Unified deep learning architecture for the detection of all catenary support components

Liu, Wenqiang; Liu, Zhigang; Nunez, Alfredo; Han, Zhiwei

DOI
10.1109/ACCESS.2020.2967831

Publication date
2020

Document Version
Final published version

Published in
IEEE Access

Received November 28, 2019, accepted January 11, 2020, date of publication January 20, 2020, date of current version January 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967831

Unified Deep Learning Architecture for the Detection of All Catenary Support Components

WENQIANG LIU[1], (Student Member, IEEE), ZHIGANG LIU[1], (Senior Member, IEEE), ALFREDO NÚÑEZ[2], (Senior Member, IEEE), AND ZHIWEI HAN[1], (Member, IEEE)

[1]School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China
[2]Section of Railway Engineering, Delft University of Technology, 2628 Delft, The Netherlands

Corresponding author: Zhigang Liu (liuzg_cd@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1734202 and Grant 51977182.

ABSTRACT With the rapid development of deep learning technologies, researchers have begun to utilize convolutional neural network (CNN)-based object detection methods to detect multiple catenary support components (CSCs). The literature has focused on the detection of specified large-scale CSCs. Additionally, CNN architectures have faced difficulties in identifying overlapping CSCs, especially small-scale components. In this paper, a unified CNN architecture is proposed for detecting all components at various scales of CSCs. First, a detection network for CSCs with large scales is proposed by optimizing and improving Faster R-CNN. Next, a cascade network for the detection of CSCs with small scales is proposed and is integrated into the detection network for CSCs with large scales to construct the unified network architecture. The experimental results demonstrate that the detection accuracy of the proposed CNN architecture can reach 92.8%; hence, it outperforms the popular CNN architectures.

INDEX TERMS High-speed railway catenary, catenary support component detection, deep learning architecture.

I. INTRODUCTION

The catenary system is an essential part of a high-speed railway, which provides a stable power transmission line for the operation of trains. Failure events in the catenary system could lead to the interruption of a complete railway line, which could cause substantial economic losses and affect the availability and safety of the railway services [1]. Therefore, for preventing failures, many monitoring systems have been proposed in the literature [2]–[6]. The effective inspecting and monitoring systems provide the railway infrastructure manager with the possibility of repairing components prior to their failure. However, due to the complexity of the catenary system (see Figure 1), which includes many components and distributes over several kilometers of railway tracks, accurately detecting and continuously monitoring all components is a massive challenge. Without the robust detection of components, the early prognosis of failures is not possible. The current research on the detection of catenary support components (CSCs) focuses mainly on the detection of a single CSC or several easy-to-detect CSCs with large scales (LSCSCs). Han *et al.* [7] presented an effective method for the detection of catenary rod-insulators (component number 2 in Figure 1). The method used a segment clustering algorithm to divide the images and detected the rod-insulators using deformable part models. Zhang *et al.* [8] used the contourlet transform to extract insulator feature information based on the anisotropy and directionality of catenary rod-insulators (component number 2 in Figure 1). The Chan-Vese model was used to detect the insulator boundaries to locate each insulator. Kang *et al.* [9] presented a novel architecture for catenary insulator (component number 2 in Figure 1) surface defect detection. The architecture consisted of two stages: The first stage used a convolutional neural network to localize the key catenary components. Then, the second stage used a deep multitask neural network to obtain the classification score and anomaly score and to determine the defect state of the insulators. Liu *et al.* [10] presented a high-precision loose strand diagnosis approach for catenary isoelectric lines (component number 7 in Figure 1) by improving the ZF-Net convolutional network. Han *et al.* [11] proposed an automatic visual inspection method for fracture detection in catenary

The associate editor coordinating the review of this manuscript and approving it for publication was Santhosh Kumar Gopalan.

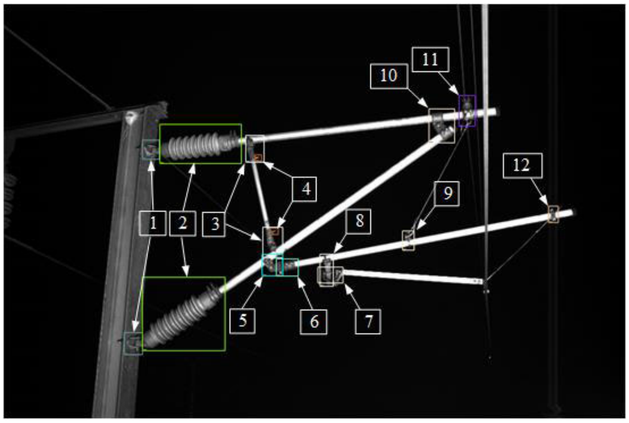


FIGURE 1. **Catenary support components. 1) Insulator base; 2) Insulator; 3) Brace sleeve; 4) Brace sleeve screw; 5) Rotary double-ear; 6) Binaural sleeve; 7) Isoelectric line; 8) Steady arm base; 9) Bracing wire hook; 10) Double sleeve connector; 11) Messenger wire base; 12) Windproof wire ring.**

clevises (components number 5 and 6 in Figure 1). This method utilized the contextual information of the regions of interest in the image to modify the faster region-based convolutional neural network to detect the clevises. A crack detection process was proposed for recognizing cracks by utilizing the wavelet entropy. Chen *et al.* [12] proposed a defect detection method for catenary fasteners (components number 3, 5, 6, and 10 in Figure 1) that uses a three-level cascade deep convolutional neural network. The first network was used to detect the LSCSCs (diagonal tubes, clevises, and joints) and the second network was used to detect the small-scale fasteners of the CSCs based on the results of the first network. Last, the located and cropped images of the components were sent to the third light-network for defect detection of the fasteners.

Zhong *et al.* [13] proposed a high-precision three-stage automatic defect detection method for catenary split pins (components 3, 5, and 6 in Figure 1) that is based on an improved deep convolutional neural network, namely, PVANET++. The detection scheme was similar to that in reference [12]. First, the LSCSCs were detected by the first network. Then, the CSCs with small scales (SSCSCs) were detected by the second network. Last, the defect state of the split pins was determined. In the literature, the proposed methods have yielded satisfactory results for the detection of a specified component. However, for the same catenary image data set, the use of a dedicated approach for the detection of each component is inefficient because the common features of the CSCs that are processed by one approach could be used to facilitate the detection of other components. Moreover, simultaneously loading these methods into the inspection system and applying them to the detection of CSCs will consume a substantial amount of system resources and increase system costs. In addition, for the detection of the overlapping SSCSCs, an independent cascade network is typically used to detect them in the literature, which will pose new problems. Therefore, research on multi-component detection under a unified architecture is crucial. With the rapid development of deep learning (DL) technology, which overcomes many shortcomings of traditional object detection methods, researchers have begun to focus on the CNN technology for multi-component detection. To the best knowledge of the authors' knowledge, the literature is focused on the detection of several easy CSCs, and the detection of all CSCs has rarely been reported [1], [14].

Many state-of-the-art deep convolutional neural networks (DCNNs) have been proposed and successfully applied to various fields [15], [16]. In the literature, there are two mainstream network architectures for multi-object detection: region-based architectures (e.g., Faster R-CNN [17] and R-FCN [18]) and regression-based architectures (SSD [19] and YOLO [20]). Both architectures have advantages and disadvantages [1]. As reported in [1] for CSC detection, the region-based Faster R-CNN can more accurately detect multiple scales of CSCs. However, it performs poorly on the detection of overlapping SSCSCs; this is because when the region proposal network (RPN) of Faster R-CNN predicts the regions of interest (RoIs) of CSCs, non-maximum suppression (NMS) is used to reduce the number of these predicted regions by merging the overlapping regions, which will cause the overlapping SSCSCs to be absorbed by LSCSCs. In contrast, the regression-based architectures can detect SSCSCs; however, the detection accuracy is not high. This is because the regression-based architectures are sensitive to the locations of objects, and they allow only one object to be detected in the divided areas.

To overcome these problems, a novel CNN architecture for the detection of all CSCs is proposed. The proposed network architecture follows the main body structure of Faster R-CNN. The main contributions of this paper are summarized below:

(1) Through the performance analysis of Faster R-CNN, the detection network of LSCSCs is proposed by optimizing and improving Faster R-CNN, including a) analyzing the feature maps of the feature extraction network (FEN) and selecting the optimal feature map; b) adding the max-pooling layer before the region proposal network (RPN) to deduce the RPN size and increase the system real-time; c) analyzing the zero fraction of the classification and regression network (CRN) and reducing the size of CRN.

(2) Inspired by the cascade strategy, the detection network of SSCSCs is proposed and integrated into the detection network of LSCSCs to build a unified network architecture, which not only compensates for the shortcomings of Faster R-CNN but also overcomes the disadvantages of traditional cascade networks.

The remainder of this paper is organized as follows: First, the functional networks of the proposed network architecture are introduced in Section II. Then, the methodology and architecture design are explained in Section III. Next, the experiment results are analyzed and discussed in Section IV. Finally, the conclusions of this work are summarized and further research is discussed in Section V.

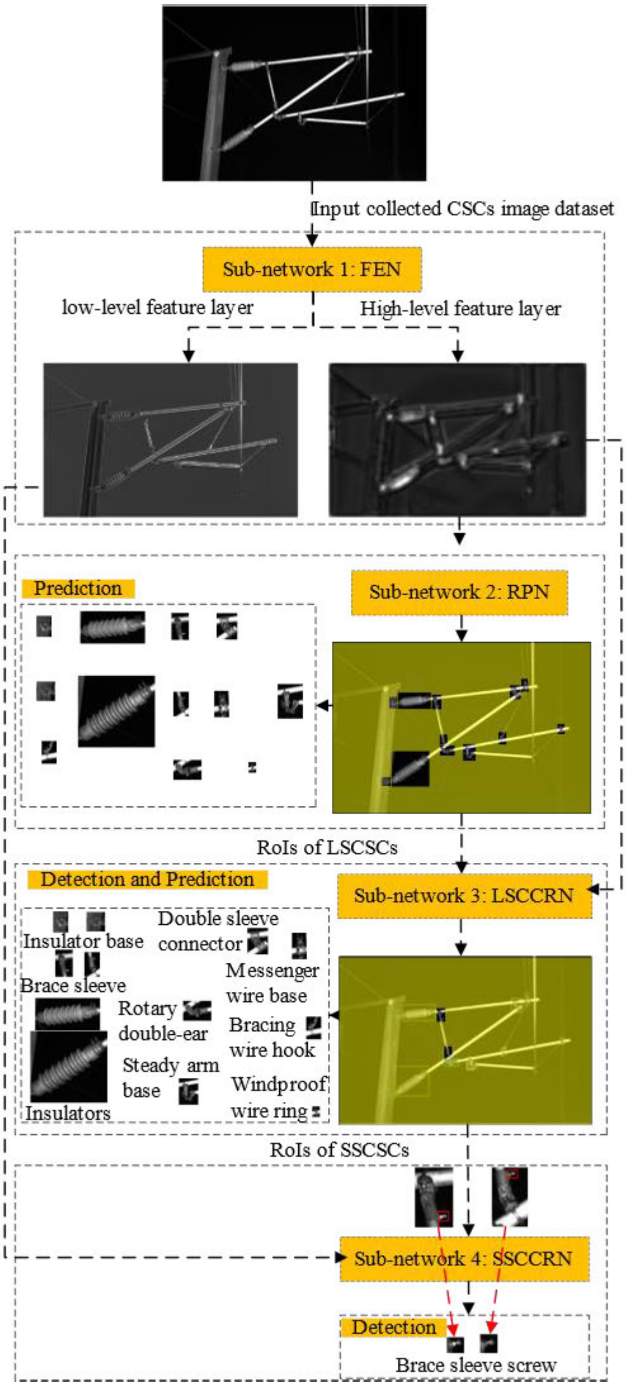


FIGURE 2. **Overview of the proposed CSCD-Architecture.**

II. OVERVIEW OF THE PROPOSED NETWORK ARCHITECTURE

An overview of the proposed network architecture, namely, the CSCD-Architecture, is presented in Figure 2. The CSCD-Architecture consists of four function networks:

1) First, catenary images are collected from the catenary inspection vehicle and stored on an on-board server. Then, the catenary images are sent to a feature extraction network (FEN) for extraction of the low-level and high-level features of all CSCs.
2) Next, the high-level features of CSCs are sent to a region proposal network (RPN) for prediction of the RoIs of the LSCSCs.
3) Then, combining the high-level features of FEN with the RoIs of the LSCSCs from RPN, the high-level features of the RoIs are obtained and sent to a classification and regression network (CRN) for detection of the LSCSCs. This network is named LSCCRN.
4) Last, according to the context relationship between LSCSCs and SSCSCs, the low-level features of FEN and the RoIs of SSCSCs from the localization results of LSCCRN are combined, and the low-level features of the RoIs are obtained and sent to a CRN for detection of the SSCSCs. This network is named SSCCRN.

III. METHODOLOGY AND ARCHITECTURE DESIGN

To build the proposed CSCD-Architecture for the detection of all CSCs, Faster R-CNN is optimized and improved to obtain the LSCCRN. Then, the cascade strategy is introduced into the proposed network architecture to obtain the SSCCRN. Faster R-CNN and the proposed CSCD-Architecture are illustrated in Figure 3, and the design principles of the proposed network architecture are described below.

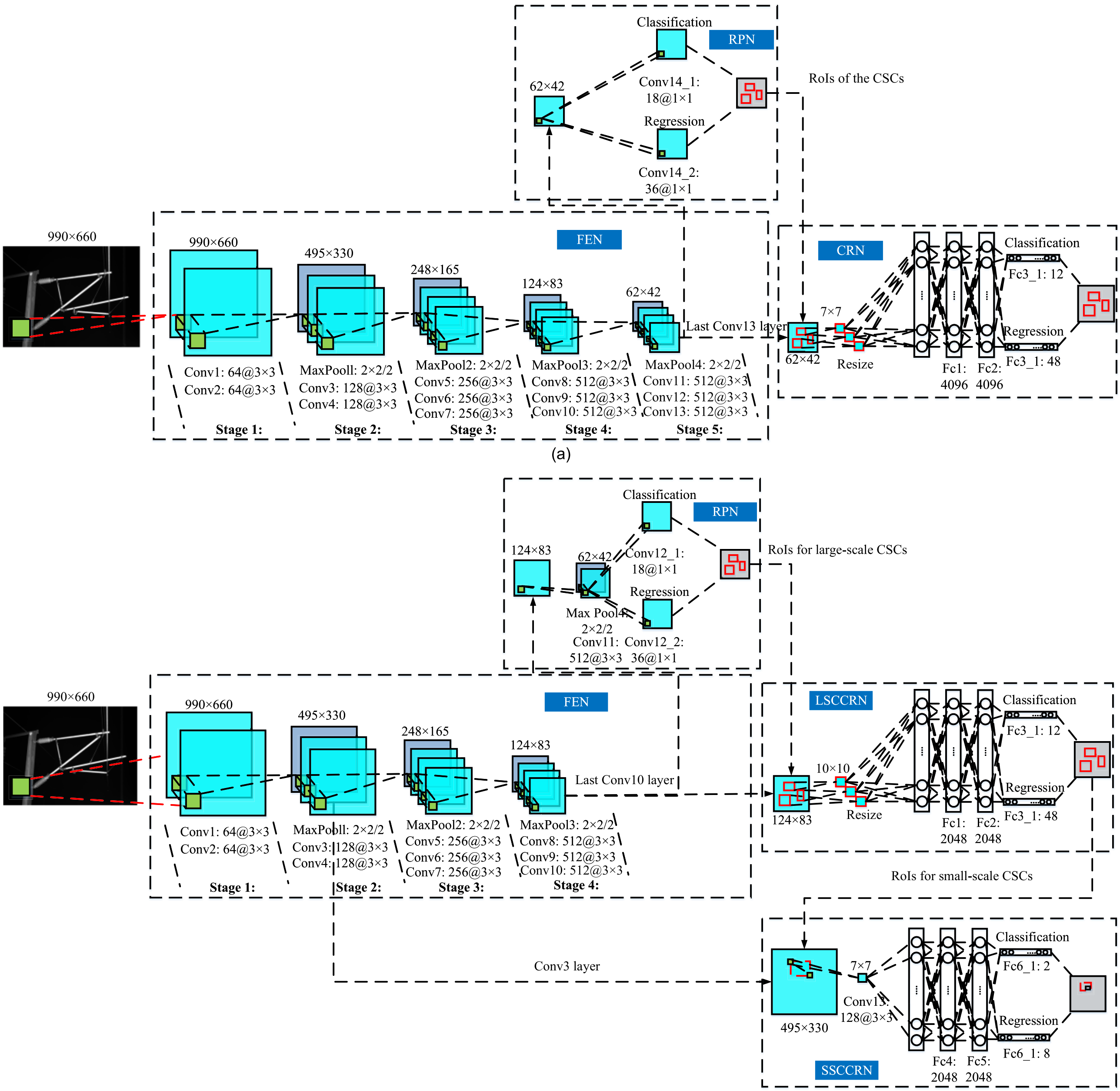
A. FEATURE EXTRACTION NETWORK

CNNs are widely used as the FEN to extract object features due to their characteristics. These features are obtained via repeated convolution and pooling operations. Theoretically, the levels of feature maps of CNNs represent various spatial resolutions and semantic information. The low-level feature maps of CNNs have a higher spatial resolution but are coarser, and they are suitable for small-scale object detection. In contrast, the high-level feature maps have more semantic information but are more abstract, and the features of objects with small scales easily disappear; they are ideal for large-scale object detection.

In Figure 3 (a), the FEN of Faster R-CNN uses the VGG16 network, which consists of five stages, to extract object features. In this paper, the Faster R-CNN is used to try to detect all CSCs, and the feature maps of various layers of the VGG16 network are presented in Figure 4. According to the analysis of the feature maps of the VGG16 network, the feature map (Conv 10) in Figure 4 (j) is sufficient for identifying CSCs. Therefore, the proposed network architecture crops the VGG16 network to remove the fifth stage of the VGG16 network and chooses Conv 10 as the last feature map, as illustrated in Figure 3 (b).

B. REGION PROPOSAL NETWORK

The RPN is widely used to reduce the computational complexity of network architectures, in combination with the characteristics of the anchor mechanism, for predicting regions of interest (RoIs) of objects, as shown in Figure 5. First, the anchors generate reference boxes of multiple predefined scales and aspect ratios to cover various scales of objects. These boxes are used as the criteria for deciding



(a)

(b)

FIGURE 3. Structural comparison between (a) Faster R-CNN and (b) the proposed CSCD-Architecture.

whether an object is present in an image and are used to calculate the bounding-box regression. Moreover, non-maximum suppression (NMS) is used to reduce the predicted RoIs by merging the overlapping RoIs based on their intersection-over-union (IoU).

In Figure 3 (a), the RPN of Faster R-CNN from top to bottom contains two sibling convolution operations for classification and regression. However, since the proposed network architecture chooses Conv 10 as the last feature map, the size of the last feature map increases and causes the RoI prediction of the RPN to take more time. To improve the efficiency of the RPN, based on an analysis of Figure 4, the feature map (Conv 11) of the VGG16 network performs excellently in predicting the RoIs of CSCs. Therefore, we add a MaxPool layer and a convolution operation to the proposed RPN to improve the real-time performance and to ensure the accuracy, as illustrated in Figure 3(b). For the catenary data set, the resolution of the original catenary images that are acquired from the inspection vehicle is 6600×4400. The resolution is too large, and it is difficult to process directly. Therefore, these images are resized from 6600×4400 to 990×660 as input images. By counting the image sizes of the resized CSCs, the size range of LSCSCs is found to be approximately 45 to 200. Therefore, in the proposed RPN, the anchors are assigned to

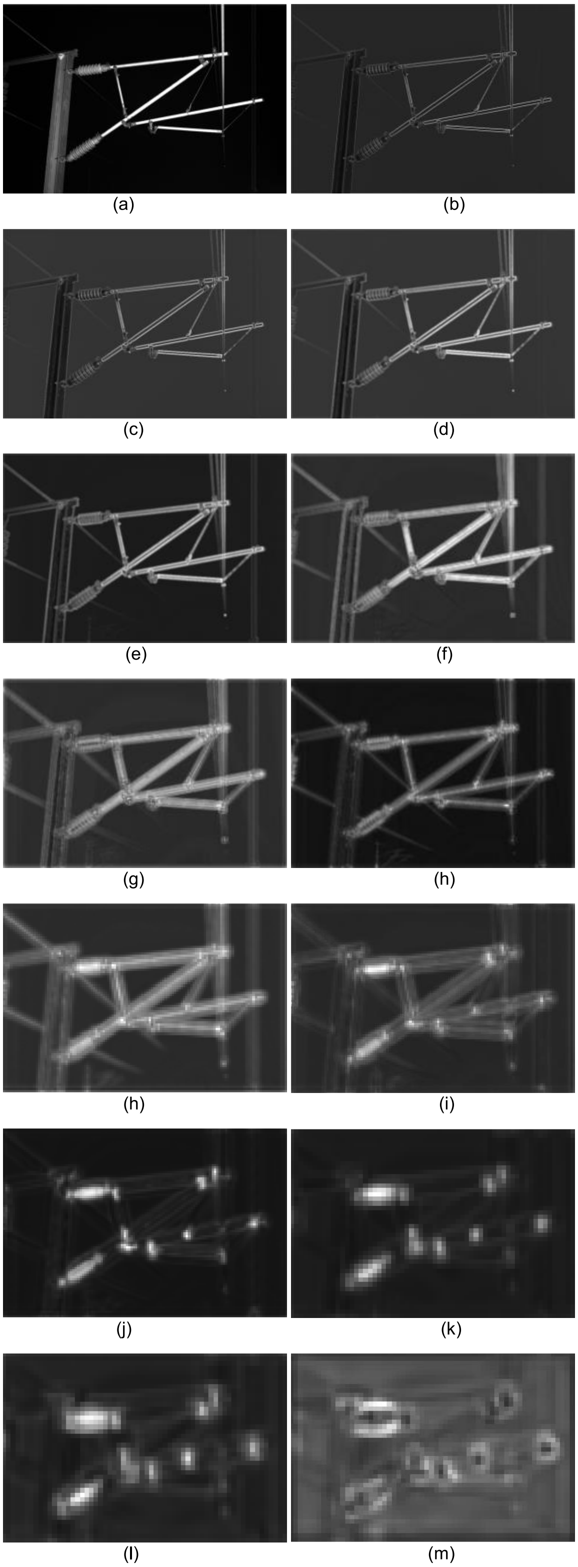


FIGURE 4. Feature maps of each layer of the VGG16 FEN. (a) is the original image; (b) is the second and bottom layer's; (m) is the last and top layer's; (c-l) are the middle layers.

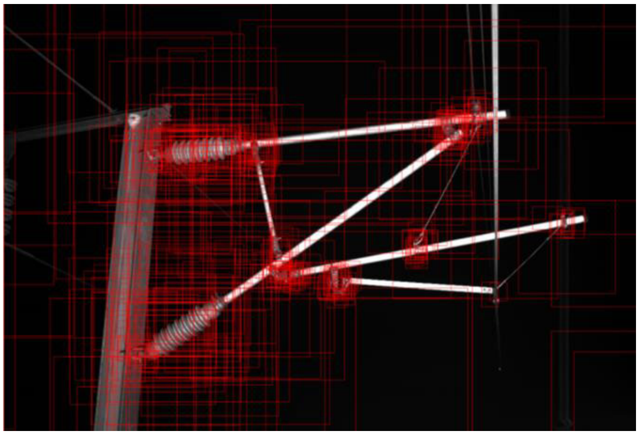


FIGURE 5. Predicted RoIs from the proposed RPN.

three scales $\{64^2, 128^2, 256^2\}$ and three aspect ratios $\{1{:}1, 1{:}2, 2{:}1\}$ of the feature map of Conv 11. When generating the proposals, a label is assigned to each anchor according to the IoU with the ground-truth. An anchor is labeled positive if it has the highest IoU with a ground-truth box or an IoU that is higher than 0.7 with any ground-truth box. An anchor is labeled negative if it has an IoU that is lower than 0.3. The remaining anchors are ignored.

With these definitions, an objective function that follows the multi-mask loss L_{RPN} is minimized. The proposed loss function is expressed as follows:

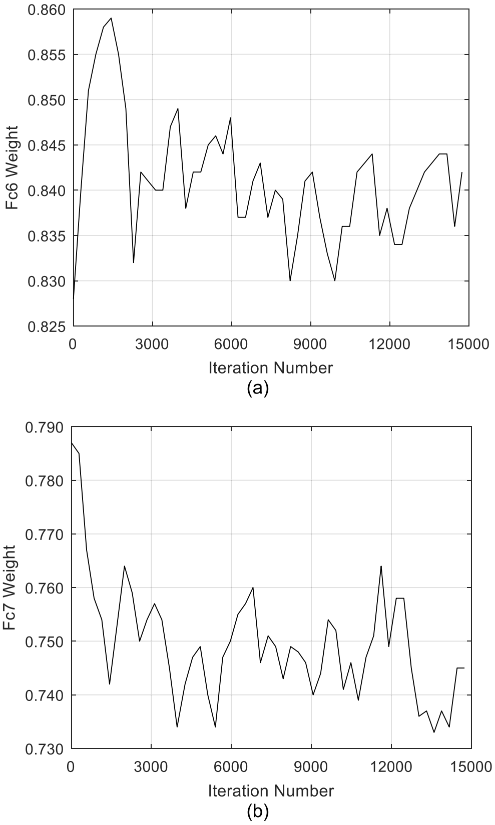
$$L_{\text{RPN}}(p_i, t_i) = \frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \tag{1}$$

where L_{cls} and L_{reg}represent the losses of classification and regression. L_{cls} is the log loss function over two classes (object vs. no object), L_{reg} denotes the robust loss function (smooth L_1) that is defined in [17], p_i is the estimated probability of an anchor being an object, p_i^* denotes the ground-truth label value, t_i is a vector that represents the four parameterized coordinates of the predicted bounding-boxes, and λ is a balancing parameter between the two task losses.

C. CLASSIFICATION AND REGRESSION NETWORK FOR LSCSCS

Fully connected (Fc) networks are typically used to classify the types of objects and to regress the locations of objects.

In Figure 3 (a), the classification and regression network (CRN) of Faster R-CNN from top to bottom involves two serial Fc operations, followed by two sibling Fc operations. By combining the predicted RoIs of RPN with the last feature map of FEN, the features of the predicted RoIs are cropped from the last feature map and are resized to a fixed size. Then, they are sent to the CRN for object detection. However, instead of the last feature map of the VGG16 network, the proposed network architecture chooses Conv 10 as the last feature map, and the resolution of the last feature map doubles. Through an analysis of the sizes of the predicted RoIs in Conv 10, the fixed size is adjusted to 10×10 from 7×7. In addition, through the analysis of the fractions of zero



(a)

(b)

FIGURE 6. Fractions of zero values in two layers of the CRN. (a) corresponds to the Fc 1 layer and (b) to the Fc 2 layer.

values in various feature maps of the CRN in Fig. 6, it is found that the average fractions of zero values are 0.745 and 0.84, respectively. Therefore, the proposed network architecture reduces the size of the CRN by half.

As discussed above, the MNS is introduced into the RPN to reduce the predicted RoIs by merging the overlapping RoIs. However, a special case for the detection of CSCs is that the small-scale brace sleeve screws and the large-scale brace sleeves overlap, as shown in Fig. 7. The brace sleeve screws would be absorbed, which could cause the brace sleeve screws not to be detected by the CRN. Therefore, the CRN can only be used to identify the LSCSCs. The proposed CRN is named LSCCRN.

Each training RoI is labeled with a ground-truth large-scale class $u_L$and a ground-truth bounding-box regression t_L^*. The RoI is labeled $u_L$if the IoU is higher than 0.5 with a class u_L, and the RoI is labeled 0 as background if the IoU is higher than 0.1 and lower than 0.4. The multi-task loss L_{LSCDN}on each labeled RoI is jointly trained with classification and bounding-box regression:

$$L_{\text{LSCDN}}(p_L, t_L^{u_L}) = L_{\text{cls}}(p_L, u_L) + \lambda\,[u_L \geq 1]\,L_{\text{reg}}(t_L^{u_L}, t_L^*) \tag{2}$$

where the Iverson bracket indicator function $[u \geq 1]$ equals 1 when $u \geq 1$ and 0 otherwise.

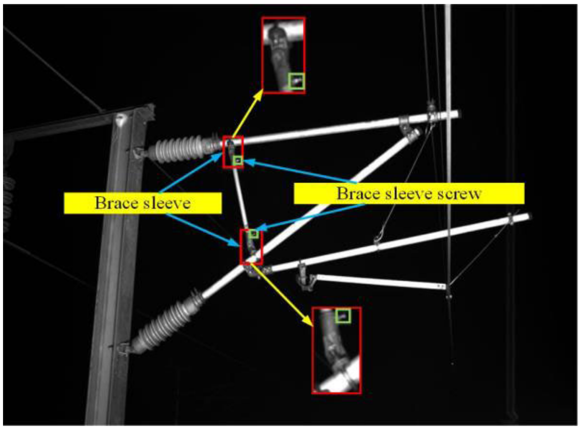


FIGURE 7. Overlap of the predicted boxes of the brace sleeves and the brace sleeve screws.

D. CLASSIFICATION AND REGRESSION NETWORK FOR SSCSCS

The proposed network architecture uses the cascade strategy to overcome the problem that Faster R-CNN is not suitable for the detection of overlapping SSCSCs while avoiding the disadvantages of traditional cascade networks.

In Figure 3 (b), a CRN of the SSCSCs (SSCCRN) is proposed, which is added behind the LSCCRN. The SSCCRN network from top to bottom includes a convolution operation and is followed by two serial Fc operations and two sibling Fc operations for classification and regression. As described above, the low-level feature maps of FEN have a higher spatial resolution but are coarser; hence, they are suitable for small-scale object detection. Thus, through analyzing the low-level feature map, the proposed network architecture chooses Conv 3 as the output feature map for the detection of the SSCSCs. According to the context structure information between the SSCSCs and the detected LSCSCs, the detected locations of the LSCSCs are selected as the RoIs of the SSCSCs. Then, the features of the RoIs of the SSCSCs are cropped from feature map Conv3. Next, a sliding window of size 7×7 slides at a stride of 1 over the cropped feature map. These windows of feature maps are cropped and sent to the SSCCRN for detection.

Each training RoI is labeled with a ground-truth large-scale class u_S and a ground-truth bounding-box regression t_S^*. The RoI is labeled u_S if the IoU is higher than 0.5 with a class u_S, and the RoI is labeled 0 as background if the IoU is lower than 0.1. The multi-task loss L_{SSCDN}on each labeled RoI is jointly trained with classification and bounding-box regression:

$$L_{\text{SSCDN}}(p_S, t_S^{u_S}) = L_{\text{cls}}(p_S, u_S) + \lambda\,[u_S \geq 1]\,L_{\text{reg}}(t_S^{u_S}, t_S^*) \tag{3}$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the performance of the proposed network architecture, comparative experiments are conducted on the catenary data set. All the experiments are conducted on a server with Intel Core i7-8700 CPU @3.70 GHz×12, GeForce GTX 1080Ti GPU, 32-GB RAM, and 2-TB hard disk. All core

TABLE 1. Classes and samples.

Types	Num. of samples	Types	Num. of samples
Insulator base	2218	Windproof wire ring	427
Insulator	2066	Messenger wire base	488
Brace sleeve	2869	Double sleeve connector	560
Brace sleeve screw	2869	Bracing wire hook	481
Rotary double-ear	1394	Isoelectric line	285
Binaural sleeve	1363	Steady arm base	488

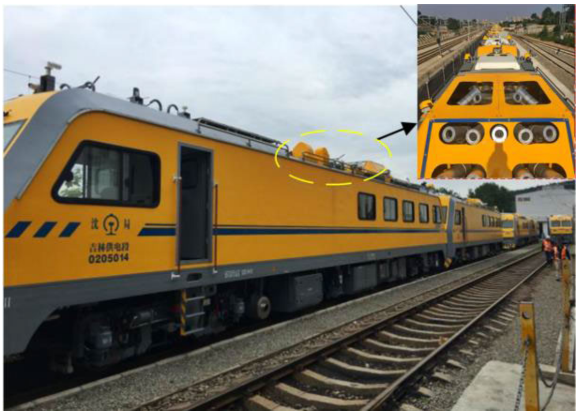


FIGURE 8. Catenary inspection vehicle.

algorithm codes are developed with TensorFlow architecture [21] on Ubuntu 17.10 system.

A. DATA SET

The catenary data set is acquired from the high-speed railway catenary inspection vehicle shown in Figure 8. The total amount of the image data set is 4644, among which the training data set is 2275, the validation data set is 975, and the test data set is 1394. The sample number of every type on the test data set is counted in Table 1.

B. TRAINING PROCEDURE

The training parameters and scheme of the proposed network architecture are set as follows.

First, to ensure that the proposed network architecture can be trained and is convergence, the proposed FEN parameters are initialized with the parameters of the VGG-16 network pre-trained on the catenary data set. And the parameters of other networks of the proposed network architecture, including RPN, LSCCRN, and SSCCRN, are initialized with random normal functions. And the Momentum algorithm is chosen as the backpropagation gradient descent method, and the term momentum and weight decay are set to 0.9 and 0.0001, respectively. The learning rate is set to 0.0001, and the max iteration is 15,000.

Next, a training scheme is designed for different object detection tasks as follows. First, detection network LSCCRN is trained 10,000 times. Then, detection network SSCCRN is trained 10,000 times. After that, these two detection networks are alternately trained 5,000 times. When the networks are trained 10,000 times, the learning rate is set to 0.00001.

C. EVALUATION METRICS

The evaluation metrics include precision (P), recall (R), F_1 score (F_1), average precision (AP), mean average precision (MAP) and time per frame (T).

$$P = \frac{TP}{TP + FP} \times 100\% \tag{4}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{5}$$

$$F_1 = \frac{2 \times P \cdot R}{P + R} \tag{6}$$

$$AP = \int_0^1 P(R)dR \tag{7}$$

$$MAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{8}$$

P Q is the number of component classes.

D. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performances of the selected parameter values and the proposed network architecture, two sets of experiments are implemented and discussed:

1) PARAMETER SELECTION AND VERIFICATION OF THE PROPOSED NETWORK ARCHITECTURE

a: DETERMINE THE ANCHOR SCALE OF RPN

The anchor scales of the RPN in Faster R-CNN are set to $\{128^2, 256^2, 512^2\}$ for the public data set by default. For the catenary data set, through determining the size range of the LSCSCs, the anchor scales of the proposed RPN are set to $\{64^2, 128^2, 256^2\}$. To evaluate the performance of the selection, a set of comparative experiments for various anchor scales are conducted. The results are presented in Table 2, Figure 9, and Figure 10.

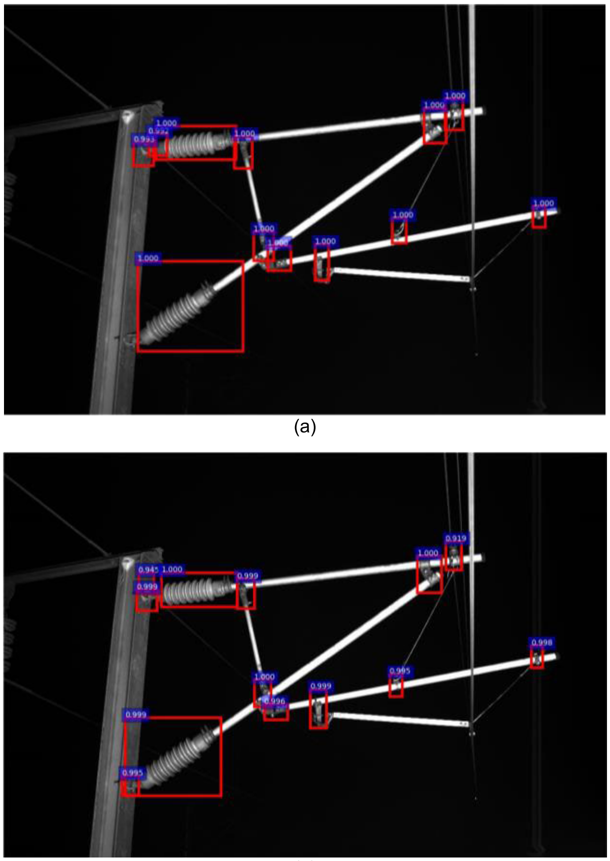
a) In Figure 9, the prediction results with the anchor scales $\{64^2, 128^2, 256^2\}$ more accurately predict the sizes and positions of the CSC bounding boxes, which lay the foundation for the next detection with LSCCRN.

b) According to Table 2, the classification accuracy MAP of LSCCRN with the anchor scales $\{64^2, 128^2, 256^2\}$ reaches 91.2% and is 1.5% higher than that with the anchor scales $\{128^2, 256^2, 512^2\}$. According to Figure 10, the classification loss and the location loss of LSCCRN with the anchor scales $\{64^2, 128^2, 256^2\}$ are lower than those with the anchor scales $\{128^2, 256^2, 512^2\}$.

These results demonstrate that the selected anchor scales are reasonable and can accurately predict the RoIs of LSCSCs.

b: CHOOSE THE FIXED SIZE OF THE FEATURES OF THE ROIS

After the local features of the RoIs of LSCSCs have been cropped, they are input into the LSCCRN for LSCSC detection. Before that, these local features should be resized to


(a)

(b)

FIGURE 9. Prediction results from the proposed RPN. (a) presents the results with anchor scales {128^2, 256^2, 512^2} and (b) the results with anchor scales {64^2, 128^2, 256^2}.

TABLE 2. Detection accuracy with various anchor scales.

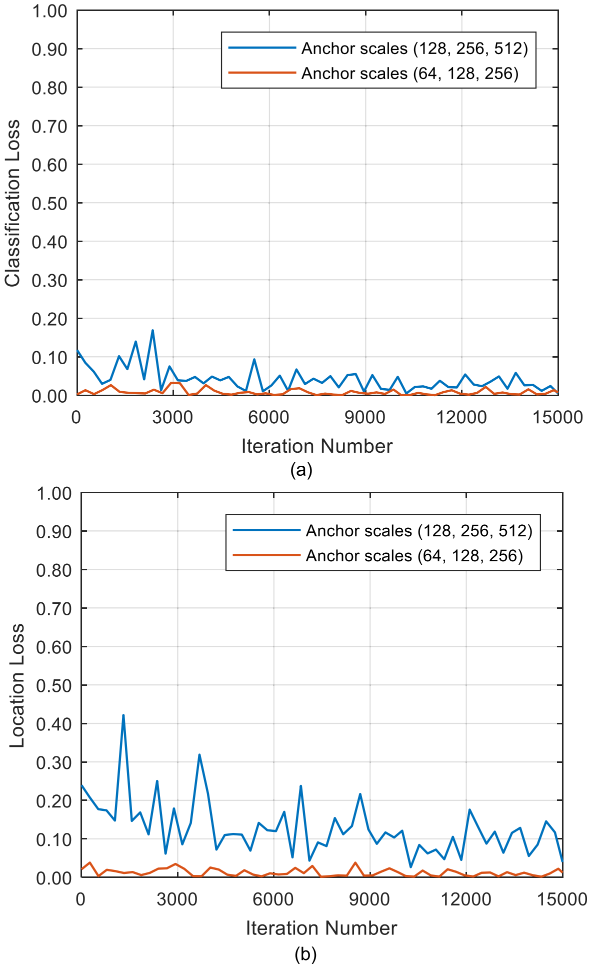
Method	MAP
Proposed RPN with anchor scales {128^2, 256^2, 512^2}	89.7%
Proposed RPN with anchor scales {64^2, 128^2, 256^2}	**91.2%**

a fixed size, which will affect the detection accuracy of LSCCRN. To choose an optimal size, a comparative experiment on eight continuous sizes, namely, {7×7, 8×8, 9×9, 10×10, 11×11, 12×12, 13×13, 14×14}, is implemented. These results are presented in Figure 10.

According to Figure 11, as the resized size increases, the detection accuracy gradually increases. When the size reaches 10×10, the detection accuracy is the highest and an accuracy of 93.1% is realized. After that, the detection accuracy begins to decrease. The figure indicates that 10×10 is the most suitable size for the detection of LSCSCs.

c: DESIGN THE STRUCTURE OF THE PROPOSED SSCCRN

Ideally, the SSCSCs can be detected directly by the proposed SSCCRN at the low-level feature layers of FEN. However, the FEN is trained for the first time to extract the features of LSCSCs, and losses the feature information of SSCSCs occur. Therefore, a convolutional operation is added into SSCCRN to extract additional feature information of SSCSCs. To evaluate the performance of the proposed SSCCRN, the detection results for the brace sleeve screw component with various feature maps are described in Table 3.



(a)

(b)

FIGURE 10. Classification loss and location loss of the LSCCRN network with two anchor scales, (a) presents the classification loss and (b) the location loss.

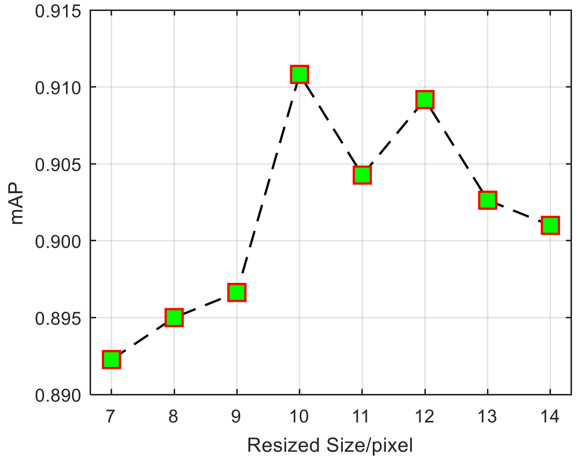


FIGURE 11. MAPs of LSCCRN with various resized sizes.

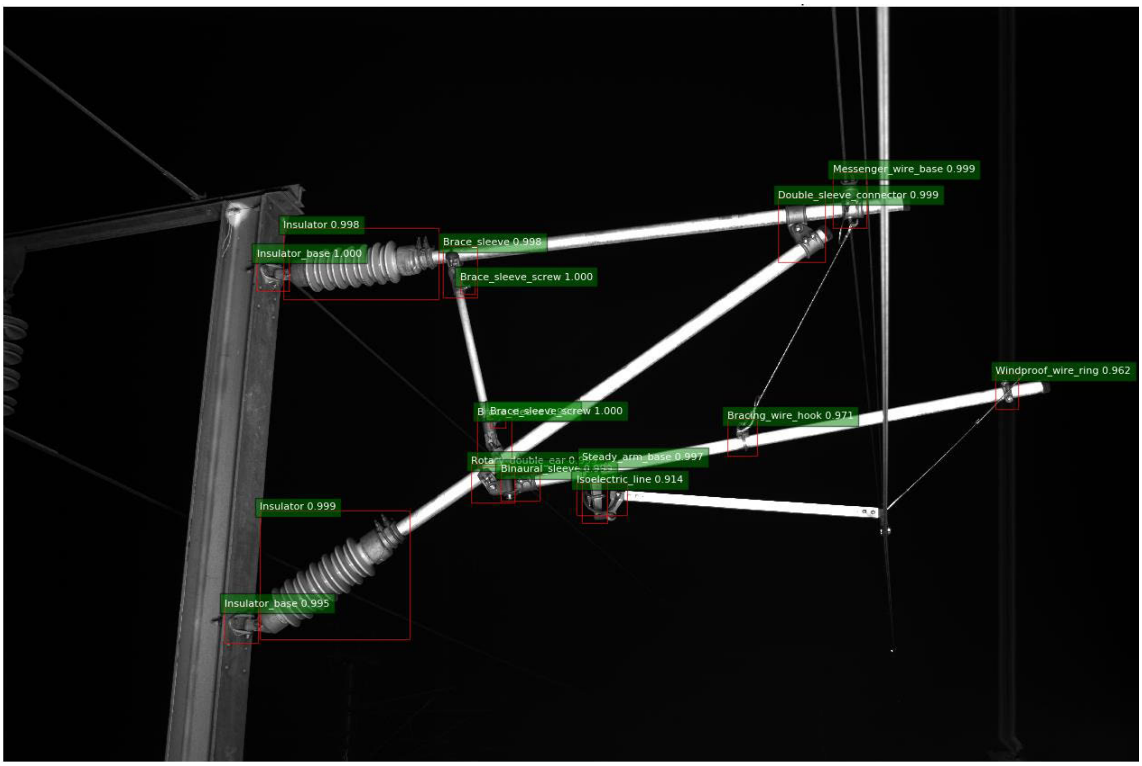


FIGURE 12. Detection example of the proposed CSCD-Architecture for the detection of all CSCs.

TABLE 3. Classification accuracy on the data set.

Method	Precision	Recall	F_1 Score	Time(s)
Conv 1	**91.6%**	**93.6%**	**92.6%**	**0.384**
Conv 2	85.9%	79.0%	82.3%	0.384
Conv 3	89.4%	54.6%	67.8%	0.386
Conv 4	90.1%	42.6%	57.8%	0.386
Conv 1+ Conv 13	94.0%	95.0%	94.5%	0.389
Conv 2+ Conv 13	92.3%	92.3%	92.3%	0.389
Conv 3+ Conv 13	**96.5%**	**97.9%**	**97.2%**	**0.391**
Conv 4+ Conv 13	95.8%	73.5%	83.2%	0.391

According to Table 3, feature map Conv 3, in combination with Conv 13, performs the best, and the precision, recall, and F_1 score are 96.5%, 97.9%, and 97.2%, respectively. Hence, these original low-level feature maps of the FEN only retain the basic shallow features of SSCSCs, and the characteristics of SSCSCs are gradually weakened as the network layers of FEN deepen. However, the proposed SSCCRN can preserve the shallow features and extract more in-depth features by adding an extra convolutional operation, which can increase the detection accuracy of SSCSCs. Moreover, the time consumption of testing is only increased by 0.007 seconds per frame.

2) COMPARISON WITH OTHER DETECTION ARCHITECTURES

(1) First, to evaluate the overall performance of the proposed network architecture, a comparative experiment with the most advanced CNN architectures is conducted. The MAPs of LSCSCs and SSCSCs are calculated in Table 4, separately. For LSCSCs and SSCSCs, the MAPs of the proposed network architecture yield the most accurate results. Specially, compared with Faster R-CNN, the LSCSC detection accuracy of the proposed LSCCRN based on the optimization and improvement of Faster R-CNN is 3.3% higher than Faster R-CNN, and the detection time consuming is also 0.009 seconds lower. In addition, by integrating the proposed SSCCRN, the SSCSC detection accuracy is significantly improved and reaches 89.4%, and the time consuming only needs extra 0.036 seconds. Although the total time-consumption of the system has increased by 0.027 seconds, the requirements of offline detection are still satisfied. A detection example is presented in Figure 12.

TABLE 4. Detection result MAPs with various architectures.

Method	MAP			Time(s)
	LSCSCs	SSCSCs	All CSCs	
YOLO	64.4%	3.8%	59.3%	0.143
SSD	80.0%	5.0%	73.7%	0.231
R-FCN	84.9%	0.0%	77.9%	0.263
Faster R-CNN	89.8%	0.0%	82.3%	0.364
Proposed	93.1%	89.4%	92.8%	0.391

Notes: The total time consuming of the proposed architecture consists of two parts: the time consuming of LSCCRN is 0.355 seconds and the time consuming of SSCCRN is 0.036 seconds.

TABLE 5. Detection result APs of the CSCs compared with the different methods.

Types / Methods	Insulator	Double sleeve connector	Steady arm base	Messenger wire base	Brace sleeve	Binaural sleeve	Rotary double-ear	Insulator base	Windproof wire ring	Bracing wire hook	Isoelectric line	Brace sleeve screw
Han *et al.* [7]	89.4%	-	-	-	-	-	-	-	-	-	-	-
Zhang *et al.* [8]	94.7%	-	-	-	-	-	-	-	-	-	-	-
Kang *et al.* [9]	99.8%	99.6%	99.4%	99.4%	-	99.6%		-	-	-	-	-
Liu *et al.* [10]	-	-	96.5%	-	-	-	-	-	-	-	-	-
Han *et al.* [11]	-	-	-	-	-	94.1%		-	-	-	-	-
Chen *et al.* [12]	-	92.2%	-	-	92.2%	92.2%		-	-	-	-	89.1%
Zhong *et al.* [13]	-	-	-	-	90.8%	94.8%		-	-	-	-	-
Liu *et al.* [14]	-	-	-	-	-	-	-	-	-	-	-	48.5%
Proposed	98.9%	99.8%	96.5%	99.4%	90.6%	90.2%	90.3%	90.6%	89.7%	88.4%	89.6%	89.4%

(2) Next, to evaluate the performance of the proposed network architecture, the APs of each CSC with various methods are listed in Table 5. According to the table, the proposed network architecture can detect each CSC, and its performance reached the same level as available methods in the literature, although the APs are not the highest. For the binaural sleeves and rotary double-ears, the APs are much lower than those in the current literature. This is because the two linked components are treated as a whole component clevis in the literature, which enhances the objects' characteristics and renders them easier to detect. For the small-scale brace sleeve screws, the proposed network architecture overcomes the problems of the cascade network [12] and realizes higher precision. All the above experimental conclusions prove that the proposed network architecture is feasible and has practical application value.

V. CONCLUSION

In this paper, a unified fast high-precision multi-scale object detection architecture, namely, CSCD-Architecture, for the detection of all CSCs is proposed. First, by optimizing and improving the FEN, RPN, and CRN of Faster R-CNN, a faster network, namely, LSCCRN, is proposed for detecting LSCSCs, which was also proven to improve the overall precision of the LSCSC detection. Furthermore, by introducing a cascade structure, a network SSCCRN is proposed for detecting SSCSCs, which overcomes the disadvantage of Faster R-CNN for the SSCSC detection. LSCCRN and SSCCRN are combined to form a unified network architecture. Moreover, in contrast to the methods from the literature on cascade networks, the proposed SSCCRN can effectively utilize information that is obtained from previous networks (FEN and LSCCRN): First, the proposed cascade network, namely, SSCCRN, can directly use the detection results of LSCSCs as the RoIs of SSCSCs and can utilize the low-level feature maps of the FEN, which can eliminate the need for intermediate storage of the large-scale CSC images in the proposed cascade network and can share the feature maps of the FEN to increase the system efficiency. Second, the samples of SSCSCs are jointly labeled with the LSCSCs to reduce the workload. Last, the SSCSCs are also detected in the original images to retain the original location information.

Although the objective of detecting all CSCs has been realized, an in-depth study on the hyper-parameters of the model should be conducted. These parameters are critical to the further improvement of the proposed network architecture. In addition, accurate object detection, as presented in this paper, is a prerequisite for effective fault detection. Further research will focus on fault detection of various types of CSCs. Methods that combine physical information of the objects and well-reported failure modes, together with data-based methodologies, are to be considered. The detection architecture that is proposed in this paper establishes a solid foundation for the subsequent fault detection of catenary support components.

REFERENCES

[1] W. Liu, Z. Liu, A. Núñez, L. Wang, K. Liu, Y. Lyu, and H. Wang, "Multi-objective performance evaluation of the detection of catenary support components using DCNNs," *IFAC-PapersOnLine*, vol. 51, no. 9, pp. 98–105, 2018.

[2] S. Gao, Z. G. Liu, and L. Yu, "Detection and monitoring system of the pantograph-catenary in high-speed railway (6C)," in *Proc. 7th Int. Conf. Power Electron. Syst. Appl. Smart Mobility, Power Transf. Secur. (PESA)*, Dec. 2017, pp. 1–7.

[3] M. Torabi, S. G. M. Mousavi, and D. Younesian, "A high accuracy imaging and measurement system for wheel diameter inspection of railroad vehicles," *IEEE Trans. Ind. Electron.*, vol. 65, no. 10, pp. 8239–8249, Oct. 2018.

[4] Z. Wang, X. Wu, G. Yu, and M. Li, "Efficient rail area detection using convolutional neural network," *IEEE Access*, vol. 6, pp. 77656–77664, 2018.

[5] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel defect detection with machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1176–1187, Apr. 2018.

[6] Z. Liu, W. Liu, and Z. Han, "A high-precision detection approach for catenary geometry parameters of electrical railway," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1798–1808, Jul. 2017.

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

[8] G. Zhang, Z. Liu, and Y. Han, "Automatic recognition for catenary insulators of high-speed railway based on contourlet transform and Chan-Vese model," *Optik*, vol. 127, no. 1, pp. 215–221, Jan. 2016.

[9] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2679–2690, Aug. 2019.

[10] 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. Inf.*, vol. 14, no. 3, pp. 1067–1077, Mar. 2018.

[11] Y. Han, Z. Liu, Y. Lyu, K. Liu, C. Li, and W. Zhang, "Deep learning-based visual ensemble method for high-speed railway catenary clevis fracture detection," *Neurocomputing*, Apr. 2019, doi: 10.1016/j.neucom.2018.10.107.

[12] J. Chen, Z. Liu, H. Wang, A. Nunez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 257–269, Feb. 2018.

[13] J. Zhong, Z. Liu, Z. Han, Y. Han, and W. Zhang, "A CNN-based defect inspection method for catenary split pins in high-speed railway," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2849–2860, Aug. 2019, doi: 10.1109/tim.2018.2871353.

[14] Z. Liu, J. Zhong, Y. Lyu, K. Liu, Y. Han, L. Wang, and W. Liu, "Location and fault detection of catenary support components based on deep learning," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2018, pp. 1–6.

[15] Y.-D. Zhang, C. Pan, J. Sun, and C. Tang, "Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU," *J. Comput. Sci.*, vol. 28, pp. 1–10, Sep. 2018.

[16] S.-H. Wang, Y.-D. Lv, Y. Sui, S. Liu, S.-J. Wang, and Y.-D. Zhang, "Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling," *J. Med. Syst.*, vol. 42, no. 1, p. 2, 2018.

[17] 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.

[18] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 379–387.

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

[20] 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.

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

WENQIANG LIU (Student Member, IEEE) received the B.S. degree in electronic information engineering from Southwest Jiaotong University, Chengdu, China, in 2013, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His current research interests include image processing, computer vision, deep learning, 3-D modeling, and virtual reality and their applications in fault detection and diagnosis in the electrified railway industry.

ZHIGANG LIU (Senior Member, IEEE) received the Ph.D. degree in power system and automation from Southwest Jiaotong University, Chengdu, China, in 2003. He is currently a Full Professor with the School of Electrical Engineering, Southwest Jiaotong University. His current research interests include the electrical relationship of vehicle-grid in high-speed railway, power quality considering grid-connect of new energies, pantograph-catenary dynamics, fault detection, status assessment, and active control. Dr. Liu was elected as a Fellow of The Institution of Engineering and Technology, in 2017. He is an Associate Editor of the journal IEEE Transactions on Instrumentation and Measurement. He is also on the Editorial Board of the journal IEEE Transactions on Vehicular Technology.

ALFREDO NÚÑEZ (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Universidad de Chile, Santiago, Chile, in 2010. He was a Postdoctoral Researcher with the Delft Center for Systems and Control, Delft, The Netherlands. Since 2013, he has been with the Section of Railway Engineering, Department of Engineering Structures, Delft University of Technology, where he is currently a tenured Assistant Professor in the topic data-based maintenance for railway infrastructure. He was a work-package Leader in the development of new sensor technologies (static-, moving-, and crowd-based sensors) for railway networks in Romania, Turkey, and Slovenia, in the European Project H2020 NeTIRail-INFRA project. He has coauthored a book and more than a 100 international journal articles and international conference papers. His current research interests include the maintenance of railway infrastructures, intelligent conditioning monitoring in railway systems, big data, risk analysis, and optimization. Dr. Núñez is on the Editorial Board of the journal *Applied Soft Computing*, Elsevier. He is also an Associate Editor of the journal IEEE Transactions on Intelligent Transportation Systems.

ZHIWEI HAN (Member, IEEE) received the Ph.D. degree in power system and its automation from Southwest Jiaotong University of China, in 2013. He is currently a Lecturer with the School of Electrical Engineering, Southwest Jiaotong University. His current research interests include modern signal processing, and computer vision and their application in railway and electric power systems.