneck block	16	1	RE	2
$80^2 \times 16$	bneck block	24	0	RE	2
$40^2 \times 24$	bneck block	24	0	RE	1
$40^2 \times 24$	bneck block	40	1	HS	2
$20^2 \times 48$	bneck block	96	1	HS	1
$20^2 \times 96$	bneck block	96	1	HS	2
$10^2 \times 96$	conv2d 1×1	15	0	RE	1

weighting coefficient of 5 to the anchor coordinate prediction loss, and reduced the confidence prediction loss of the anchor that does not contain the object. The loss weight coefficient was set to 0.5. In the formula, i represents the grid number, j represents the number of anchor prediction value. The PLN loss can be written as:

$$\lambda_{coord} = 5, \quad \lambda_{noobj} = 0.5$$

$$\begin{aligned} L_1 = {} & \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \bigcap_{ij}^{obj} \left(x_i - x_i^{\wedge}\right)^2 + \left(y_i - y_i^{\wedge}\right)^2 \\ & + \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \bigcap_{ij}^{obj} \left(\sqrt{w_i} - \sqrt{w_i^{\wedge}}\right)^2 \\ & + \left(\sqrt{h_i} - \sqrt{h_i^{\wedge}}\right)^2 \end{aligned} \tag{3}$$

$$\begin{aligned} L_2 = {} & \sum_{i=0}^{s^2} \sum_{j=0}^{B} \bigcap_{ij}^{obj} \left(C_i - C_i^{\wedge}\right)^2 \\ & + \lambda_{noobj} \sum_{i=0}^{s^2} \sum_{j=0}^{B} \bigcap_{ij}^{noobj} \left(C_i - C_i^{\wedge}\right)^2 \end{aligned} \tag{4}$$

$$L = L_1 + L_2 \tag{5}$$

C. DROPPER LOCATION NETWORK

After obtaining the location of the pantograph, according to the relationship of the OCS and the pantograph, the dropper and the contact line, and the dropper image can appear in the rectangular area containing the image of the pantograph. Therefore, according to the location of the pantograph, the candidate location of the dropper was determined, and the dropper location was performed in the area. The backbone network of the DLN was a dropper image feature extraction network (DFEN). The DLN used the same end-to-end object location structure, confidence calculation method and the loss function as PLN. The DLN is shown in the Fig.8, and the detailed DFEN parameters are shown in the Table 2.

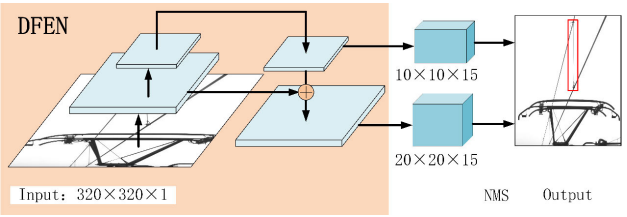


FIGURE 8. The dropper location network (DLN).

TABLE 2. Dropper image feature extraction network (DFEN).

Input	Operator	Out	SE	NL	S
$320^2 \times 1$	conv2d 3×3	16	0	HS	2
$160^2 \times 16$	bneck block	16	1	RE	2
$80^2 \times 16$	bneck block	24	0	RE	2
$40^2 \times 24$	bneck block	24	0	RE	1
$40^2 \times 24$	bneck block	40	1	HS	2
$20^2 \times 40$	bneck block	40	1	HS	1
$20^2 \times 40$	bneck block	48	1	HS	1
$20^2 \times 48$	bneck block	96	1	HS	1
$20^2 \times 96$	bneck block	96	1	HS	2
$10^2 \times 96$	bneck block	96	1	HS	1
$10^2 \times 96$	bneck block	192	1	HS	1
$10^2 \times 192$	conv2d 1×1	15	0	RE	1

The method adopted two-level feature layer, and extracts 20 × 20 and 10 × 10 layers from DFEN to perform dropper location at different scales. The 10 × 10 layer feature information was mainly for large object location, and the 20 × 20 layer consisted of its own layer plus the result of 10 × 10 up-sampling, mainly for small object locating. Finally, the dropper location result was output to the dropper recognition network for fault recognition. In this paper, two kinds of anchors with different sizes were set for different scale prediction layers, one for small object prediction and another for large object prediction, with a total of six anchors.

The DPLN can detect multiple droppers in the OCS image. This paper selects the dropper closer to the camera as the input of the dropper fault recognition network. This makes the dropper image clearer, saves computational resources, and can improve the recognition accuracy of the dropper fault. During the running of the railway, the dropper appears in front of the camera from far to near, and the dropper is closer to the upper edge of the image. As shown in Fig.9 and Fig.10. As the train moves forward, the center point of the dropper image (yellow box) is closer to the upper edge of the image. When the next dropper (blue box) appears in the image, this method filter it out according to the distance (red line). According to the difference between the single-track railway and the double-track railway, this method considers the situation in two cases, as shown by the algorithm: select dropper image.

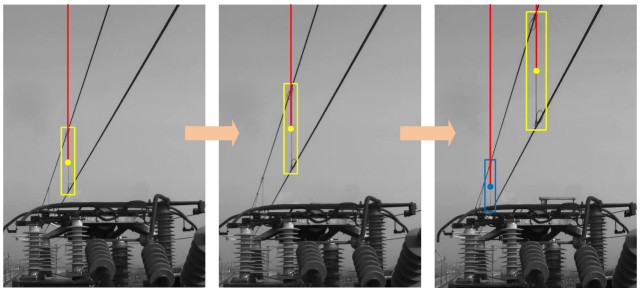


FIGURE 9. The location changes of the dropper in the single-track railway.

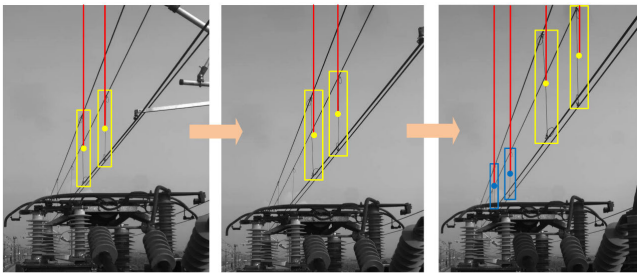


FIGURE 10. The location changes of the dropper in the double-track railway.

Algorithm 1 Select Dropper Image

Input:
(1) All the droppers detected by the DPLN
(2) Type of railway line: T
Output:
Dropper for the dropper fault recognition network

If T = single-track;
For i = 0:1:N(number of image frames);
  Calculate the distance from the dropper to the upper edge of the image;
  Save the closest dropper and the distance d(i);
End If
  d(i)< d(i+1) and d(i)< d(i−1);
Return the closest dropper of the i frame
If T = double-track;
For j = 0:1:N(number of image frames);
  Calculate the distance from the dropper to the upper edge of the image;
  Save the two closest droppers and the distance m(j), n(j);
End If
  m(j)+n(j)< m(j+1)+n(j+1) and m(j)+n(j)< m(j−1)+n(j−1);
Return the two closest droppers of the j frame

D. K-MEANS ANCHOR CLUSTER

In the conventional object location method, the anchor is obtained by a multi-scale sliding window traversal method or a selective search, and then locating is performed. However, this method runs slower and the effect is not the best. In this paper, the k-means clustering algorithm was used to analyze the image of the pantograph and the dropper, and the IOU was clustered as the distance to obtain the anchor size that matches the feature distribution of the pantograph and the dropper. The k-means clustering steps are as follows:

Step 1: The width and height of the real bounding box of the image sample were taken as a sample point (w_n, h_n), $n \in \{1, 2, ..., N\}$, and the center point of the object frame is (x_n, y_n), $n \in \{1, 2, ..., N\}$, and the sample points of all the images were composed into a set of image sample points.

Step 2: This paper randomly selected the K group points (W_m, H_m), $m \in \{1, 2, ..., k\}$ in the set as the cluster center.

Step 3: Calculated the distance (d) from all points in the sample point set to the K cluster centers, and assigned the sample points to the nearest cluster center to obtain K point clusters.

$$d = 1 - IOU\left[(x_n, y_n, w_n, w_n), (x_n, y_n, W_m, H_m)\right] \quad (6)$$

Step 4: Recalculated the center point of the cluster. The calculation formula is as follows, N_m indicates the number of object frames in the number m cluster.

$$\begin{cases} W_m^{\wedge} = \dfrac{1}{N_m} \sum w_m \\ H_m^{\wedge} = \dfrac{1}{N_m} \sum h_m \end{cases} \quad (7)$$

Step 5: Repeated steps 3 and 4 until the center of the cluster stopped moving and K parameters of the cluster center were obtained, which is the width and height of the anchor.

Because k-means algorithm is sensitive to the location of random initial point K, this paper solved it independently for three times, and chose its average mean as the final result, as the anchor value of PLN and DLN. It can be seen that Table 3 is composed of DLN anchors and PLN anchors. The length and width units of anchors is pixels, and the image size is 320 × 320. The DLN contain six anchors, and the PLN contain three anchors, which correspond to the candidate anchor for the pantograph location and the dropper location network. The approximate size distribution of the anchor is shown in Fig.11.

E. DROPPER FAULT RECOGNITION NETWORK

This paper proposed a dropper fault recognition network (DFRN) based on depthwise separable convolution. According to the China High-speed Railway OCS Operation and Maintenance Rules [25] and Reference [13], [14], the typical faults of the dropper include breakage, slack, and disappearance. DFRN can recognize the normal, breakage, slack and disappearance of the dropper. Focal Loss [26] was used as its loss function, and the loss function was optimized by SGD algorithm. Finally, the result was output through Softmax layer. The network structure is shown in Fig.12.

Through the analysis of samples, it can be known that the sample distribution was extremely unbalanced, and the number of faulty samples was far less than the normal number of samples.

TABLE 3. Anchors of dropper progressive location network.

Anchors	DLN anchors												PLN anchors					
Type	anchor1		anchor2		anchor3		anchor4		anchor5		anchor6		anchor7		anchor8		anchor9	
320^2	x1	y1	x2	y2	x3	y3	x4	y4	x5	y5	x6	y6	x7	y7	x8	y8	x9	y9
One	9	37	10	56	11	66	13	75	15	103	23	133	135	93	131	143	145	53
Two	8	34	11	59	12	64	13	79	16	101	21	132	133	104	130	145	141	50
Three	10	46	12	55	13	75	13	78	17	108	25	142	136	96	134	142	145	55
Mean	9	39	11	57	12	68	13	77	16	104	23	136	135	98	132	143	144	53

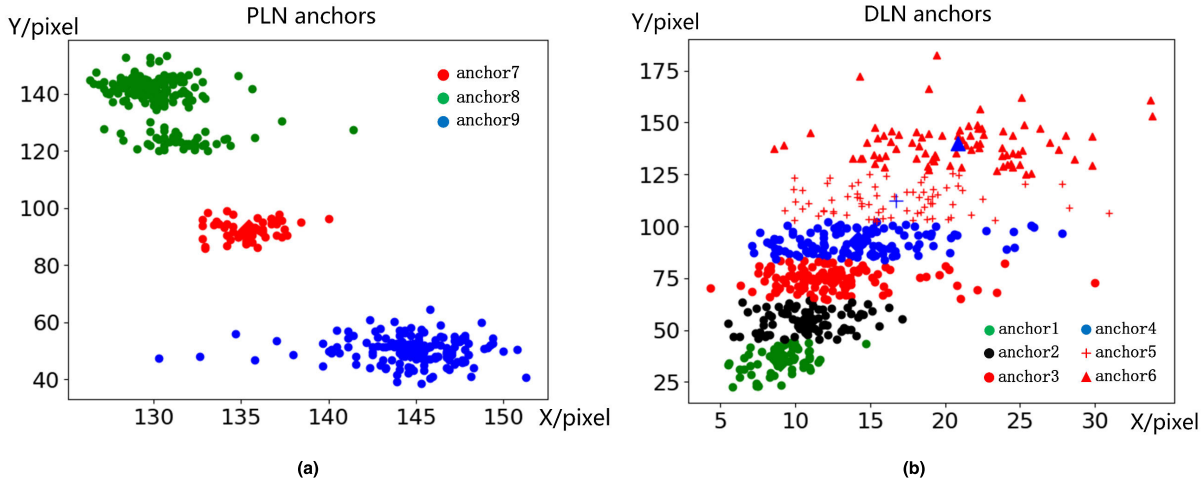


FIGURE 11. Size distribution of DPLN anchors: (a) size distribution of PLN anchors, (b) size distribution of DLN anchors. It is a schematic diagram of the anchor size distribution of our image. It corresponds to the size of the anchor in Table 3.

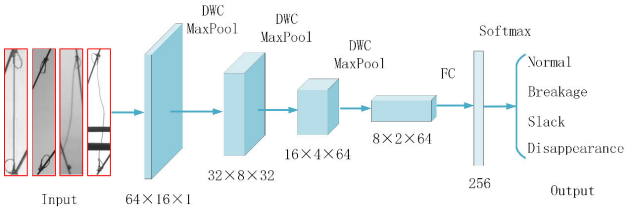


FIGURE 12. The dropper fault recognition network.

In order to solve the problem of data imbalance, in the data aspect, dropper fault samples were generated by data enhancement and DCGAN [27], [28]. The fault samples were enhanced by image mirroring, rotating, stretching and contrast changing. The faulty samples were generated by the DCGAN to make the distribution of the samples more balanced. In the algorithm level, the cross-entropy function was usually used as the loss function of the recognition network, and the minimization of the cross-entropy function was taken as the optimization objective of the model. However, the direction of network optimization for minimizing cross-entropy function was to minimize the loss of all types of samples. In this process, the loss of normal samples and fault samples was equal, which was not suitable for the imbalance of dropper sample distribution. Focal loss was proposed as the loss function of the network. The Focal loss function is as follows:

$$L_{fl} = \begin{cases} -\alpha \times y_1^{\gamma} \times logy_1 \\ -\beta \times y_2^{\gamma} \times logy_2 \\ -\chi \times y_3^{\gamma} \times logy_3 \\ -\delta \times y_4^{\gamma} \times logy_4 \end{cases}$$
$$y_1 + y_2 + y_3 + y_4 = 1 \tag{8}$$

L_{fl}: Represents the loss function of network

y_1, y_2, y_3, y_4: Represent the output values of the four states

This paper introduced $\alpha, \beta, \chi, \delta$ as loss weight coefficients of normal, breakage, slack and disappearance of dropper. This paper introduced the gamma coefficient to reduce the mining strength of difficult samples by reducing the balance network of normal sample loss weight coefficients. This paper made the network pay more attention to the characteristics of difficult samples and made the model more popular.

This paper used the depthwise separable convolution module to extract the dropper image features. After obtaining the dropper image from the DPLN, this paper used bilinear interpolation to modify the image size to 64 × 16, which met to the sample shape of the dropper image. In this paper, a momentum-driven optimizer SGD was used to optimize the

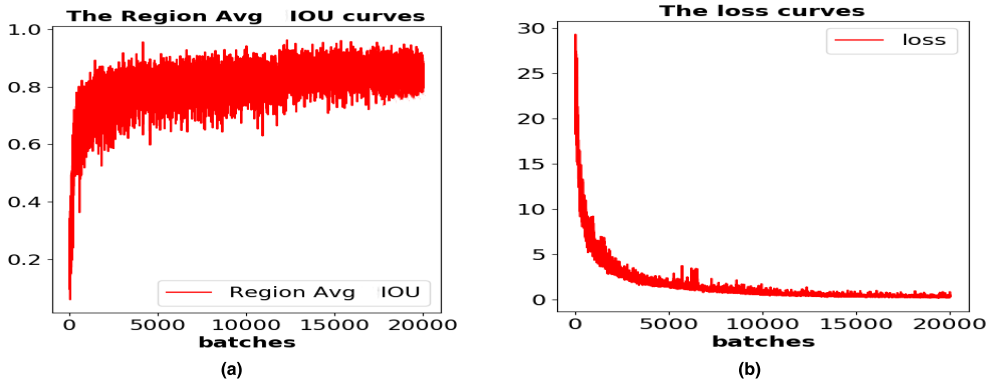


FIGURE 13. DPLN training curve: (a) the region AVG-IOU curves, (b) the loss curves.

TABLE 4. The experimental data.

Data set	OCS images		Dropper images	
Type of images	Normal	Extreme	Normal	Faulty
Number of images	69240	6211	89500	951+4755
Size of images	1536×2048		64×16	

TABLE 5. The results in the normal background.

Method	HOG	YoloV3	F-R	DPLN
Number of images	66000	66000	66000	66000
mAP(%)with IOU>0.6	79.42	93.20	**95.40**	95.32
mAP(%)with IOU>0.7	67.20	88.74	92.13	**93.15**
mAP(%)with IOU>0.8	58.80	83.42	88.63	**90.48**
Time(ms)	107.6	112.8	632.7	**50.4**

loss function Focal loss, so that the loss function approached the global minimum.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

This paper used OOM equipment to acquire 75451 OCS images with a size of 1536 × 2048, then guided by a professional railway maintenance worker, and used the LabelImg tool for data labeling. It included 69240 images of normal background and 6211 images of extremely complex background. It contained more than 90000 droppers, about 89500 normal droppers, and 951 fault droppers. This paper used data enhancement and DCGAN to generate 4755 fault droppers, a total of 95206 pieces of experimental images. As shown in the Table 4. Detailed data on location and recognition experiments were described below.

The experimental environment was described as follows: Deep learning open source framework TensorFlow, Ubuntu 16.04, the embedded artificial intelligence platform NVIDIA JetsonTX2 was used as the platform for experiment to ensure that the method can still guarantee good real-time performance on embedded devices.

A. DROPPER LOCATION EXPERIMENT RESULTS

This paper used 70451 OCS images as a testing dataset and 5000 images as a training dataset to perform dropper progressive location network training. This paper used the mAP to evaluate the accuracy of the method. The formula is as follows:

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$mAP = \int_0^1 P(R)dR \tag{11}$$

Among them, Recall is the recall rate (R), Precision is the accuracy rate (P), TP is the number of droppers that are correctly located by the network. FN is the number of droppers that are not correctly located by the network, FP is the number of incorrectly located droppers, TP+FN is the total number of actual droppers, TP+FP is the total number of droppers that the network has located.

In the dropper location experiment, 5,000 images were used for training. The number of batches is 32 and a total of 20000 batches were trained. During the training process, the IOU and the loss had been constantly changing. The parameters of the training process are shown in Fig.13.

70000 OCS images of the testing dataset were divided into 66000 normal backgrounds and 4000 extremely complex backgrounds. DPLN and YoloV3, Faster-RCNN (F-R) and HOG [28] were compared under the same conditions. The results in the normal background are shown in Table 5. This paper used PLN to find the candidate region for the dropper location and perform the DLN in this region, which filtered out the other interference outside the region. DPLN's location accuracy is slightly better than F-R, and HOG has the worst performance. For practical applications, real-time performance is very important in the dropper fault detection task. DPLN takes an average of 50.4ms for each image, and

FIGURE 14. Failure in the extreme background.

TABLE 6. The results in the extreme background.

Method	HOG	YoloV3	F-R	DPLN
Number of images	4000	4000	4000	4000
mAP(%)with IOU>0.6	61.12	82.40	87.84	**89.10**
mAP(%)with IOU>0.7	47.20	75.80	79.13	**80.64**
mAP(%)with IOU>0.8	28.80	61.28	69.90	**72.40**
Time(ms)	112.5	112.5	633.2	**51.0**

TABLE 7. Comparison of dropper fault location.

Type of fault		Slack	Breakage	Disappearance
Number of images		210	145	96
mAP(%) with IOU>0.7	DPLN	**85.58**	**84.80**	81.79
	F-R	84.10	83.72	**81.90**
	YoloV3	79.90	79.49	77.10
	HOG	52.30	44.50	35.00

DPLN takes the least amount of time, while F-R takes about 12 times. The results in the extremely complex background are shown in Table 6. Due to the influence of tunnel environment, bridge background and architectural background, and multiple complex backgrounds, the accuracy of the network had been significantly reduced. There was even a situation where it could not be located, as shown in Fig.14. This paper conducted a dropper location comparison experiment on the faulty dropper. The results are shown in Table 7. Due to the scarcity of the faulty sample, and the dropper characteristics under the type of disappearance are not obvious enough, and the network mainly relies on the feature of the dropper clip for location, it can be seen that the mAP of the disappearance is lower than the others.

In the case where the IOU is greater than 0.7, the AP curve of all image backgrounds is shown in Fig.15. The comparison of location results is shown in Fig.16, the red box is the result of the location, and the blue box is the candidate region for the dropper. Experiments showed that DPLN performed better than other methods.

After the DLN obtained the dropper from the OCS image, this paper used the algorithm (Select Dropper Image) to select the dropper and used 10,000 images for testing, including 6240 single-track images and 3760 double-track images. The experimental results showed that the accuracy of the

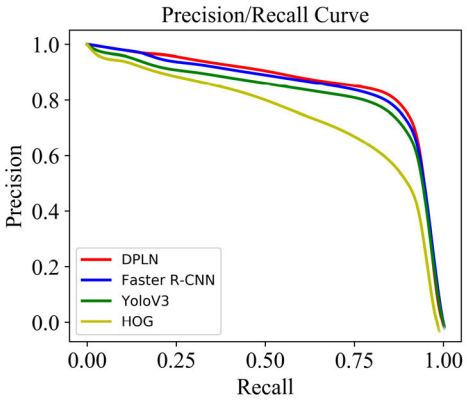


FIGURE 15. The AP curve of all image backgrounds (IOU>0.7).

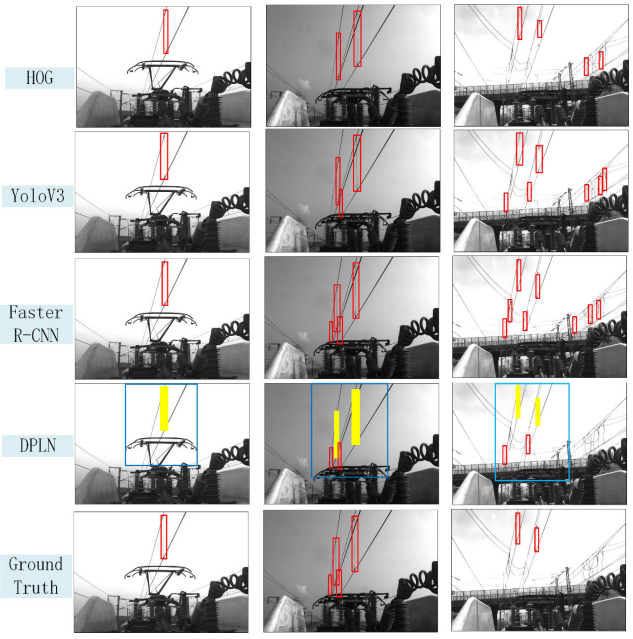


FIGURE 16. A comparison of the DPLN and YoloV3, Faster-RCNN, HOG locating results.

algorithm is 99.21%. As shown in Fig.16, the yellow areas are the selected dropper images, which are used as input to the dropper fault recognition network.

B. FAULT RECOGNITION EXPERIMENT RESULTS

This paper used 91000 images as the testing dataset to verify the accuracy of the algorithm, and used 4206 images as the training dataset. The training dataset contained 1500 normal, 1306 slack, 818 breakage, and 582 disappearance images. Accuracy of DFRN is equal to the number of correctly recognized dropper fault (Correct Numbers) divided by the total number of droppers (Total Numbers).

$$Accuracy = \frac{Correct\ Numbers}{Total\ Numbers} \tag{12}$$

In the original data, there were only 951 dropper fault samples, and 4755 fault samples were generated by data

TABLE 8. Results of dropper fault recognition.

Background of image		Simple background				Complex background			
Type of fault		Normal	Slack	Breakage	Disappearance	Normal	Slack	Breakage	Disappearance
Number of images		40000	822	420	368	48000	578	580	232
Accuracy (%)	DFRN	96.32	93.92	93.33	94.30	93.92	90.66	89.66	**92.67**
	VGG	**97.10**	**94.53**	**95.23**	**94.57**	**94.62**	**91.18**	**91.03**	91.81
	ResNet	95.88	93.31	92.38	93.21	93.50	89.79	88.45	90.95

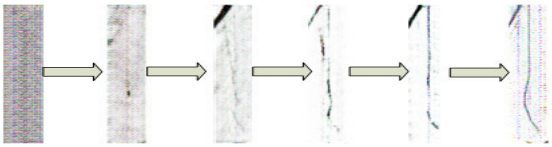


FIGURE 17. The process of generating samples through DCGAN.

enhancement and DCGAN, which can effectively improve the generalization ability of the DFRN and prevent model over-fitting. The process of generating samples through DCGAN is shown in Fig.17.

During the training process, the number of batches is 64, and the initial learning rate (LR) was set to 0.005. As the number of training steps increased, the learning rate was dynamically attenuated, the attenuation coefficient (DecayRate) was 0.9, and the dynamic learning rate was (UpdateLR), the attenuation expression is as follow:

$$UpdateLR = LR \times DecayRate^{Step} \tag{13}$$

In the training process, this paper used 20% of the 4206 samples as the validation set in the network training process. After 1400 trainings had been carried out, the accuracy rate of validation set was about 96%. This paper compared with the classical image recognition network VGG [30] and ResNet [31] algorithms, and recognized the dropper fault under simple background and complex background. Results of dropper fault recognition are shown in Table 8. Since the dropper fault recognition task is relatively easy, it can be seen that the classification accuracy of the three methods is almost the same. The recognition accuracy of the three types of faults is lower than the normal type. Since the faulty samples are less than the normal samples, it is easier to classify the normal type.

This paper used data collected from real-running trains for algorithm testing, and compared the average accuracy and real-time performance with the classical algorithms of deep learning image recognition VGG and ResNet. The results are shown in Table 9. The average accuracy of VGG is 95.66%, which is slightly higher than DFRN, the average accuracy of DPRN is 94.92%, and the ResNet accuracy is 94.47%. DFRN takes 12.27ms to process each image. Because this method used a more efficient network, although the accuracy is not the highest, the real-time performance has been greatly improved.

TABLE 9. Comparison of fault recognition methods.

Method	Accuracy（%）	Time（ms）
VGG	**95.66**	122.80
ResNet	94.47	34.54
DFRN	94.92	**12.27**

C. LIMITATIONS

There are some limitations of this study and further research is still needed. The first is that the complex background image results in a lower accuracy of the dropper progressive location network, and the accuracy of dropper fault recognition in complex backgrounds is lower than in a simple background. In extreme cases, such as tunnels and bridges, this method is completely unable to locate the dropper, as shown in Fig.14. The second is the scarcity of the dropper fault images, which makes it difficult to further improve the accuracy of fault detection. Due to the scarcity of fault data, it is difficult to unify location and classification tasks into one network, and the accuracy of location and recognition is low. Our method needs more fault data to train and enhance the generalization performance of the model.

V. CONCLUSION

This paper proposed a dropper progressive location network (DPLN) to locate the dropper. First, the pantograph location network was used to locate the pantograph, and the candidate region of the dropper was determined. Then, this paper used the dropper location network to obtain the dropper image. The dropper fault recognition network (DFRN) was used to recognize four types of droppers: normal, slack, breakage, and disappearance. In this paper, the depthwise separable convolution and end-to-end object location structure were used to construct the DPLN and DFRN. The data enhancement, DCGAN and Focal loss were used to effectively solve the problem of unbalanced sample types during training. The experimental results proved that the proposed method can effectively solve the problem of dropper fault

detection in the actual scene. Further research will be required to improve the accuracy of dropper fault detection in complex backgrounds. When there is a large amount of fault data in the future, further research is needed to unify the location and classification tasks into one network.

REFERENCES

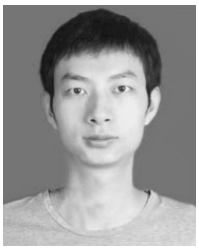
[1] C. J. Cho and H. Ko, "Video-based dynamic stagger measurement of railway overhead power lines using rotation-invariant feature matching," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1294–1304, Jun. 2015.
[2] A. K. Dubey and Z. A. Jaffery, "Maximally stable extremal region marking-based railway track surface defect sensing," *IEEE Sensors J.*, vol. 16, no. 24, pp. 9047–9052, Dec. 2016.
[3] D. Zhan, D. Jing, M. Wu, D. Zhang, L. Yu, and T. Chen, "An accurate and efficient vision measurement approach for railway catenary geometry parameters," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 12, pp. 2841–2853, Dec. 2018.
[4] Z. Liu, W. Liu, and Z. Han, "A high-precision detection approach for catenary geometry parameters of electrical railway," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1798–1808, Jul. 2017.
[5] I. Aydin, M. Karaköse, and E. Akin, "A robust anomaly detection in pantograph-catenary system based on mean-shift tracking and foreground detection," in *Proc. IEEE SMC*, Manchester, U.K., Oct. 2013, pp. 4444–4449.
[6] I. Aydin, M. Karakose, and E. Akin, "A new contactless fault diagnosis approach for pantograph-catenary system using pattern recognition and image processing methods," *Adv. Elect. Comput. Eng.*, vol. 14, no. 3, pp. 79–88, Aug. 2014.
[7] Z. Liu, L. Wang, C. Li, and Z. Han, "A high-precision loose strands diagnosis approach for isoelectric line in high-speed railway," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1067–1077, Mar. 2017.
[8] J. Chen, Z. Liu, H. Wang, A. Nunez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 257–269, Feb. 2018.
[9] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2679–2690, Aug. 2019.
[10] C. Petitjean, L. Heutte, R. Kouadio, and V. Delcourt, "A top-down approach for automatic dropper extraction in catenary scenes," in *Proc. IbPRIA*, Póvoa de Varzim, Portugal, 2009, pp. 225–232.
[11] R. Bai, "Detection and recognition of catenary string and pantograph skateboard based on image processing technology," M.S. thesis, Dept. Control Eng., SWJT Univ., Chengdu, China, 2017.
[12] G. Wu, "Fault Detection Algorithm of Support and Suspension Devices of Catenary in High-speed Railway," M.S. thesis, Dept. Elect. Eng., ST Univ., Shijiazhuang, China, 2016.
[13] Y. Xu, "Application of image processing in detecting defects of catenary hanger," M.S. thesis, Dept. Elect. Eng., SWJT Univ., Chengdu, China, 2018.
[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.
[15] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.
[16] 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. 2016.
[17] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.
[18] 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.*, Jun. 2016, pp. 779–788.
[19] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," Apr. 2018, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767
[20] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jun. 2017, pp. 1800–1807.
[21] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," Apr. 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861
[22] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," May 2019, *arXiv:1905.02244*. [Online]. Available: https://arxiv.org/abs/1905.02244
[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520.
[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.
[25] *China High-speed Railway OCS Operation and Maintenance Rules, China Railway*, Standard 362, 2015.
[26] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2999–3007.
[27] J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2672–2680.
[28] Y. Yu, Z. Gong, P. Zhong, and J. Shan, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2016, pp. 97–108.
[29] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, vol. 1, Jun. 2005, pp. 886–893.
[30] K. Simonyan and A. Zisserman, "Deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1–14.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

SHIWANG LIU is currently pursuing the master's degree in electrical engineering with Southwest Jiaotong University.

His current research interests include vision inspection technology, fault detection, image processing, deep learning, and their application in the railway industry.

LONG YU received the Ph.D. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2008.

He is currently an Associate Professor with the School of Electrical Engineering, Southwest Jiaotong University. His current research interests include vision measurement, machine learning, computer vision, and their application in the railway industry.



DONGKAI ZHANG received the M.S. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, where he is currently pursuing the Ph.D. degree in electrical engineering.

His current research interests include vision inspection technology, fault diagnosis, image processing, and machine learning.