

Deep Architecture for High-Speed Railway Insulator Surface Defect Detection: Denoising Autoencoder With Multitask Learning

Gaoqiang Kang, Shibin Gao^{ID}, Long Yu^{ID}, and Dongkai Zhang

Abstract—The insulator is an important catenary component that maintains the insulation between the catenary and earth. Due to the long-term impact of railway vehicles and the environment, defects in the insulator are inevitable. Recently, automatic catenary inspection using computer vision and pattern recognition has been introduced to improve the safety of railway operation. However, achieving full automation of insulator defect detection is still very challenging due to the visual complexity of defects and the small number of defective insulators. To overcome these problems, this paper proposes a novel insulator surface defect detection system using a deep convolutional neural network (CNN). The proposed system consists of two stages. First, a Faster R-CNN network is adopted to localize the key catenary components, and the image areas that contain the insulators are obtained. Then, the classification score and anomaly score are determined from a deep multitask neural network that is composed of a deep material classifier and a deep denoising autoencoder. The defect state is determined by analyzing the classification score and anomaly score. Experiments of the catenary insulator defect detection along the Hefei–Fuzhou high-speed railway line indicate that the system can achieve high detection accuracy.

Index Terms—Catenary, deep denoising autoencoder (DDAE), defect detection, high-speed railway, insulator, multitask learning (MTL).

I. INTRODUCTION

AS A KEY piece of equipment of the traction power supply system in electrified railway, the catenary directly affects the safety and reliability of the railway transport. The catenary operates in the open-air environment and suffers from the mechanical and electrical impact of the pantograph, making it a weak part in the traction power supply system [1]. Thus, to ensure safe operation of the traction power supply, it is essential to monitor the condition of the railway catenary,

Manuscript received February 12, 2018; revised August 9, 2018; accepted August 14, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant U1434203, in part by the National Key Research and Development Plan of China under Grant 2017YFB1201202, and in part by the China Railway Corporation Science and Technology Research and Development Project under Grant 2015J008-A. The Associate Editor coordinating the review process for this paper was Dr. Zheng Liu. (*Corresponding author: Long Yu.*)

G. Kang is with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China, and also with the School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China (e-mail: kanggqyy@163.com).

S. Gao, L. Yu, and D. Zhang are with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China (e-mail: gao_shi_bin@126.com; yulong.swjtu@163.com; fengmitu@163.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2018.2868490

especially for high-speed railway. Traditionally, a catenary inspection is performed by trained personnel who walk along the railway network to check its operating status [2]. However, a manual inspection is slow and unreliable and cannot meet the growing requirements of the high-speed railway operation. Therefore, it is necessary to develop automatic inspection systems to improve the efficiency and reliability of catenary inspection.

Currently, due to the advantages of high efficiency and reliability, automatic inspection systems based on the computer vision technology have been gradually applied to railway inspections, such as for rail, vehicle, and catenary [3]–[5]. The insulator, as an important component of the catenary, not only maintains the insulation between the catenary and earth but also plays a role in supporting the catenary. Therefore, the insulator defect detection is very important. As a distributed system, the catenary is generally inspected using cameras installed on the inspection vehicle. The inspection vehicle takes images dynamically, which are inevitably influenced by complex background, light intensity, and shooting attitude. Therefore, there are two key tasks to accomplish insulator defect detection: object detection and defect detection, in other words, localizing the insulator from the complex background and detecting defect based on the localized insulator image.

Many methods have been proposed for object detection using handcrafted features as well as deep learning approaches. In [6], the maximally stable extremal region technique was employed to detect three classes of railway track surface defect regions. Liao and An [7] combined local features into a multiscale local feature descriptor to detect the insulator. Oberweger *et al.* [8] designed a part-based model with a tailored circular descriptor and a voting scheme for insulator localization. Han *et al.* [9] proposed an insulator detection method based on deformable part models, which is capable of handling intraclass variability and performs well with occlusion. These methods have achieved relatively good results in object detection using handcrafted features. However, in recent years, object detection algorithms based on deep learning [10]–[14], using learned features, have made a breakthrough in accuracy, scalability, and robustness. Gibert *et al.* [15] trained a deep convolutional neural network (DCNN) end to end for the material classification and railway track elements detection. Liu *et al.* [16] used a Faster R-CNN network to localize the isoelectric line and achieved a state-of-the-art result compared to that of the approaches

using handcrafted features. Inspired by these works, a Faster R-CNN network is used in this paper to localize insulators.

In the insulator defect detection stage, image segmentation is an important issue. In [17], DCNN was employed to extract the multiscale feature maps of the input image and a classifier to produce segmentation results. In [15], a deep material classifier (DMC) was employed to segment the railway track elements and achieved a state-of-the-art result. Inspired by these works, a DMC is designed to segment the insulators. After segmentation, two types of approaches based on machine learning can be used to detect defects: supervised and unsupervised. The supervised learning approach, which requires defective samples for classifier training, has been widely used in railway infrastructure defect detection [18]–[20]. However, the number of defective insulators in practice is very small, which is not enough to train a robust classifier. Fortunately, defect detection can be treated as an anomaly detection problem [21] and solved using the unsupervised learning approach that does not need defective samples for training. Autoencoder [22], which can learn low-dimensional, nonlinear features of the input data from unlabeled data sets, has been widely used in anomaly detection. In [23], the acoustic novelty was detected based on the reconstruction error between the input and the output of the denoising autoencoder. The use of autoencoder for outlier detection was studied in [24]. Liu *et al.* [25] employed a deep sparse autoencoder to extract hidden features of driving behavior and recognize distinctive driving behavior patterns. Hasan *et al.* [26] applied a deep autoencoder to learn the regular dynamics in videos and identify irregularity. In [27], it was shown that denoising autoencoders can achieve higher anomaly detection accuracy than the basic autoencoders. We propose using a deep denoising autoencoder (DDAE) for the insulator surface defect detection. To further improve the performance of the DDAE, we introduce the idea of multitask learning (MTL) [28], [29] in our study; the DDAE and the DMC are integrated together in a deep architecture and trained jointly. We demonstrate that the supervised DMC and the unsupervised DDAE can benefit from each other in the MTL framework.

In this paper, a new solution of the insulator surface defect detection is provided. The main contribution of this paper is that a DMC and a DDAE are integrated into a deep multitask neural network (DMNN) architecture, needing no defective samples for training to accomplish simultaneous segmentation and defect detection. The additional contribution is the introduction of a metric that can quantitatively assess the surface defects of the insulators.

This paper is organized as follows. Section II overviews the defect detection system. The localization of key components is described in Section III. The DMNN is theoretically described in Section IV. The experimental results and analysis are summarized in Section V. Section VI presents the conclusion and future work.

II. SYSTEM OVERVIEW

The inspection system, containing two groups of the camera and auxiliary light devices, is installed on the roof of the inspection vehicle, as shown in Fig. 1. Two groups of

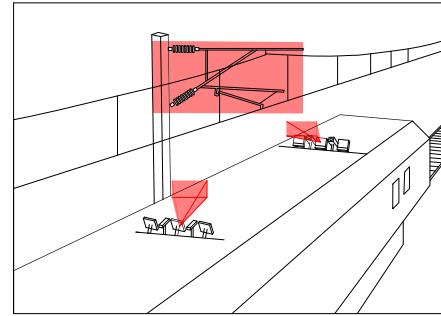


Fig. 1. Sketch map of the catenary inspection device.

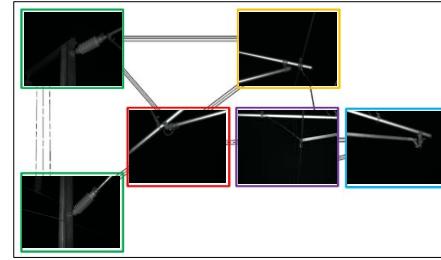


Fig. 2. Image captured by different cameras of the inspection device.

high-resolution cameras take photographs of the catenary from two opposite directions, respectively.

To further improve the resolution of the cameras with a resolution of 4920×3280 pixels are, respectively, responsible for a part of the catenary imaging. Images that captured by the inspection vehicle at night are shown in Fig. 2.

During the inspection, the captured images, kilometer marks, and other information are stored in a vehicle database for offline defect detection.

The proposed insulator surface defect detection system contains two main modules: components localization and insulator surface defect detection. Fig. 3 shows the diagram of the detection system.

A. Key Components Localization

The purpose of key components localization is to localize and extract six kinds of key components including insulators in the catenary images captured by different cameras. The images of the key components are captured dynamically by the inspection vehicle and have multiple scales and complex backgrounds. To localize the key components in the captured images, the Faster R-CNN network that performs well in both speed and accuracy is adopted.

B. Insulator Surface Defect Detection

As a key component, once the insulator is localized, the impact of the background is greatly reduced. However, there are still the following problems.

- 1) The number of defective samples is not sufficient to train a robust classifier.
- 2) The visual complexity of defects makes it difficult if not impossible to construct a precise model.
- 3) The gray-scale values of insulators are similar to those of other catenary devices in the background.

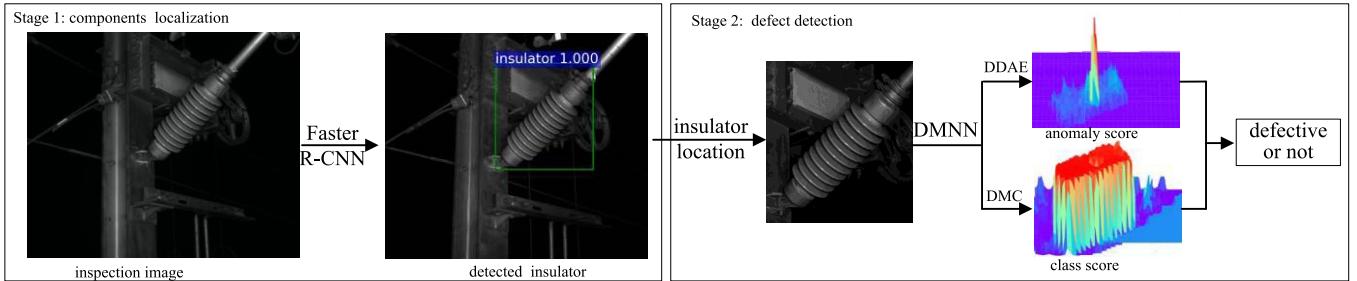


Fig. 3. Diagram of the proposed two-stage detection system that consists of the components localization and defect detection.

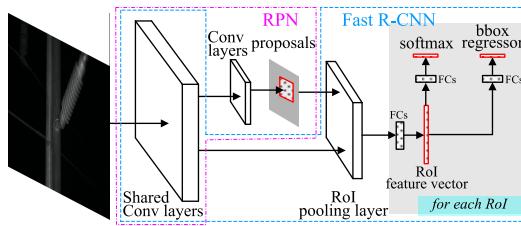


Fig. 4. Architecture of the Faster R-CNN.

To overcome these problems, the DMNN composed of the DMC and the DDAE is proposed to detect insulator defects in a sliding window way. When the DMNN slides on the image, the DMC and DDAE give the classification score and anomaly score of the insulator, respectively. Then, the defect state of the insulator is identified by analyzing these scores.

III. KEY COMPONENTS LOCALIZATION

As discussed previously, we first localize the insulator in a catenary image and then identify the insulator's defect state. As one of the key components, insulators are localized by the same Faster R-CNN network with other key components. The Faster R-CNN network consists of two CNNs: the regional proposal network (RPN) and the Fast R-CNN object detection network [13], as shown in Fig. 4. These two deep neural networks share full-image convolutional layers, which improve the detection efficiency.

The RPN is a fully convolutional network that takes the image as the input and directly generates the region proposals that may contain the object to be detected. The RPN network slides through the feature map of the shared convolutional layers with a 3×3 spatial sliding window and simultaneously predicts k anchors of different scales and aspect ratios at each sliding position. For a convolutional feature map of size $W \times H$, there is a total of $W \times H \times k$ anchors. Then, according to the score of the anchors being an object or not, the top 300 anchors are selected as the region proposals and are fed to the region of interest (RoI) pooling layer of the Fast R-CNN. The RoI pooling layer uses max pooling to convert the features of the region proposals into fixed-length feature vectors. Finally, the softmax probabilities and per-class bounding-box regression offsets are output by the fully connected layers. After training, the two networks form a unified Faster R-CNN network that is used to localize key catenary components.

IV. DEFECT DETECTION

In this section, we describe the proposed DMNN for the insulator surface defect detection. As shown in Fig. 5, the network is composed of the DMC and DDAE which share convolutional layers. The detailed parameters of the DMNN are given in Table I.

We detect the insulator surface defects by sliding the DMNN over the insulator images, and the DMC determines whether the patch at the current sliding position belongs to the insulator or not. Meanwhile, the autoencoder reconstructs the patch and outputs corresponding reconstruction error. Then, the insulator's classification score and reconstruction error map are obtained and used to calculate the anomaly score. Finally, the defect state of the insulator is identified by analyzing the classification score and anomaly score. The proposed method is a multitasking learning framework that learns two tasks at the same time, in which the DMC and DDAE can benefit from each other. In addition, as an unsupervised learning framework, training the network does not require images of defective insulators.

A. Deep Material Classifier

Once an insulator is localized, the impact of the background is greatly reduced. However, due to the gray-scale values similarity between insulators and other catenary equipment, as shown in Fig. 6, an accurate segmentation of the insulators is still a very difficult task. Inspired by Gibert *et al.* [15] and Farabet *et al.* [17], we segment the insulators using a deep classifier that slides over the insulator images.

The DMC has a total of four convolutional and two fully connected layers and shares the first two convolutional layers with the DDAE. In the convolutional layers, the input data are convoluted with the linear filters, and the feature map is obtained from the nonlinear activation function. Each feature map contains a kind of feature, which shares the same filter parameters, and different feature maps use different filter parameters to extract different features.

To prevent overfitting, we use dropout [30] regularization on the layer Fc1_c with a ratio of 0.3. In addition, the rectified linear units' activation function is used in all layers.

The loss function of the classifier is as follows:

$$L_c = - \sum_i ((1 - p_i) \log(1 - \hat{p}_i) + p_i \log \hat{p}_i) \quad (1)$$

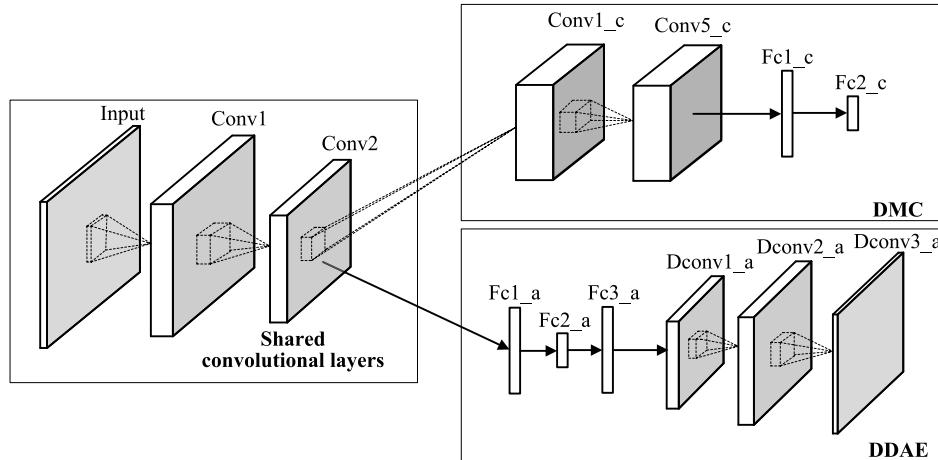


Fig. 5. Architecture of the DMNN that is composed of the DMC and DDAE sharing convolutional layers.

TABLE I
DETAILED PARAMETERS OF THE DMNN

Layer	Input	Kernel size	Stride	Output
Conv1	1×32×32	3×3	1	64×32×32
Conv2	64×32×32	3×3	2	48×16×32
Conv4_c	48×16×16	3×3	2	128×8×8
Conv5_c	128×8×8	3×3	1	128×8×8
Fc1_c	128×8×8	1×1	1	1024×1×1
Fc2_c	1024×1×1	1×1	1	2×1×1
Fc1_a	32×16×16	1×1	1	256×1×1
Fc2_a	256×1×1	1×1	1	30×1×1
Fc3_a	30×1×1	1×1	1	256×1×1
Dconv1_a	1×16×16	3×3	1	48×16×16
Dconv2_a	48×16×16	3×3	2	64×32×32
Dconv3_a	64×32×32	3×3	1	1×32×32

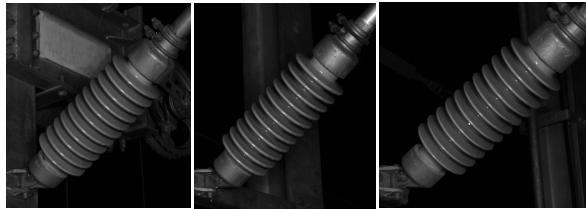


Fig. 6. Insulator images localized by the Faster R-CNN.

where p_i and \hat{p}_i are, respectively, the ground truth and predicted probabilities of the input patch being an insulator patch.

B. Deep Denoising Autoencoder

The goal of the autoencoder is to detect defects in the insulator image based on the reconstruction error. When there is a defective region in the input image, the reconstruction error of the defective area will be larger than that of the normal area. Due to the fact that direct reconstruction of the full image is intractable, and the defective area is usually a small part of the insulator surface, we instead reconstruct overlapping patches of the insulator image to detect surface defects.

A basic autoencoder is a neural network with only one hidden layer, which is trained to reconstruct its input with as little error as possible. The autoencoder network consists of two parts: an encoder $z = f(x)$ and a decoder $y = g(z)$, where z is the hidden layer that describes the low-dimensional feature of the input x . The autoencoder is trained using the backpropagation algorithm to minimize the loss

$$e(x, y) = \frac{1}{N} \sum_{i=1}^N \|x_i - g(f(x_i))\|^2. \quad (2)$$

However, this Euclidean loss encourages the autoencoder to learn an identity function, which is detrimental to the defect detection. The denoising autoencoder changes the loss and instead minimizes

$$\tilde{e}(x, y) = \frac{1}{N} \sum_{i=1}^N \|x_i - g(f(\tilde{x}_i))\|^2 \quad (3)$$

where \tilde{x}_i is the corrupted version of the corresponding x_i , which is obtained by corrupting x_i with some form of noise. Therefore, the denoising autoencoder can undo the corruption and increase the reconstruction error of defective patches relative to that of normal ones, which is beneficial for the defect detection.

The revolutionary success of deep neural network architectures has shown that the depth of the neural networks offers many advantages. In comparison with a shallow autoencoder, a deep autoencoder can not only reconstruct the input with a lower error but also can learn hierarchical features by its hidden layers, which substantially improve its ability to solve tasks.

The DDAE for insulator surface defect detection consists of two convolutional layers, three deconvolutional layers, and three fully connected layers, as shown in Fig. 5. Similar to (3), we use a loss with the L₂ regularization to train the denoising autoencoder

$$\tilde{e}_D(x, y) = \frac{1}{N} \sum_{i=1}^N \|x_i - g(f(\tilde{x}_i))\|^2 + \gamma \|W\|^2 \quad (4)$$

where W is the parameter of the DDAE and γ is a coefficient to balance the loss and the regularization.

C. Multitask Learning

The MTL is a transfer learning technique that learns two or more related tasks at the same time with the aim of a mutual benefit [28]. In the MTL framework, knowledge learned for one task is transferred to the other tasks, which helps the related tasks to be learned more effectively. Then, each task can benefit from reusing knowledge that has been learned in the training of other tasks.

The material classification and defect detection are two related tasks since both of them need to extract features of the insulator image. The material classifier needs the features to distinguish the insulator and background, and the DDAE needs them to reconstruct the patches. If we build a deep architecture for each task, they cannot benefit from each other, and the detection speed will be much slower. In our architecture, we unified the two neural networks into one by sharing convolutional layers, which can improve the features learned by these layers. Thus, it is reasonable to expect better results for each task when training them cooperatively.

Traditionally, the loss function L_M in MTL is the weighted sum of that in single-task learning [15], [29]

$$L_M = \sum_t^T \lambda_t \sum_i^{N_t} E_t(f(x_{ti}, y_{ti})) \quad (5)$$

where T is the number of tasks, y_{ti} is the ground truth label of the training sample x_{ti} , f is the multioutput function learned by the multitask neural network, E_t is the loss function of the task t , and λ_t is the weighting factor to balance different tasks.

However, in our DMNN, the DMC and DDAE require different training sets. The training of the DMC requires both insulator patches and background patches, whereas the DDAE only requires insulator patches for training. This difference means that the weighted sum of loss function is not suitable for our learning task. Therefore, we train the DMNN in an alternating way. The training procedure is composed of two phases: the first phase consists of updating the DMC, and the second phase consists of updating the DDAE. The details of the training algorithm are shown in Algorithm 1.

Algorithm 1 Pseudocode for Training the MTL

1: Input:

- X_c : training set for the DMC, including both insulator patches and background patches.
- X_d : training set for the DDAE, including only insulator patches of X_c .
- K_c : number of classifier training iterations per mini-batch.
- K_d : number of DDAE training iterations per mini-batch.

2: for number of training iterations do

- Sample mini-batch of m examples from the training set X_c .

3: for $k = 1$ to K_c do

- Update the DMC by minimizing the loss:

$$L_c = - \sum_i ((1 - p_i) \log(1 - \hat{p}_i) + p_i \log \hat{p}_i)$$

end for

- Sample mini-batch of m examples from the training set X_d .

- add random noise to the m sampled examples

4: for $k = 1$ to K_d do

- Update the DDAE by minimizing the loss:

$$\tilde{e}_D(x, y) = \frac{1}{N} \sum_{i=1}^N \|x_i - g(f(\tilde{x}_i))\|^2 + \gamma \|W\|^2$$

end for

end for

After training, the DMC and DDAE share convolutional layers and form a unified deep neural network DMNN to detect insulator surface defects.

Note that when the DMNN slides on the insulator to detect defects, too large sliding window is hard to reconstruct, while too small sliding window cannot contain sufficient structural information, both of which are detrimental to the defect detection. The size of the sliding window should be chosen to contain enough structural information in each window and make the window as small as possible. K_c and K_d are used to control the iterations for the DMC and the DDAE in per minibatch, respectively. The purpose of this strategy is to balance the two networks.

D. Decision

Once the DMNN slides through the insulator image, a classification score map and a reconstruction error map will be produced by the DMC and DDAE, respectively. Whether the patch in the sliding window is normal or defective is determined by the reconstruction error and classification score. In the decision phase, patches with a classification score larger than a predefined threshold T_c are classified as insulator patches and used for the defect detection.

Since the catenary in different line sections has different structures, the insulators in different images have different poses and sizes. However, the insulator image patches in the same image have similar features, and it is reasonable to assume that they have a similar reconstruction error.

TABLE II
COMPARISON OF DIFFERENT OBJECT DETECTION APPROACHES

Localizing methods	mAP (%)					
	Insulator	swivel clevis	clamp	Messenger wire holder	steady arm support	clevis end holder
Faster R-CNN ZF	99.5	99.4	99.2	98.9	98.7	99.3
Faster R-CNN VGG16	99.8	99.6	99.5	99.4	99.4	99.6
YOLO	98.8	97.6	98.9	98.5	98.9	98.4
SSD	99.0	98.6	98.4	98.8	98.3	99.3
YOLOv2	99.2	99.5	99.1	98.9	99.5	99.4



Fig. 7. KCIS-01 catenary inspection vehicle.

Thus, we define an adaptive anomaly score to quantitatively assess the surface defects of the insulators

$$s_i = \frac{e_i}{\text{median}(e_1, \dots, e_i, \dots, e_N)} \quad (6)$$

where N is the number of insulator patches in an insulator image. Then, during operations, we compute the anomaly score for each patch and consider patch i anomalous when s_i exceeds a predefined threshold T_a . This is equivalent to consider the patch anomalous if it has an indicator falling outside a confidence region around $\text{median}(e_1, \dots, e_i, \dots, e_N)$.

V. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of the proposed insulator defect detection system, we tested it on a catenary image data set of the Hefei–Fuzhou high-speed railway line. The data set was collected by the KCIS-01 catenary inspection vehicle, as shown in Fig. 7.

The data set contains approximately 18 000 key catenary components images with a resolution of 4920×3280 pixels. The experimental environment is as follows: Ubuntu 16.04, Python 2.7, deep learning framework Tensorflow, Intel Core i7-7600, and GTX 1070 GPU with 8-GB memory.

A. Key Catenary Components Localization

In this section, we evaluate the performance of the localizing algorithm based on the Faster R-CNN. The key catenary components' image data set was randomly divided into two parts: 12 000 images for training and 6 000 images for testing. The components, including the insulator, swivel clevis, clamp, messenger wire holder, steady arm support, and clevis end holder, were selected as key catenary components to be localized in the experiment. These parts play an important role in

the catenary system, and their risk of damage is comparatively high. Representative localization results of the key components are shown in Fig. 8.

We compared the performance of several deep learning object detection approaches, including the single shot multibox detector [14], YOLOv2 [31], and Faster R-CNN [13]. To make a fair comparison, all these algorithms used the same training data set and annotations. After training, we set the intersection-over-union threshold at 0.7 and used the mean average precision (mAP) metric to evaluate the object detection algorithms. The localization results of key components are summarized in Table II.

It can be found that the deep learning object detection approaches perform well in catenary component localization. Compared with the other listed methods, the Faster R-CNN VGG16 outperforms other approaches in all the components localizing except steady arm support, which also has comparable mAP with YOLOv2 for steady arm support localization and was selected to localize the key components.

B. Material Classification and Defect Detection

In this section, we evaluate the performance of the proposed defect detector DMNN. In the experiment, we selected 1000 representative insulator images that have been localized by the Faster R-CNN and then annotated the boundaries of these insulators manually. In each insulator foreground patches within the insulator boundary and 300 background patches outside the insulator boundary were randomly selected to create three data sets containing 500 000 patches each. The size of the patches in the three data sets is 16×16 , 32×32 , and 48×48 pixels, respectively. As shown in Algorithm 1, in our multitask architecture, the DMC and DDAE require different training sets. From each data set, we randomly selected 400 000 patches as the training set of the DMC, and the remaining 100 000 patches were used as the DMC test sets. The 200 000 foreground patches were used as the training set of the DDAE. In addition, the 20 000 patches including 10 000 synthetic defect patches were used as the validation set to analyze the effect of key parameters on the performance of the DDAE.

We trained the DMNN using Algorithm 1 with a batch size of 128, a momentum of 0.9, and a weight decay of 5×10^{-5} for a total of 200 epochs. The learning rate was initially set to 0.001 and decays by a factor of 0.5 every 50 epochs. The performance of the DMNN was evaluated on a real-life defective insulator image data set.

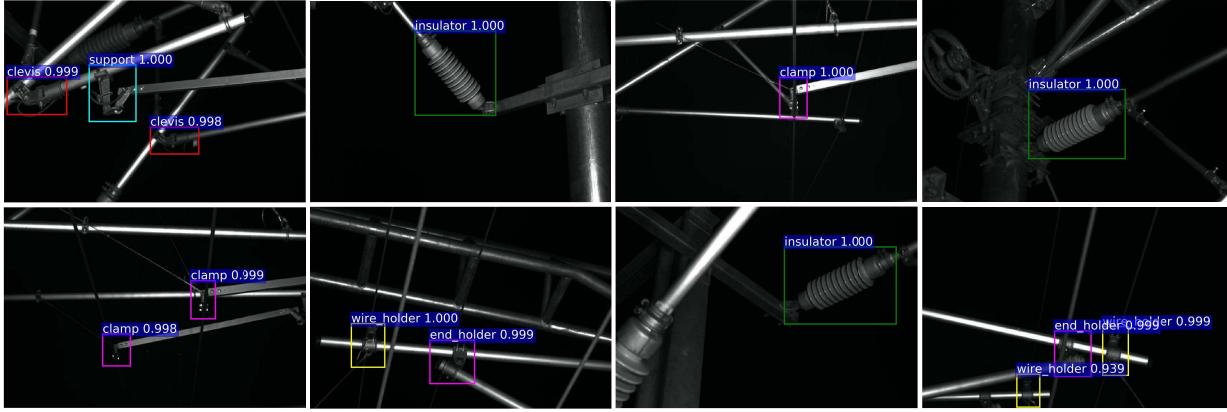


Fig. 8. Representative key components localization results by the Faster R-CNN.

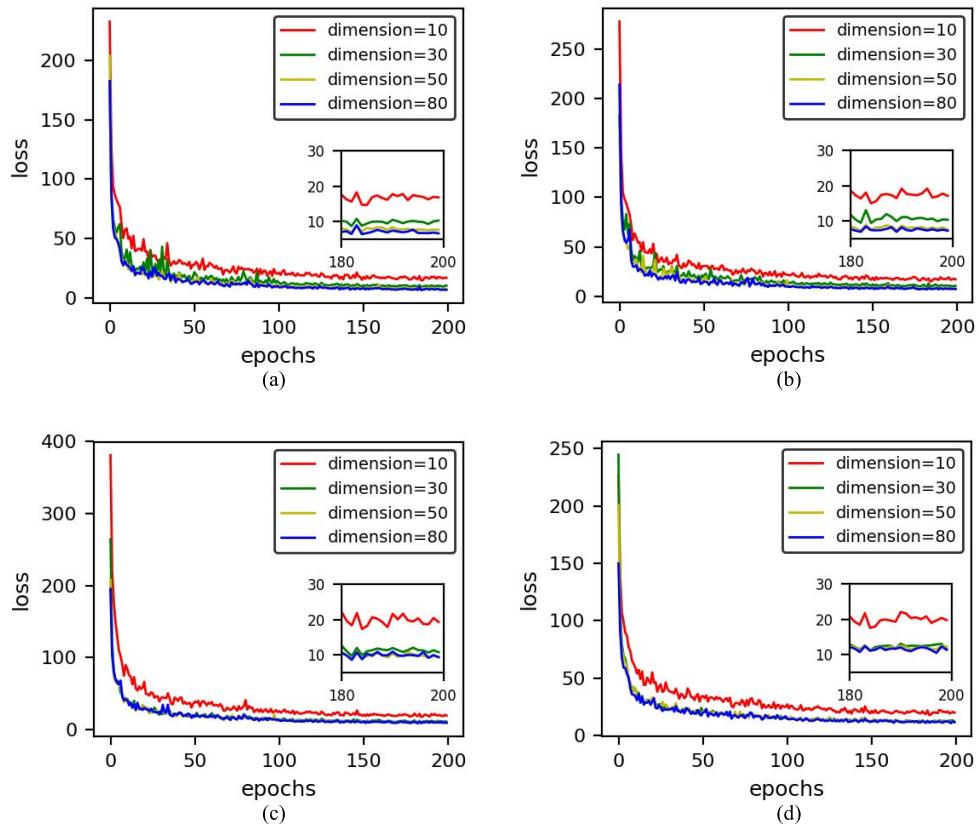


Fig. 9. Under different noise intensities, training losses of the DDAE with different Fc2_a dimensions. (a) No noise. (b) SNR = 20 dB. (c) SNR = 12 dB. (d) SNR = 8 dB.

1) Analysis of Parameters of DDAE: The three key parameters that have significant impacts on the performance of DDAE are the Fc2_a dimension, the intensity of noise added to the training set, and the size of insulator patches. First, the Fc2_a dimension is one of the key factors that determine the reconstruction capacity of the autoencoder. If the autoencoder is allowed to have too much capacity, the autoencoder can learn to perform the copying of the input, without extracting useful information of the input data. On the contrary, if the dimension of the Fc2_a layer is too low, the reconstruction capability is weak, resulting in a large reconstruction error,

which is not conducive to fault detection. Then, the denoising autoencoder can undo the corruption and increase the reconstruction error of defective patches relative to that of normal ones. However, deficient noise intensity does not significantly improve the fault detection capability, while excessive noise intensity can damage the structural information of the image. Finally, the size of insulator patches determines the difficulty of reconstruction, which also has an important impact on defect detection.

In this section, we thoroughly analyze the effect of the three key parameters. First, we analyzed the first two key

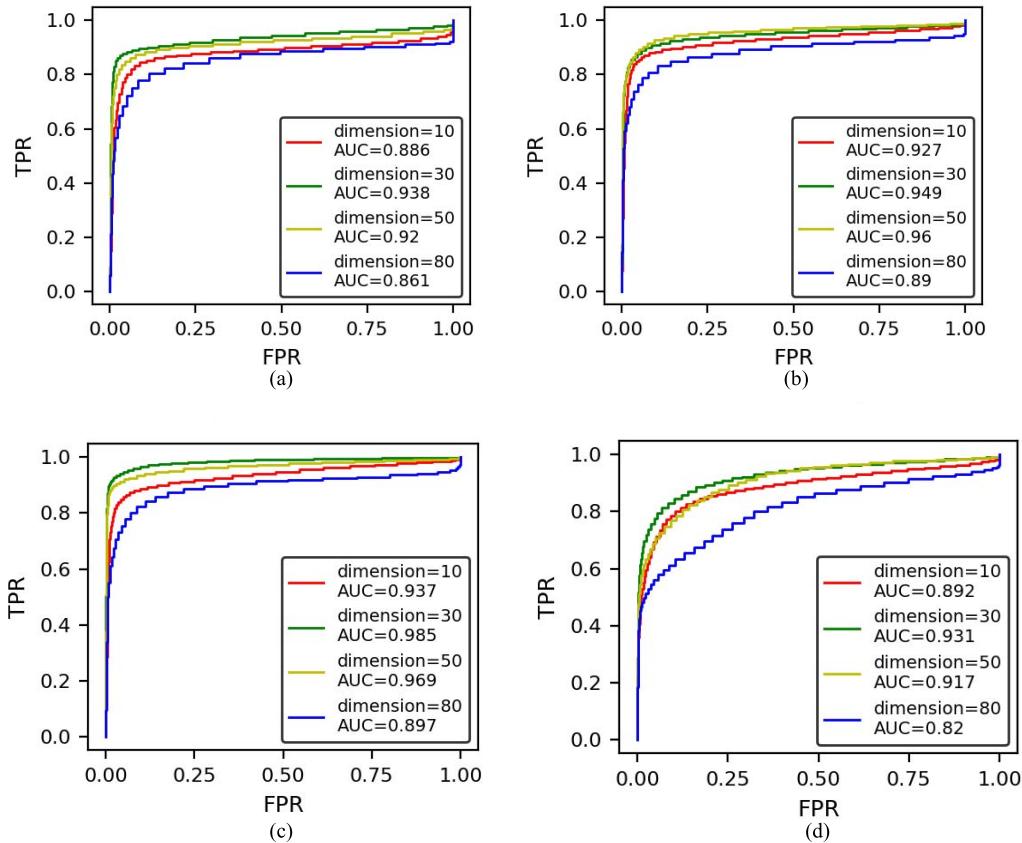


Fig. 10. Under different noise intensities, the ROC curves for the DDAE with different Fc2_a dimension, and the corresponding AUC values reported in the legend. (a) No noise. (b) SNR = 20 dB. (c) SNR = 12 dB. (d) SNR = 8 dB.

parameters using the data set containing patches of size 32×32 pixels. The networks with different Fc2_a dimensions $d = \{10, 30, 50, 80\}$ were trained on both the clean training set and the noisy training sets with different signal to noise ratio $\text{SNR} = \{20, 12, 8\}$ dB. The SNR is defined as follows:

$$\text{SNR} = 20 \log_{10} \left(\frac{A_s}{A_n} \right) \quad (7)$$

where A_s is the maximum gray-scale value of the images in the training set, and A_n is the maximum expectation of the noise added to the training set. Given the SNR and A_s , the corresponding A_n can be calculated according to (7). To prevent the network from overfitting to some certain kinds of defect, the following noise addition process was used to increase the diversity of the noise. First, randomly sample from the uniform distribution $U(0, A_n)$ and $U(0.1 A_n, 0.3 A_n)$ to obtain E_i and σ_i , respectively. Then, the noise was obtained by randomly sampling from the normal distribution $N(E_i, \sigma_i)$. Finally, the noise was added to the image i of the current minibatch. Fig. 9 investigates the effects of the Fc2_a dimension and intensity of noise on the training loss of the DDAE.

It can be found that the training losses increase with the increase in noise intensity and decrease with the increase in Fc2_a dimension. However, when the Fc2_a dimension of the

network exceeds 30, the decrease in training loss is no longer obvious, especially when the SNR is less than 12.

In the testing phase, the real-life defective samples are rare and not sufficient to analyze the key parameters of DDAE. However, when the insulator is dirty or damaged, the corresponding area in the image will be darker than normal. We synthesized the defective patches by darkening a 5×5 pixels region in the normal patches and used them to analyze the key parameters of DDAE. The validation set containing 10000 synthetic defect patches and 10000 normal insulator patches were used to analyze the role of the two key parameters. The evaluation metric used was the receiver operating characteristic (ROC) curve that is obtained by plotting the true positive rate against the false positive rate for different discrimination thresholds. Moreover, the area under the ROC curve (AUC) was used as a single quantitative performance indicator.

It can be observed in Fig. 10 that both the Fc2_a dimension and the noise intensity have significant impacts on the defect detection performance of the DDAE. The DDAE with the Fc2_a dimension of 30 achieved the best performance under all noise intensities except for SNR = 20 dB. In contrast, although the networks with the Fc2_a dimension of more than 30 have obtained lower training losses, their performances in defect detection have not been improved accordingly. The networks with the Fc2_a dimension of 80 have the poorest

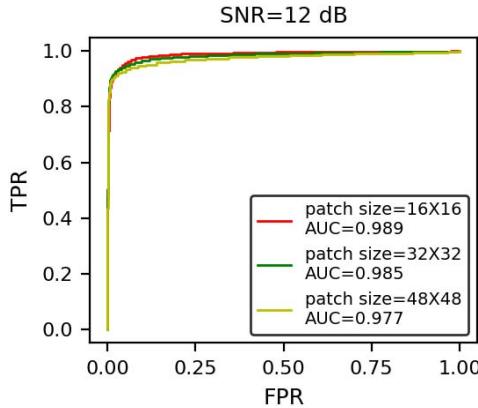


Fig. 11. Comparison of the best performance of the DDAEs trained using patches of different sizes.

defect detection performance under all noise intensities. This result is due to differences in the reconstruction capabilities of the DDAE with different Fc_2_a dimensions. When the dimension of the latent space is too high, the DDAE tends to learn merely an identity function. When the dimension of the latent space is too low, the DDAE cannot reconstruct the normal patches perfectly. Both of these conditions are not good for defect detection.

When trained using training set with $SNR = 12$ dB, the DDAE achieved the best performance compared with those trained with other noise intensities. When no or too much noise was added to the training set, the DDAEs did not perform as well. This result is because the DDAE can learn more useful features than the basic autoencoder and excessive noise can cause information loss of insulator patches.

Finally, using the data set containing patches of 16×16 and 48×48 pixels, the DDAEs with different Fc_2_a dimensions were trained on training sets with different SNRs. They achieved the best detection results when the Fc_2_a dimension was 10 and 30 and the SNR was 12 and 12, respectively. Comparison of the best performances of the DDAEs trained using patches of different sizes is shown in Fig. 11.

The performance of DDAE was improved as patch size decreased. This result is because the reconstruction difficulty will decrease with the decrease in patch size, which is more conducive to the defect detection.

2) Material Classification: In this section, we first analyze how different patch sizes affect the result of DMC. The DMCs were trained on the three different data sets containing patches of different sizes, and their performances were compared in Fig. 12.

It can be observed the performance of DMC was improved as patch size increased. This result is because the too small patch cannot contain enough structural information that is essential for classification.

In addition to the proposed DMC, we have also implemented and evaluated the following alternative methods. To make a fair comparison, all these algorithms use the same training data set containing patches of 32×32 pixels for material classification.

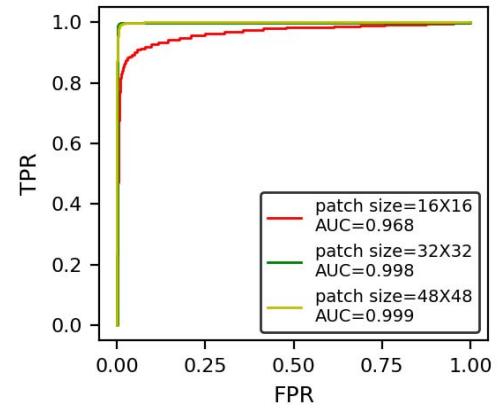


Fig. 12. Comparison of the performances of the DMCs trained using patches of different sizes.

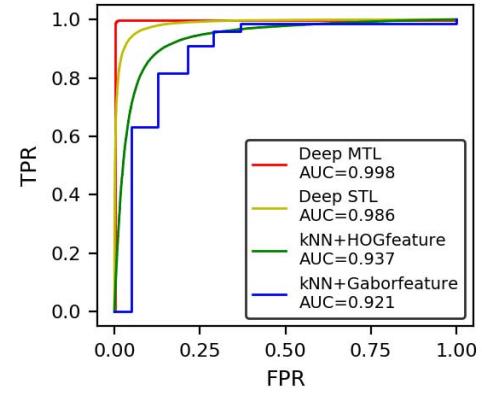


Fig. 13. ROC curves for different classifiers with the corresponding AUC values reported in the legend.

a) Deep MTL: The proposed DMC described in Section IV.

b) Deep STL: Single-task deep neural network that has the same architecture as the DMC.

c) HOG feature with a k-nearest neighbor: As a local handcraft feature descriptor, the histogram of gradient feature [33] is invariant to light and rotation. To select the optimal dimension, the histogram of oriented gradient (HOG) features of different dimensions were extracted to train and test the k-nearest neighbor (kNN) classifier. The optimal number of dimension is found to be 324.

d) Gabor feature with a k-nearest neighbor: To extract the feature of the images, a bank of filters was used to filter each image. As in [33], the mean and standard deviation of the output of each filter are computed to build a feature descriptor. To select the optimal feature, filter banks of different scales and orientations were used to extract features for the training and testing of the kNN classifier. When the filter bank is with six scales and six orientations, filters achieved the best result.

It can be observed in Fig. 13 that the proposed method is the most accurate, followed by the deep single-task learning (STL) network and HOG with the kNN classifier method. The Gabor feature performs poorly in our data set. Both deep classifiers have achieved better performance than that of the handcraft features-based material classifier. In addition, although STL

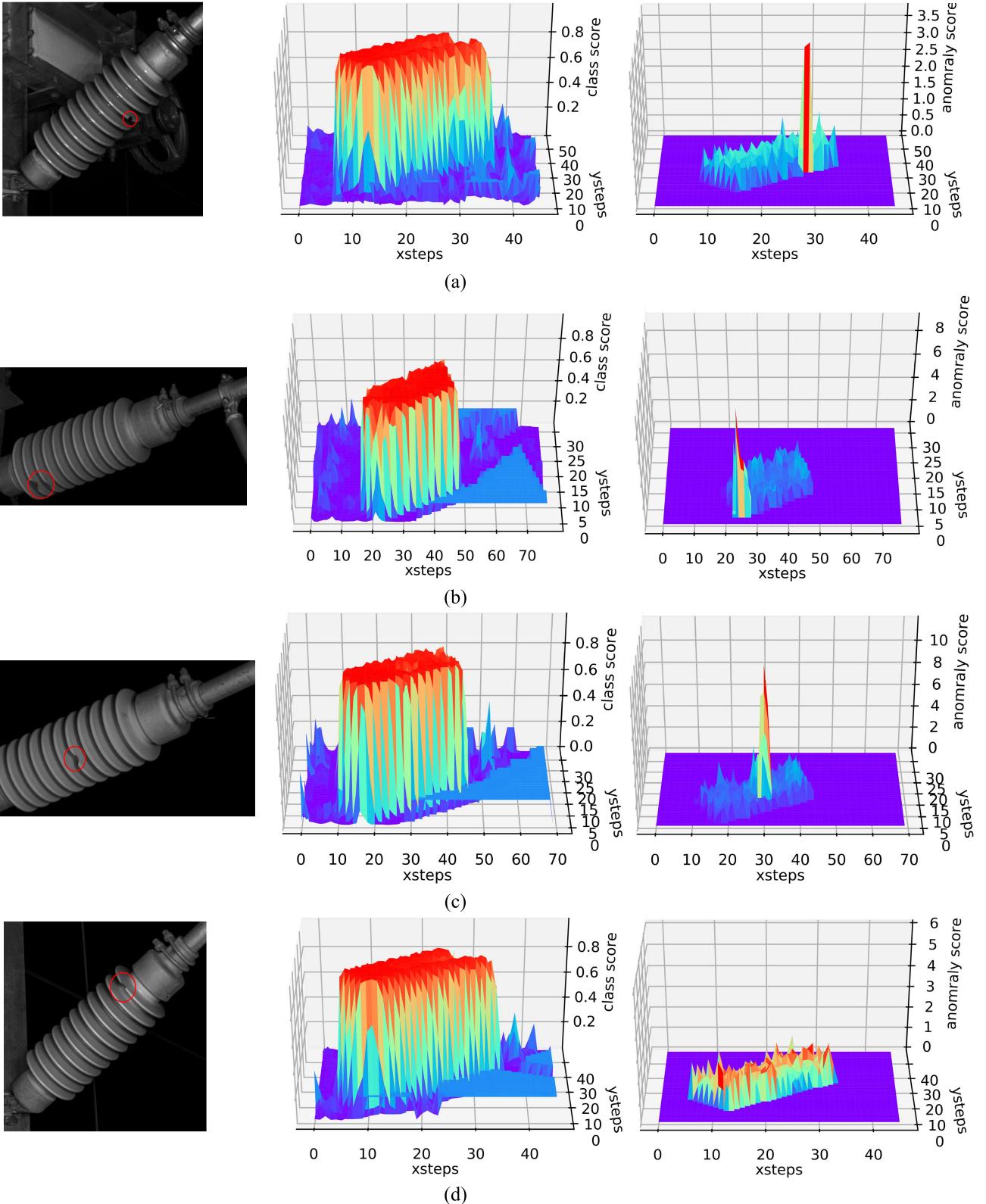


Fig. 14. Classification scores and anomaly scores of the real-life defective insulators. (a) and (b) Dirty insulators. (c) Damaged insulator. (d) Failure case.

has the same architecture as the DMC, the performance of STL is still worse than that of the DMC, which indicates that the deep classifier can benefit from MTL.

3) Defect Detection: According to the above-mentioned experimental results, the DMNN with the Fc2_a dimension

of 30, in which the DDAE was trained with a noise intensity of SNR = 12 dB and patches of size 32×32 pixels, was chosen to detect the real-life insulator defects. In the testing set, there were 1000 insulator images, including 72 real-life defective images located by the Faster R-CNN. The step size

TABLE III

COMPARISON OF THREE ANOMALY DETECTION APPROACHES

Detection approaches	tp	fp	fn	F1-score
SPC	62	11	10	0.86
SDAE	65	13	7	0.87
DDAE	71	6	1	0.95

of the sliding window in the experiment is 12 pixels. When the classification score of the sliding window was greater than 0.85, we classified it into insulator patch and calculated its anomaly score according to (6). Fig. 14 shows examples of the material classification and defect detection results of the insulator images. The first column is the insulator images located by the Faster R-CNN, and the second and third columns are their corresponding classification and anomaly scores.

It can be found that the insulator patches obtain much higher classification scores than those of the background patches, and the insulators are successfully segmented from the background. The anomaly score figures clearly depict the low anomaly scores of normal patches and the high anomaly scores of the defective patches such as damaged or dirty. This indicates that the anomaly score can be used to effectively detect insulator defects. However, as shown in Fig. 14(d), the DMNN cannot detect the defect when the defective area is small and the gray value of the defect area only slightly differs from that of the normal area.

To further evaluate the performance of our method, we compared our algorithm with two similar defect detection approaches that do not require defective patches for training. To make a fair comparison, all the algorithms used the same training set and the same anomaly score defined in (6). All these algorithms detect insulator defects based on the material classification results of the DMC. The evaluation metric was the F1-score

$$F_1 - \text{score} = \frac{2 \times tp}{2 \times tp + fn + fp} \quad (8)$$

where tp is the number of correctly detected defective insulators, fp is the number of normal insulators misclassified as defective insulators, and fn is the number of defective insulators misclassified as normal insulators.

a) **SPC:** Sparse coding [34] first learns an overcomplete dictionary D of normal patches, and then determines whether the patches are defective or not by the reconstruction error. To select the optimal dimension of D , the dictionary with different dimensions (from 1000 to 5000) was trained to detect defects, and the 2000-dimensional dictionary achieved the best result.

b) **SDAE:** Single-task deep autoencoder without denoising training, which has the same architecture as the DDAE.

c) **DDAE:** The proposed method described in Section IV.

It can be found in Table III that the DDAE only misses one defect and falsely reported six defects. Compared with other approaches listed, the DDAE has achieved the best performance. In addition, although the single-task deep autoencoder (SDAE) has the same architecture as the DDAE in the MTL, the performance of the SDAE is still worse than that of the DDAE, which indicates that the deep autoencoder can benefit from MTL and denoising training.

VI. CONCLUSION

This paper presented a method to detect railway insulator surface defects. The proposed two-stage architecture can automatically localize the six kinds of key components and detect the insulators surface defects in the captured catenary images. The Faster R-CNN was applied to localize the key components. A multitask DCNN DMNN that integrates a DMC and a DDAE was proposed to segment insulators and detect their surface defects at the same time. We thoroughly analyzed the parameters of the DDAE and compared the DMC and DDAE with the related alternative methods. The experimental results demonstrate that the DMNN has a superior performance over those of the other tested alternatives and can effectively detect insulator surface defects. Thus, the proposed approach can be implemented in the catenary inspection system to detect the insulator surface defects. Further research focuses on improving the shared feature extraction network of the DMNN and applying it to detect defects of other key catenary components.

REFERENCES

- [1] E. Karakose, M. T. Gencoglu, M. Karakose, I. Aydin, and E. Akin, "A new experimental approach using image processing-based tracking for an efficient fault diagnosis in pantograph–catenary systems," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 635–643, Apr. 2017.
- [2] *Fittings for Overhead Contact System in Electrification Railway*, Standard TB/T 2075.1, China Railway Industrial, 2010.
- [3] 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.
- [4] Q. Li and S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2189–2199, Aug. 2012.
- [5] L. Liu, F. Zhou, and Y. He, "Automated visual inspection system for bogie block key under complex freight train environment," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 2–14, Jan. 2016.
- [6] 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.
- [7] S. Liao and J. An, "A robust insulator detection algorithm based on local features and spatial orders for aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 963–967, May 2015.
- [8] M. Oberweger, A. Wendel, and H. Bischof, "Visual recognition and fault detection for power line insulators," in *Proc. 19th CVWW*, 2014, pp. 1–8.
- [9] Y. Han, Z. Liu, D.-J. Lee, G. Zhang, and M. Deng, "High-speed railway rod-insulator detection using segment clustering and deformable part models," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 3852–3856.
- [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.*, Jun. 2016, pp. 779–788.
- [11] 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.*, Jun. 2014, pp. 580–587.
- [12] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1440–1448.
- [13] 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.
- [14] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [15] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.
- [16] 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.

- [17] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [18] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.
- [19] S. Faghhih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter, "Deep convolutional neural networks for detection of rail surface defects," in *Proc. Int. Joint Conf. Neural Netw.*, 2016, pp. 2584–2589.
- [20] J. Chen, Z. Liu, H. Wang, A. Núñez, 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.
- [21] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009.
- [22] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [23] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 1996–2000.
- [24] H. Sohn, K. Worden, and C. R. Farrar, "Statistical damage classification under changing environmental and operational conditions," *J. Intell. Mater. Syst. Struct.*, vol. 13, no. 9, pp. 561–574, Sep. 2002.
- [25] H. Liu, T. Taniguchi, Y. Tanaka, K. Takenaka, and T. Bando, "Visualization of driving behavior based on hidden feature extraction by using deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2477–2489, Sep. 2017.
- [26] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 733–742.
- [27] T. Grozdić and S. T. Jovičić, "Whispered speech recognition using deep denoising autoencoder and inverse filtering," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 12, pp. 2313–2322, Dec. 2017.
- [28] R. Caruana, *Multitask Learning*. New York, NY, USA: Springer-Verlag, 1998.
- [29] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.
- [30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [31] J. Redmon and A. Farhadi. (Dec. 2016). "YOLO9000: Better, faster, stronger." [Online]. Available: <https://arxiv.org/abs/1612.08242>
- [32] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.
- [33] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
- [34] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2013, pp. 1–6.



Gaoqiang Kang 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 image processing, machine learning, computer vision, and fault detection.



Shibin Gao received the Ph.D. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China.

Since 1998, he has been a Full Professor with the Department of Electrical Engineering, Southwest Jiaotong University. His current research interests include power system protection and automation, online monitoring of electrical equipment, rail transit traction power supply system security, and railway infrastructure service status monitoring.



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.