

Received July 6, 2021, accepted August 4, 2021, date of publication August 17, 2021, date of current version August 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3105419

Railway Insulator Detection Based on Adaptive Cascaded Convolutional Neural Network

ZAIXING WANG, XIAOZHONG LIU¹, HUAYI PENG¹, LIJUN ZHENG¹, JINHUI GAO,
AND YUFAN BAO

School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730073, China

Corresponding authors: Zaixing Wang (55232793@qq.com) and Xiaozhong Liu (rossi.liu66@gmail.com)

ABSTRACT Insulator failure is one of the important causes of railway power transmission accidents. In the automatic detection system of railway insulators, the detection and classification of insulator faults is a challenging task due to the complex background, small insulators and unobvious failures. In this article, we propose a railway insulator fault detection network based on convolutional neural network, which can detect faulty insulators from images with high resolution and complex background. The insulator fault detection network realizes the position detection and fault classification of the insulator by cascading the detection network and the fault classification network. The method of cascading two networks can reduce the amount of network calculations and improve the accuracy of fault classification. The insulator detection network uses low-resolution images for position detection, and this method can prevent the detection network from paying too much attention to the details of the image, thereby reducing the amount of network calculations. The fault classification network uses high-resolution insulator images for fault classification. The high-resolution images in this method have rich detailed information, which helps to improve the accuracy of fault classification. The trained insulator detection network and the fault classification network are cascaded to form an insulator fault detection network. The precision, recall and mAP values of the insulator fault detection network are 94.10%, 92.88% and 93.46% respectively. Experiment shows that this network cascading method can significantly improve the accuracy and robustness of insulator fault detection.

INDEX TERMS Convolutional neural network, target detection, railway insulator, fault classification, cascade network.

I. INTRODUCTION

The insulator is located between the arm and the pillar in the railway catenary, and has been exposed to the atmosphere for a long time; it not only has to withstand wind and sun, but also withstand strong electric field and strong mechanical stress. Therefore, insulators are prone to failure, resulting in reduced insulation strength, which threatens the normal operation of railway electrical systems. According to statistics from relevant national departments, railway accidents caused by insulator failures have increased year by year, and have become the main factor leading to potential safety hazards in the power system [1], so it is very important to detect insulators. At present, the detection methods of insulators in railways can be divided into two categories: power detection and non-power detection [2]. The principle of power

detection is to detect the leakage current of the insulator, and use electrical methods to detect whether the insulator has leakage current. This detection method is susceptible to electromagnetic interference caused by arcs, which affects the accuracy of judgment. The non-electricity detection is based on the image information of the insulator, and the method for processing image is used to detect the faulty insulator. The advantages of this type of detection method are non-contact, fast response and good linearity. The insulator detection method in this article belongs to the non-electricity detection method.

The non-electricity detection method uses image processing technology to extract the characteristics of the insulator, such as color, texture, shape, etc., to distinguish the insulator and the background in the image. Zhang *et al.* [3] used HIS model to identify tempered glass insulators and classify them. Bharata Reddy *et al.* [4] used Discrete Orthogonal S-Transform (DOST) and Adaptive Network-based Fuzzy

The associate editor coordinating the review of this manuscript and approving it for publication was Fanbiao Li¹.

Inference System (ANFIS) to identify the position of the insulator and extract the color features, and then use K-means clustering and Support Vector Machine (SVM) to classify insulators. Li *et al.* [5] proposed a method for detecting insulators based on texture. The method first uses contour projection to search for the position of the insulator in the image, then obtains the characteristics of the insulator from the image, and uses SVM for classification. Oberweger *et al.* [6] proposed an algorithm based on local gradient descriptor and local voting mechanism to detect insulators. The algorithm also uses descriptors supported by elliptic space to check whether the insulators are faulty. Wu *et al.* [7] proposed a texture segmentation algorithm. The method first divides the insulator image into multiple sub-regions with smooth contours, and then uses Grey-Level Co-occurrence Matrix (GLCM) to extract the texture features of the insulator, with the contour of the insulator segmented by Grey level co-occurrence integrated algorithm (GLCIA). Wang *et al.* [8] proposed an insulator identification method that combines shape, color and texture features. This method uses dominant color components to identify insulators and detects insulator drop-off defects through texture features. The above methods all require artificial design of insulator feature extractors, whose performance depends on the complexity of the background and the image quality, and is not very robust to complex and changeable background environments. For different types of insulators, corresponding algorithms need to be designed, and more parameters need to be adjusted when optimizing the algorithm.

Compared with traditional image processing algorithms, deep learning technology has the characteristics of automatic feature extraction, strong adaptability and high upper limit. Among various deep learning techniques, the Convolutional Neural Networks (CNN) have good translation invariance and robustness, so they are widely used in tasks such as target detection and image segmentation. Many models based on convolutional neural networks have achieved good results in target detection tasks, such as Faster R-CNN [9], YOLO [10]–[12], and SSD [13]. These models have also been successfully applied to areas such as autonomous driving, face recognition and cell detection, and achieved good results. At present, some researches have used the classic convolutional neural network model for insulator fault detection. Zhang *et al.* [14] proposed a catenary insulator detection model based on the GAN model. It can be successfully trained with only normal insulator images. Detect surface defects of real insulators. In the method proposed by Varghese *et al.* [15], the GoogLeNet pre-trained model is used for wire, tower and insulator detection, and the output of the CNN model is processed by the method of spectral clustering. Gao [16] proposed a CNN model that combines target detection and image segmentation to segment and classify insulator images from transmission tower images. Liu *et al.* [17] combined the location information of the catenary support components with the improved SSD network and proposed a new detection method for the

catenary support components to achieve rapid positioning of 12 components in the catenary. Kang *et al.* [18] proposed a new insulator detection model based on the Faster R-CNN model and deep multitask neural network (DMNN), which can simultaneously segment insulators and indicate defect detection. Huo *et al.* [19] improved the accuracy of catenary insulator detection by adding deconvolution to faster R-CNN. Insulators in railway catenary image data have the characteristics of complex background, high damage rate and unobvious failures, which increase the difficulty of detecting insulator failures. And image target recognition algorithms based on convolutional neural networks are all general algorithms. If you want to better complete the task of detecting catenary insulators, the algorithms need to be improved.

In this article, a cascaded split detection network (CSDN) for railway insulator detection and classification based on convolutional neural network is established, which is mainly composed of insulator detection network and fault classification network. The insulator detection network merges the feature maps with different semantic information through the feature fusion module [20], and then multiplies the fused feature map with the attention map through the multi-region adaptive module, so that the features in the feature map are automatically enhanced [21]–[23], and finally the RPN module [9] predicts the position information of the insulator. In the fault classification network, we use the location information of the insulators and the original contact net to obtain high-definition insulator images, and then use the vgg16 network to extract insulator features and classify the insulator status. The network proposed in this paper is divided into two steps during training. In the first step, the detection network is trained using the detection data set, after the training is completed, the detection network can accurately detect the position of the insulator. In the second step, the classification data set is used to train the fault classification network; after the training is completed, the fault classification network can correctly classify the insulators.

II. METHODOLOGY

The CSDN model we designed is shown in Figure 1, which consists of a detection network and a classification network. The detection network mainly includes three modules: a multi-layer fusion module, a multi-area self-adaptation module, and a proposed area network. The detection network realizes the output of the insulator position through the function of different modules. The classification network is composed of 13 layers of convolutional layers and 3 layers of fully connected layers. It performs feature extraction through the convolutional layers to realize the classification of insulator states.

When the detection network and classification network are both trained, they are cascaded together to form our CSDN model. The detection process of the CSDN model is shown in Figure 1. First, input the reduced catenary image into the detection network, and output the position information

of the insulator. Secondly, input the $W \times H$ catenary image and insulator position information into the crop layer, output the cropped insulator image, and fill the insulator image with black. Finally, input the cropped insulator image into the classification network, output the predicted state of the insulator, and draw the final result map.

In the CSDN model, the detection network uses the reduced catenary image to detect the position of the insulator, which greatly reduces the amount of calculation of the detection network and improves the detection speed. In the classification network, the classification accuracy is improved by using insulator images with rich details. The CSDN model realizes fast and accurate insulator position positioning and status classification by connecting the trained detection network and the classification network in series.

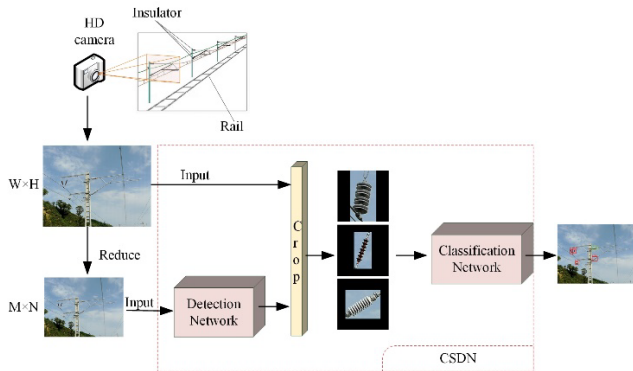


FIGURE 1. CSDN model testing process.

A. DETECTION NETWORK

As shown in Figure 2, the detection network is composed of a multi-layer fusion module, a multi-area adaptive module and an RPN module, and its function is to detect the position of the insulator. The multi-layer fusion module is composed of 5 blocks and a fusion layer, and its function is to synthesize and output image semantic information of different levels output by different blocks to form a feature map with rich feature information. The multi-region adaptive module is composed of a down-sampling layer, a convolutional layer, and an up-sampling layer, and its function is to adaptively strengthen the feature map output by the multi-layer fusion module, so that the insulator characteristics in the image are enhanced. The RPN module is composed of RPN, ROI and FC, and its function is to judge the position of insulator in the feature map, so that the detection network can accurately locate the insulator.

1) MULTI-LAYER FUSION MODULE

The function of the multi-layer fusion module is to enrich the information of the feature map and improve the performance of the detection network. As shown in Figure 2, the multi-layer fusion module structure consists of five Blocks and a concatenate layer. Block1 and Block2 are

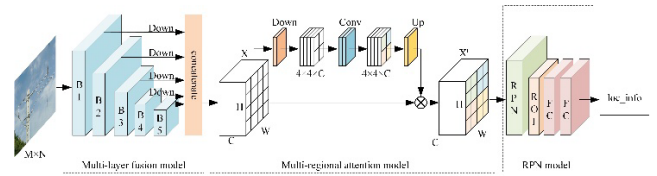


FIGURE 2. Schematic diagram of detection network. B1 represents block1, B2 represents block2 and so on, Down represents down-sampling, X represents feature map, X' represents feature map after multi-region adaptation, C represents the channel number of feature map, H represents the height of feature map, W represents the width of the feature map, \otimes Represents multiply pixel by pixel, loc_info represents location information.

composed of two convolution layers and a maximum pooling layer respectively, Block3 and Block4 are composed of three convolutional layers and a maximum pooling layer respectively, and Block5 is composed of two convolutional layers. In the multi-layer fusion module, the size of the convolution kernel is set to 3×3 , the step size is set to 1, the pooling layer size is both 2×2 , and the step size is 2. When detecting, first input the reduced contact net image to the block for convolution operation. Then, the output characteristic maps of Block1 to Block4 are down-sampled respectively. Finally, the different features are merged together in the concatenate layer.

All the convolution kernels in the multi-layer fusion module are smaller in size. Such a small-sized convolution kernel has the characteristics of fewer parameters and less calculation, which can increase the speed of the detection network. In the convolutional neural network, the abstraction degree of features extracted from different network depths is different. When the network is shallow, features such as texture and details can be extracted, and when the network is deep, features such as contour shape can be extracted [24]. In the multi-layer fusion module, we merge the outputs of different depth networks together to enrich the semantic information contained in the feature map. In order to maintain the consistency of the feature map size before the feature map fusion, we down-sample the output feature maps of Block1 to Block4. In network training, migration learning can reduce the time of network training [25], so the weight of the vgg16 pre-trained model on ImageNet is used when initializing the convolutional layer in the multi-layer fusion module [26].

2) MULTI-REGION ADAPTIVE MODULE

The function of the multi-region adaptive module is to make multiple regions in the feature map adaptively optimize the features. The structure can be seen from Figure 2, which consists of a down-sampling layer, two convolutional layers with a size of 1×1 , and an up-sampling layer. When detecting, the feature map x with the size of $W \times H \times C$ is first input into the down-sampling layer, and after the processing of the down-sampling layer, the size of the feature map x becomes $4 \times 4 \times C$. Then, the output feature map of the down-sampling

layer is sent to the two convolutional layers with a size of 1×1 to calculate the attention map. Sigmoid function and ReLU function are used as activation functions respectively in the first and second convolution layers. Finally, the calculated attention map is multiplied pixel by pixel with the feature map x after the up-sampling layer.

When calculating the attention map, we use the average pooling layer to perform average pooling in multiple regions of the feature map. Compared with global average pooling and global maximum pooling [27], [28], multi-region average pooling can retain more details in the feature map. In order to ensure the size consistency of the attention map and the feature map when multiplied, we up-sample the attention map.

3) RPN MODULE

The function of the RPN module is to propose a region proposal for the feature map for the position regression of the insulator. The structure can be seen from Figure 2, which consists of an RPN, a layer of ROI and two fully connected layers. In the RPN module, the cross-entropy loss function is used as the loss function to distinguish anchors from the foreground or the background. The expression is as follows.

$$L_{cls}p(p_i^*, p_i) = -\frac{1}{N} \sum_i [p_i^* \log(p_i) + (1-p_i^*) \log(1-p_i)] \quad (1)$$

In formula (1), p_i^* represents the probability of positive and negative samples, p_i^* is 1 for positive p_i^* samples, and 0 for negative samples. p_i Represents the probability of predicting that the anchor belongs to the foreground. The loss function expression used when returning from the anchor to the precise region proposal is as follows.

$$L_{reg}p(p_i^*, t_i, t_i^*) = \frac{1}{N} \sum_i [p_i^* L_{reg}(t_i, t_i^*)] \quad (2)$$

In formula (2), p_i^* represents the probability of positive and negative samples, and it is 1 for positive samples and 0 for negative samples. Only positive samples return to the bounding box. t_i Represents the predicted Bounding Box coordinate, and t_i^* the coordinate of Ground Truth Bounding Box. $L_{reg}()$ represents the Smooth L1 loss function, and the expression is as follows.

$$L_{reg}(t_i, t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2 & |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5 & otherwise \end{cases} \quad (3)$$

The Smooth L1 loss function is used when calculating the loss of the prediction box. The Smooth L1 loss function has the following two characteristics: (a) When the difference between prediction box and Ground truth is too large, the gradient value will not be too large; (b) When the difference between the prediction box and Ground truth is small, the gradient value is small enough. Convolutional neural network training requires stability and can be trained to high accuracy. The gradient value is an important factor that affects the

stability of network training. If the gradient value is too large, the training will be unstable and it is difficult to achieve high accuracy. So based on our requirements for network training and the characteristics of the Smooth L1 loss function, we use the Smooth L1 loss function when calculating the loss of the prediction box

B. FAULT CLASSIFICATION NETWORK

The function of the classification network is to classify the status of the insulators. In this network, the status of the insulators can be divided into: normal, damaged and missing. The structure of the classification network is shown in Figure 3, which is divided into a clipping layer, a feature extraction layer and a prediction layer. The output of the catenary image and the detection network (the position information of insulators) is used as the input of crop layer. In the crop layer, the position information of the insulator is amplified first, and then projected into the catenary image to intercept the insulator image, and then the insulator image is adjusted to 360×360 pixel size using the method of filling, and then input into the feature extraction layer. In the feature extraction layer, Vgg16 network [29] is used to extract the insulator features and classify the insulator state. There are three inputs in the Prediction layer: insulation position information, insulator state classification and catenary image. The output of the predict layer is a catenary image marked with insulator position and insulator state information.

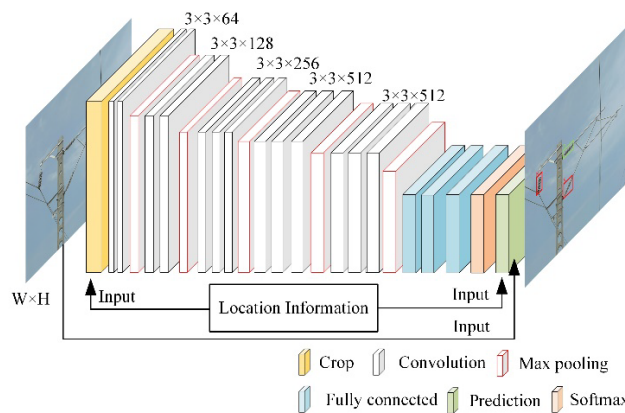


FIGURE 3. Schematic diagram of fault classification network.

Joining the classification network can improve the accuracy of insulator state classification. The input image of the classification network is a high-definition insulator image, which contains rich detailed information of the insulator and is beneficial to improve the accuracy of the network. In the classification network, the vgg16 network with a deeper network structure is used to extract the insulator features. The extracted features contain high-level semantic information such as contours and shapes, which are beneficial to image classification.

When training the fault classification network, vgg16 is trained using the classification data set alone. After the

training is completed, the crop layer and prediction layer are added. The cross-entropy loss function is used as the loss function in the fault classification network, and its function expression is:

$$L = -\frac{1}{N} \sum_{i \in M} [y_i \log p_i + (1 - y_i) \log (1 - p_i)] \quad (4)$$

In formula (4), M represents the collection of images in the training set, y_i represents the label, the value of positive samples is 1, and the value of negative samples is 0. p_i Indicates that the output value of the softmax layer is between 0 and 1. When the difference between the predicted value and the label value increases, the gradient of the cross-entropy loss function will also increase, and the parameter adjustment amplitude will increase accordingly to achieve the purpose of rapid convergence when the parameters are updated

III. EXPERIMENTAL RESULTS

In this section, The proposed network will be evaluated by detecting images taken by high-definition cameras. First, the main contents of the experiment include the description of the data set, the evaluation method and the experiment setting. Then, according to the experimental results, the influence of the model or the network is analyzed and conclusions are drawn. In the experimental settings, two experiments were carried out, namely, the performance test of the detection network and the test of different input sizes of the detection network, verifying the influence of the multi-layer fusion module, the multi-region adaptive module and different input sizes on the insulator detection network. In addition, a fault classification network test experiment was set up to test the performance of the fault classification network. Finally, a CSDN performance test experiment was carried out to compare the performance of our detection network and the current mainstream detection network in railway insulator fault detection and classification tasks. The results of the above experiments are detailed in C.

A. DATA SET PREPARATION

All catenary images are provided by China Railway Group Co., Ltd. for the monitoring and testing of catenary. The pixel size of the catenary image is 3968×2976 , and the number is 1100, each image containing at least three insulators. In the entire data set, the weather in the image includes sunny, cloudy, and rainy days, and the background includes hills and plains. Among them, 1000 images are used to make the detection data set and classification data set, and the other 100 images are used to make the evaluation data set. The detection data set is used to train, verify and test the detection network; the classification data set is used to train, verify and test the classification network; the evaluation data set is used to evaluate the performance of the CSDN.

The 1000 contact net images with a pixel size of 3968×2976 are reduced to 500×375 images, and then the data set is expanded by the method for mirror transformation and

brightness adjustment to make the detection data set. Mirror transformation and brightness adjustment cannot only expand the data, but also simulate the real railway environment. After data expansion, the test data set is composed of 4000 catenary images with a pixel size of 500×375 ; the test data set is divided into training set, validation set and test set. Part of the image of the detection data set is shown in Figure 4 a.

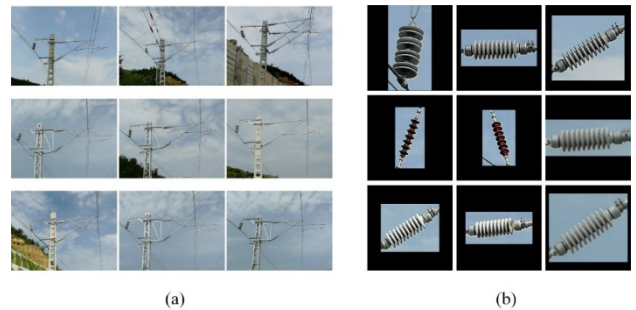


FIGURE 4. Part of the image in the data set. (a) The picture is the detection data set, (b) the picture is the classification data set.

The insulator coordinate information of the training set in the detection data set is enlarged and projected back to the original catenary image with a size of 3968×2976 , and then the insulator image is intercepted to make a classification data set. A total of 2,000 insulator images were intercepted, including 1,851 insulators in normal state, 81 insulators damaged, and 68 insulators missing. Because the insulator is damaged and the number of missing images of the insulator is small, it will cause data imbalance, which will affect the network performance. Therefore, the method for data expansion (such as mirroring, rotation and brightness adjustment) is used to expand the images of damaged insulators and missing insulators to reduce the impact of data imbalance. After the data set is expanded, the total number of images in the classified data set is 2447. Since the intercepted insulator images are inconsistent in size and cannot be input into the classification network, the images in the classification data set are filled with 360×360 images. Part of the image of the prepared classification data set is shown in Figure 4 (b). The evaluation data set consists of 100 original catenary images with a pixel size of 3968×2976 , which are used to evaluate the performance of CSDN. The specific image data division of the detection data set, classification data set and evaluation data set is shown in Table 1.

TABLE 1. Data set division.

	Training set	Validation set	Test set	Total
Detection data set	2660	1140	200	4000
Classification data set	1757	490	200	2447
Evaluation data set	-	-	100	100

B. IMPLEMENTATION

The parameters of the detection network in this article are set as follows: the basic size of anchors is (64, 128, 256), and

the scaling ratio of anchors is (0.5, 1, 2); the initial learning rate is 0.001, the momentum is 0.9, the training batch is 64, the IOU threshold in the RPN module is 0.6, and the NMS threshold is 0.7. In the multi-layer fusion module, the weights of vgg16 on ImageNet are used to initialize the convolutional layer, and other convolutional layers in the detection network are initialized with He initialization [30]. In order to evaluate the performance of the detection network, four widely used metrics are applied, namely AP, Recall, Precision and PR curves. Considering that the railway companies pay different attention to AP, Recall and Precision, we have introduced the APR indicator [31]. The calculation formula of APR is:

$$APR = 0.5 * AP + 0.3 * Recall + 0.2 * Precision \quad (5)$$

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

The parameters of the classification network are set as follows: the initial learning rate is 0.005, the momentum is 0.9, the training batch is 128, and the number of iterations is 6,000. All convolutional layers are initialized using HE.

Three metrics were used to evaluate the performance of the classification network, namely Precision, Recall and F1- Score. The calculation formula of F1- Score is:

$$F1 = \frac{2Precision * Recall}{Precision + Recall} \quad (6)$$

C. EXPERIMENTAL RESULTS

1) MULTI-LAYER FUSION MODULE

The model of the detection network is shown in Figure 2. In this section, the influence of the multi-layer fusion module and the multi-region adaptive module on the detection network are tested through experiments. For this, four models are trained separately: MF_0, F_3, MF_5 and MFA. In the three detection network models of MF_0, MF_3 and MF_5, the three numbers 0, 3, and 5 indicate the number of blocks fused in the fusion module, and none of these three detection network models contain a multi-region adaptive module. The MFA detection network model includes a multi-layer fusion module, a multi-region adaptive module and an RPN module, and the output of all blocks is fused in the multi-layer fusion module. The four models are trained using the same training set of the detection data set, and the model is output after training for 2000epoch, and then tested on the test set. The test results are shown in Table 2 and Figure 5.

It can be seen from Table 2 that MF_5 and MF_3 are improved in Precision, Recall, AP and APR compared with MF_0, indicating that the method of fusing the feature maps output by different blocks can improve the performance of the detection network. This is because the feature maps output by different convolutional layers contain different levels of information, and fusing these different feature maps together can

TABLE 2. Test results of different models.

Net	Precision(%)	Recall(%)	AP(%)	APR(%)	Time(ms)
MF_0	88.21	87.89	89.15	88.58	40.1
MF_3	91.21	89.70	91.24	90.77	65.9
MF_5	91.88	90.41	91.35	91.17	68.6
MFA	94.62	94.87	94.23	94.50	82.7

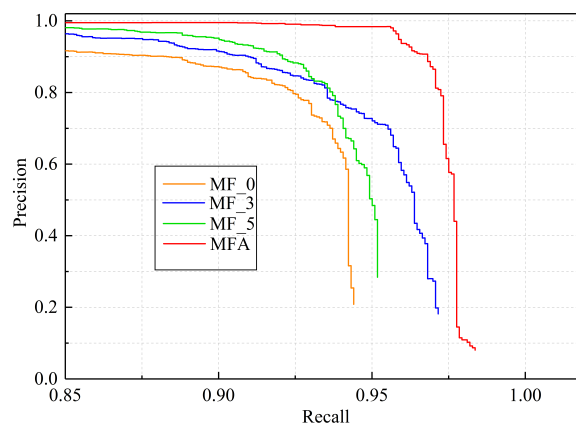


FIGURE 5. PR curves of different models.

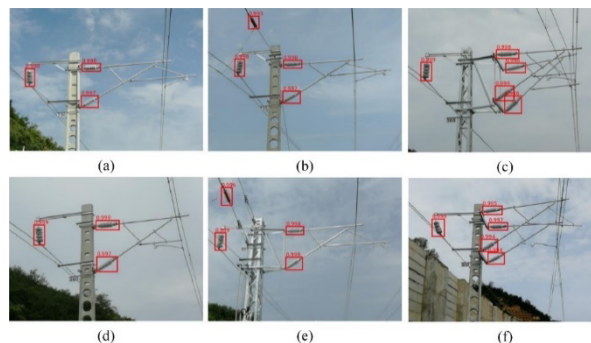


FIGURE 6. Test results of the detection network.

enrich the semantic information of the feature maps. Compared with MF_3, the performance of MF_5 has improved but not much, indicating that the feature map extracted by the previous convolutional layer has a low contribution to improving the performance of the detection network. Compared with the other three models, MFA has the best performance in the four indicators of Precision, Recall, AP and APR, and the improvement of Recall is the largest, indicating that the multi-region adaptive module improves the performance of the detection network. This is because in the multi-region adaptive module, we multiply the attention map with the original feature map, which can strengthen the region of interest of the network and retain smaller features. The Time indicator is the time it takes to test each image, and the Time of the four model tests of MF_0, F_3, MF_5, and MFA increase in turn. This is because the more complex the model, the more time it takes to detect.

The Precision-recall curve is shown in Figure 5. It can be seen that the Precision of the MF_0 model remains above 90% when the recall changes from 0 to 87%; The Precision of MF_3 and MF_5 models remained above 90% when the recall range was 0 to 90%, and the Precision of MFA model remained high when the recall range was 0 to 95%. This shows that the multi-region adaptive module improves the detection network performance more than the multi-layer fusion module. Since the MFA model achieves a good compromise between high precision and high recall, and can accomplish the insulator detection task well, the MFA model will be used as our detection network.

Figure 6 is a partial test result of the MFA model on the test set, where the number above the red box represents the score of the insulator. It can be seen from Figure 6 that the scores of the insulators are all above 99% and the positions of the frames are more accurate. Figure (a) is an image taken on a sunny day and Figure (b) is an image taken on a rainy day. It can be seen from the detection results that the detection network performs very well under different weather conditions, and that Figures (b) and (e) both contain two different types of insulators. It can be seen from the detection results that the detection network has a good detection effect for different types of insulators, and that the background in figure (c) is relatively simple in the sky and the background in figure (f) has protective walls, plants and hills. It can be seen from the inspection results that the inspection network performs well regardless of its complex background or single performance.

It can be seen from the above experimental results that the AP value and APR value of our designed detection network are 94.23% and 94.50% respectively, and they have good detection results for different types of insulators under different weather backgrounds. It shows that the detection network we designed can well complete the task of detecting the position of the railway catenary insulator.

2) DETECTION NETWORK IMAGE INPUT TEST OF DIFFERENT SIZES

We use images with long side lengths of 500, 400, 300, and 200 to train the detection network, and also change the long side lengths of the images in the test set to 500, 400, 300, and 200 respectively. The test results of the impact of the test image size on the detection network are shown in Table 3 and Figure 7. It can be seen from Table 3 that when the input size is 200, AP and APR are 35.82% and 37.97%, respectively, which are both low, because the smaller the image size, the more difficult it is to detect insulators. In Table 3, when the input size is increased from 200 to 500 in turn, both AP and APR values increase sequentially, and when the input size is 500, AP is 94.23%, APR is 94.50%; therefore, we set the input size of the test network to 500. It can be seen from Table 3 that when the size is increased from 200 to 500, the detection time also increases from 49.1ms to 82.7ms, indicating that the larger the image size, the longer the detection time.

TABLE 3. Performance test results of different input sizes of the detection network.

Size	Precision(%)	Recall(%)	AP(%)	APR(%)	Time(ms)
200	38.56	41.17	35.82	37.97	49.1
300	79.81	81.35	81.44	81.09	68.4
400	90.29	92.75	90.32	91.04	73.5
500	94.62	94.87	94.23	94.50	82.7

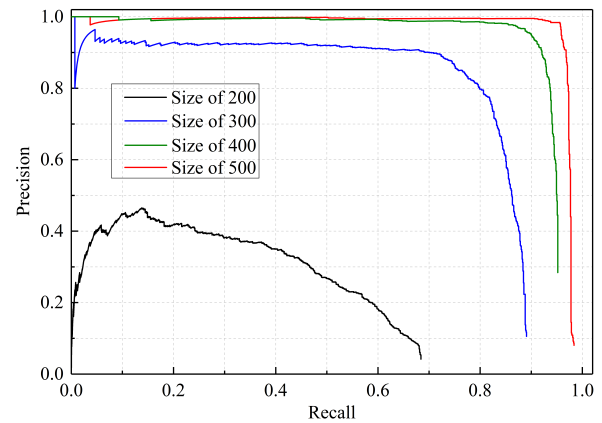


FIGURE 7. Test results of detecting different size inputs of the network.

Figure 7 shows the PR curves of models trained with different input sizes on the test set. It can be seen that the detection network trained with an image size of 200 performs poorly, that the performance of the detection network becomes better as the image size increases, and that the detection network trained with an image size of 500 performs best.

From the above experimental results, it can be seen that as the size of the input image of the detection network increases, the network detection effect becomes better, and the detection time is also increasing. In order to balance the network detection effect and detection time, we set the long side size of the input image of the detection network to 500.

3) FAULT CLASSIFICATION NETWORK TEST

The model of the classification network is shown in Figure 3. In this section, the training set in the classification data set will be used to train the classification network. The trend of loss and accuracy during training is shown in Figure 8. Figure 8 (a) shows the loss change curve during the training of the classification network. It can be seen from the figure that the loss of the classification network is continuously decreasing. When the loss of the classification network is reduced to less than 0.1 after 5000steps of training. Figure (b) in Figure 8 shows the accuracy of the classification network during training. From Figure (b), it can be seen that the accuracy of the network after 500step training reaches 0.9 or more, and after 6000step training, the accuracy can reach 0.95 or more. Figure 8 shows that the classification network converges and performs better for insulator state classification. Insulators can be divided into three states: normal, damaged

TABLE 4. Test results of classification network on the test set.

Condition	Precision	Recall	F1-Score
Normal	96.95%	95.48%	96.09%
Damaged	91.89%	89.47%	90.66%
Missing	84.37%	93.10%	88.52%

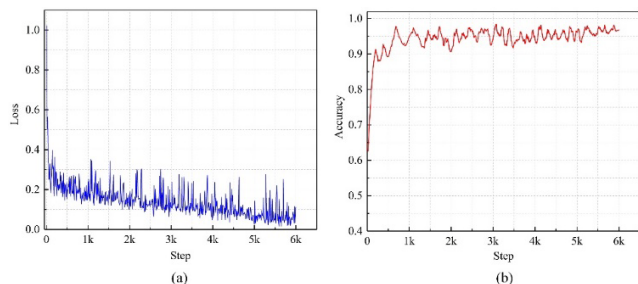


FIGURE 8. Shows the results of the classification network training after 6000 steps. (a) Is the loss curve after 6000 steps, and (b) is the accuracy curve after 6000 steps.

and missing. The test results of the classification network for different states are shown in Table 4.

Table 4 shows the performance of the classification network on the test set. From Table 4, it can be seen that the F1-Score of the normal type of insulator is 96.09%, the F1- Score of the damaged type of insulator is 90.66%, and the F1-Score of the missing type of insulator is 88.52%. The main reason for the lower F1-Score of the classification network for the damaged and missing states is that the number of insulator images in the damaged and missing states in the training data set is small.

From the above experimental results, it can be seen that the classification network has a good classification effect for the three states of insulators, and the F1-Score of the three states is above 88%, which can be competent for the classification task.

4) CSDN PERFORMANCE TEST

Connect the trained detection network and the classification network to form the CSDN, and use the evaluation data set to evaluate the CSDN. In order to evaluate the impact of insulator background images on the performance of CSDN, 20 images with backgrounds of vegetation, buildings and sky were selected from the evaluation data set, and the trained CSDN was used for testing. The test results are shown in Table 5, and some test diagrams are shown in Fig. 9.

TABLE 5. Test results of CSDN under different backgrounds.

Backgrounds	Precision	Recall	mAP	APR
Vegetation	94.08%	92.89%	93.12%	93.24%
Sky	94.60%	93.75%	93.88%	93.99%
Buildings	92.19%	90.23%	91.50%	91.26%

As shown in Table 5, when the background of the insulator is all sky, the mAP value reaches 93.88%, which is the highest mAP value among the three background images. The background is the insulator image of vegetation, and the mAP value is 93.12% during the test. The background is the insulator image of the building, and the mAP value is 91.50% during the test, which is the lowest score among the three background images. The test results show that when the background image of the insulator is vegetation or the sky, it has a small impact on the CSDN model, and when the background is a building, it has a greater impact on the CSDN model.



FIGURE 9. Test results of the CSDN model in different contexts.

The background of the insulator in Figure 9 (a) is vegetation, the background in (b) is all sky, and the background in (c) is buildings, and the light intensity is also different in the 9 results. In the figure (a), the background of the three images is vegetation, and the light intensity increases from left to right. It can be seen from the detection effect that the insulator is accurately detected and the state is correctly classified. In the figure (b), the background of the three images is the sky, and the light intensity increases from left to right. It can be seen from the detection effect that the insulator is accurately detected and the state is correctly classified. In the figure (c), the background of the three images is the background, and the light intensity increases from left to right. It can be seen from the detection results that a few insulators have not been detected. This is because the color of the background is too close to the color of the insulator, which increases the difficulty of network detection. It can be seen that the light intensity has a small impact on the network, but the image background has a greater impact on the detection model.

In order to further test the performance of CSDN, we compared the trained CSDN model with the detection methods in the literature [17], [18], and also compared with the classic target detection framework: Faster R-CNN, SSD and YOLOv3. The models in the literature [17], [18] and the classic target detection framework all use the data set in

TABLE 6. Comparison results of CSDN and different models.

Net	Size	Precision(%)	Recall(%)	mAP(%)	APR(%)	Time(ms)
CSCNET[17]	3968×2976	85.40	88.34	86.24	86.70	359
Faster R-CNN+DMNN[18]	3968×2976	91.35	89.73	90.63	90.50	932
Faster R-CNN	3968×2976	89.28	81.09	88.48	86.42	813
SSD	3968×2976	87.88	82.79	82.88	83.85	274
YOLOv3	3968×2976	81.53	68.57	71.50	72.63	148
CSDN	1984×1488	90.65	89.22	89.31	89.55	195
CSDN	2976×2232	92.78	91.31	91.89	91.89	219
CSDN(ours)	3968×2976	94.10	92.88	93.46	93.41	231

this article for training and testing. Table 6 summarizes the comparison results.

As shown in Table 6, the mAP and APR values of the CSDN model on the evaluation data are 93.46% and 93.41%, respectively, which are better than other comparative models. It shows that the detection network and the cascade method of detection network and classification network proposed in this paper are beneficial to improve the effect of insulator position detection and state classification. The detection time of an image in the CSDN model is 231ms. Compared with the SSD and YOLOv3 designed based on the one-stage method, the test time is not much different. The main reason is that the CSDN model reduced the original catenary image with a size of 3968×2976 to 500×375 for detecting the position of the insulator, which greatly accelerated the speed of detecting the position of the insulator. At the same time, the insulator position information is used to obtain a more informative insulator image from the original catenary image, and VGG16 is used for classification to ensure the accuracy of the insulator state classification. As shown in Table 6, in the CSDN model, the larger the input image size, the better the detection effect, but the detection time will also increase. The reason is that high-resolution input images contain more detailed information, which is conducive to the classification of the network, but also increases the amount of network calculations. It can be seen from Table 6 that the Faster R-CNN and Faster R-CNN+DMNN models designed based on the two-stage method have higher mAP and APR values, but the detection time is longer. The detection model designed based on the one-stage method has a short detection time, but both mAP and APR values are low. The CSDN model we designed is better than the detection model designed based on the two-stage method in both mAP and APR values, and the detection speed is also close to the detection model designed based on the one-stage method.

Figure 10 shows a partial detection result diagram of CSDN. The red box in the figure indicates that the insulator state is normal, the green box indicates that the insulator state is damaged, and the yellow box indicates that the insulator state is missing. In Figure 10 (a), (b) and (d), insulators in two states are normal and damaged respectively; Figure (c) insulators in two states, missing and normal. It can

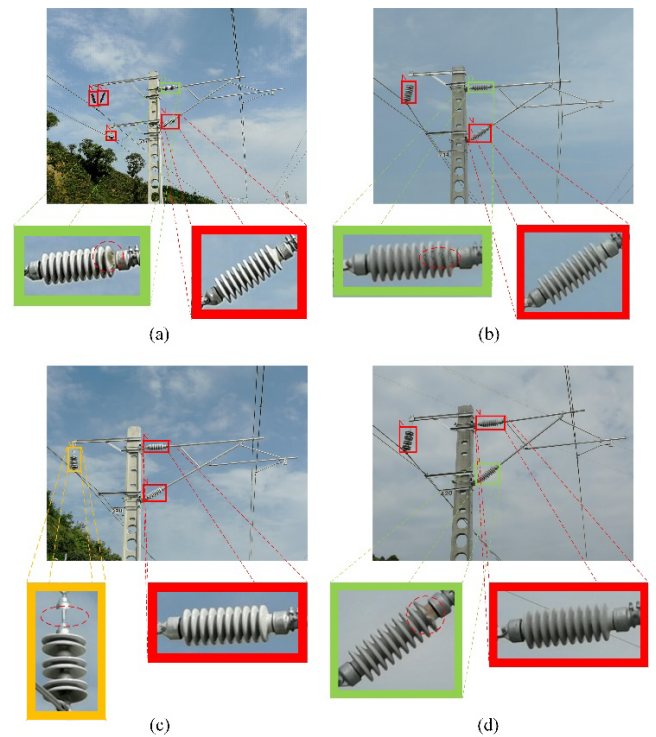


FIGURE 10. CSDN test results on the evaluation data set, the red box, green box, and yellow box in the figure represent normal, damaged, and missing insulators.

be seen from the above experimental results that the precision value, recall value, mAP value and APR value of CSDN are 94.10%, 92.88%, 93.46%, and 93.41% respectively, which are higher than the other three models, indicating that our network is in contact with railways. The grid insulator fault detection task has a high accuracy rate and good robustness.

IV. CONCLUSION

In this paper, we have designed an insulator detection and fault classification network based on deep convolutional neural networks to check the status of insulators in railway catenary images taken by high-definition cameras. In order to solve the problem that the original image pixels are too large

to directly compress the insulator details, we propose a CSDN detection framework. In our proposed detection framework, two convolutional network connection methods are used to detect the position of the insulator and classify its state. First, the reduced image is used to train the detection network so that the detection network can quickly and accurately detect the position of the insulator. Then, the insulator position information detected by the detection network is used to project back to the catenary image to intercept the insulator image to train the classification network. After the two networks are trained, they are connected together to form our CSDN detection framework. In our detection framework, the detection network input is a reduced contact network image, which can speed up detection and ensure detection accuracy. In the detection network, we have added a multi-layer fusion module to integrate different levels of semantic information to improve the precision of the detection network, and also added a multi-region adaptive module for adaptive optimization of features, which improves the recall rate of the detection network. The trained detection network has achieved 94.23% of the insulator detection AP and 94.50% of the APR. In our detection framework, for the classification network, the insulator images intercepted in the catenary image are used for training, which improves the accuracy of classification. The trained classification network has F1-Scores of 96.09%, 90.66%, and 88.52% for normal, damaged and missing states, respectively. The trained detection network and the classification network are connected to form our CSDN detection framework, with the test result of the detection framework on the evaluation data set being 93.46% and the PRA being 93.41%. In future work, the data set needs to be further expanded so that the model can identify more catenary failures.

REFERENCES

- [1] L. Pang, "Research of catenary insulator damage detection technology based on machine vision," M.S. thesis, Dept. Electron. Eng., LZJT Univ., Shanxi, China, 2017.
- [2] T. Zhang, *Maintenance and Management of Railway Catenary*. Beijing, China: Railway Press, 2006, pp. 62–67.
- [3] X. Zhang, J. An, and F. Chen, "A simple method of tempered glass insulator recognition from airborne image," in *Proc. Int. Conf. Optoelectron. Image Process. (ICOIP)*, vol. 1, Nov. 2010, pp. 127–130.
- [4] M. J. B. Reddy, B. K. Chandra, and D. Mohanta, "A DOST based approach for the condition monitoring of 11 kV distribution line insulators," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 18, no. 2, pp. 588–595, Apr. 2011.
- [5] B. Li, D. Wu, Y. Cong, Y. Xia, and Y. Tang, "A method of insulator detection from video sequence," in *Proc. 4th Int. Symp. Inf. Sci. Eng. (ISISE)*, Dec. 2012, pp. 386–389.
- [6] M. Oberweger, A. Wendel, and H. Bischof, "Visual recognition and fault detection for power line insulators," in *Proc. 19th Comput. Vis. Winter Workshop*, 2014, pp. 1–8.
- [7] Q. Wu, J. An, and B. Lin, "A texture segmentation algorithm based on PCA and global minimization active contour model for aerial insulator images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1509–1518, Oct. 2012.
- [8] W. Wang, Y. Wang, J. Han, and Y. Liu, "Recognition and drop-off detection of insulator based on aerial image," in *Proc. 9th Int. Symp. Comput. Intell. Des. (ISCID)*, vol. 1, Dec. 2016, pp. 162–167.
- [9] 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.
- [10] 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)*, Jun. 2016, pp. 779–788.
- [11] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
- [12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," 2015, *arXiv:1512.02325*. [Online]. Available: <http://arxiv.org/abs/1512.02325>
- [14] D. Zhang, S. Gao, L. Yu, G. Kang, X. Wei, and D. Zhan, "DefGAN: Defect detection GANs with latent space pitting for high-speed railway insulator," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.
- [15] A. Varghese, J. Gubbi, H. Sharma, and P. Balamuralidhar, "Power infrastructure monitoring and damage detection using drone captured images," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1681–1687.
- [16] F. Gao, J. Wang, Z. Kong, J. Wu, N. Feng, S. Wang, P. Hu, Z. Li, H. Huang, and J. Li, "Recognition of insulator explosion based on deep learning," in *Proc. Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2017, pp. 79–82.
- [17] Z. Liu, K. Liu, J. Zhong, Z. Han, and W. Zhang, "A high-precision positioning approach for catenary support components with multiscale difference," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 700–711, Mar. 2020.
- [18] 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.
- [19] H. Hongtao, M. Yongzhi, Y. Kun, and L. Yifan, "Research on detection algorithm of catenary insulator based on improved faster R-CNN," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2020, pp. 3814–3818.
- [20] H. Jiang, C. Zhang, and M. Wu, "Pedestrian detection based on multi-scale fusion features," in *Proc. IEEE Int. Conf. Netw. Infrastruct. Digit. Content (IC-NIDC)*, Aug. 2018, pp. 329–333.
- [21] Y. Zhu, C. Zhao, H. Guo, J. Wang, X. Zhao, and H. Lu, "Attention CoupleNet: Fully convolutional attention coupling network for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 113–126, Jan. 2019.
- [22] Z. Lan, Q. Huang, F. Chen, and Y. Meng, "Aerial image semantic segmentation using spatial and channel attention," in *Proc. 4th Int. Conf. Image, Vis. Comput. (ICIVC)*, Jul. 2019, pp. 316–320.
- [23] R. H. Md. Rafi, B. Tang, Q. Du, and N. H. Younan, "Attention-based domain adaptation for hyperspectral image classification," in *Dig. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 67–70.
- [24] X. Ying, Q. Wang, X. Li, M. Yu, H. Jiang, J. Gao, Z. Liu, and R. Yu, "Multi-attention object detection model in remote sensing images based on multi-scale," *IEEE Access*, vol. 7, pp. 94508–94519, 2019.
- [25] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, Jun. 2007, pp. 759–766.
- [26] N. M. Zamri, G. F. Ling, P. Y. Han, and O. S. Yin, "Vision-based human action recognition on pre-trained AlexNet," in *Proc. 9th IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Nov. 2019, pp. 1–5.
- [27] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2017, *arXiv:1709.01507*. [Online]. Available: <http://arxiv.org/abs/1709.01507>
- [28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*. [Online]. Available: <http://arxiv.org/abs/1807.06521>
- [29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [31] H. Jiang, X. Qiu, J. Chen, X. Liu, X. Miao, and S. Zhuang, "Insulator fault detection in aerial images based on ensemble learning with multi-level perception," *IEEE Access*, vol. 7, pp. 61797–61810, 2019.



ZAXING WANG received the Ph.D. degree from Lanzhou University, Lanzhou, China, in 2008. He is currently an Associate Professor at Lanzhou Jiaotong University, Lanzhou. His research interests include novel device simulation and model, computer software and application, radio electronics, and automation technology.



LIJUN ZHENG received the B.Sc. degree from Lanzhou Jiaotong University, Lanzhou, China, in 2015, where she is currently pursuing the M.S. degree. Her research interests include semiconductor power device and technological fundamentals of semiconductor device.



XIAOZHONG LIU received the B.Sc. degree from Lanzhou Jiaotong University, Lanzhou, Gansu, China, in 2017, where he is currently pursuing the M.S. degree. His research interests include computer graphics, computer vision, machine learning, and image processing.



JINHUI GAO received the B.Sc. degree from Xi'an University of Posts and Telecommunications, Xi'an, China, in 2018. He is currently pursuing the M.S. degree with Lanzhou Jiaotong University, Lanzhou, China. His research interests include semiconductor-device and FPGA.



HUAYI PENG received the B.Sc. degree from Harbin University of Science and Technology, Harbin, China, in 2019. He is currently pursuing the M.S. degree with Lanzhou Jiaotong University, Lanzhou, China. His research interests include semiconductor power device, computer graphics, computer vision, machine learning, and image processing.



YUFAN BAO received the B.Sc. degree from Chongqing University of Posts and Telecommunications, Chongqing, China, in 2017. He is currently pursuing the M.E. degree with Lanzhou Jiaotong University, Lanzhou, China. His research interests include semiconductor power device and technological fundamentals of semiconductor device.

...