DIRECTIONAL-AWARE AUTOMATIC DEFECT DETECTION IN HIGH-SPEED RAILWAY CATENARY SYSTEM

Huimin Liu, Yi Chang, Student Member, IEEE, Ziqin Li, Sheng Zhong, and Luxin Yan, Member, IEEE*

National Key Laboratory of Science and Technology on Multi-spectral Information Processing,
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China
{*liuhuimin, yichang, liziqin, zhongsheng, yanluxin*}@*hust.edu.cn*

ABSTRACT

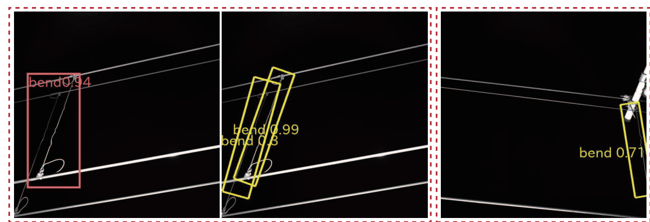
It is crucial to detect the defect objects that could be a hidden danger in high-speed catenary system. Instead of detecting objects based on horizontal rectangle, we propose an end-to-end trainable directional-aware defect detection network (D3-Net) which can automatic select the defective components. D3-Net is composed of two streams, including a directional-aware locator that regresses an inclined rectangle and a channel-wise classifier to diagnose the defects of object. Specifically, we make full use of the directionality characteristic of the man-made objects in the localization stream, which can predict the inclined rectangle with an angle parameter. Besides, we introduce a channel attention module (CAM) in the classification stream to obtain discriminative features for better distinguishing the normal and defective objects. Experimental results on two datasets, Dropper and Insulator, demonstrate that our proposed model outperforms the traditional horizontal detection methods.

Index Terms— directional-aware, attention, defect detection, dropper, catenary system

1. INTRODUCTION

With the increase of a large number of new railway lines, the high-speed railway is coming into the period of maintenance from large-scale construction. In this paper, we mainly focus on detecting the defect objects that possess intrinsic directionality, such as the dropper installed between the messenger wire and the contact wire, as shown in Fig. 1. Because the dropper operates in the open-air environment and suffers from dynamic excitation, it tends to be loosened and fallen off over the time, which has become a big hidden danger affecting traffic safety [1].

In order to ensure safe operation of the traction power supply, it is essential to monitor the status of the dropper. Currently, non-contact detection and monitoring system for the



(a) (b)

Fig. 1. Comparison of the detection results based on horizontal and inclined rectangle. The horizontal predicted boxes are missing when the droppers are densely arranged (a) and partially shielded (b).

catenary state based on computer vision technologies have been widely used in high-speed railway [2, 3, 4, 5, 6, 7]. However, there is little research on the automatic defect detection of the dropper and suspension devices. Moreover, previous defect detection methods basically used hand-crafted features, such as maximally stable extremal region technique [3], part-based model with a tailored circular descriptor [6] and histogram of oriented [8], which are less robust for disturb, shield and illumination.

In recent years, convolutional neural network (CNN) methods have achieved excellent results in many computer vision applications, such as recognition and object detection [9, 10, 11]. In the field of catenary system, several defect detection methods based on CNN have also been proposed [12, 13, 14, 15, 16, 17]. Although these approaches replaced the hand-crafted features with more powerful features extracted by CNN, they trained models in multi-stage pipelines that are slow and resource consumption. Liu *et al.* [16] proposed Faster-RCNN for the localization of isoelectric line, then the image segmentation was carried out based on the Markov random filed model and followed with the fault diagnosis model. Chen *et al.* [13] put forward a three-stage cascaded framework to detect fasteners on the catenary support device, including two detectors to localize key components and a classifier to diagnose their status. These methods above have some limitations, on the one hand, *the training*

This work was supported in part by the Projects of the National Natural Science Foundation of China under Grant No.61571207 and No.61433007, and in part by Hubei Provincial Natural Science Foundation of China under Grant No.2018CFA089.

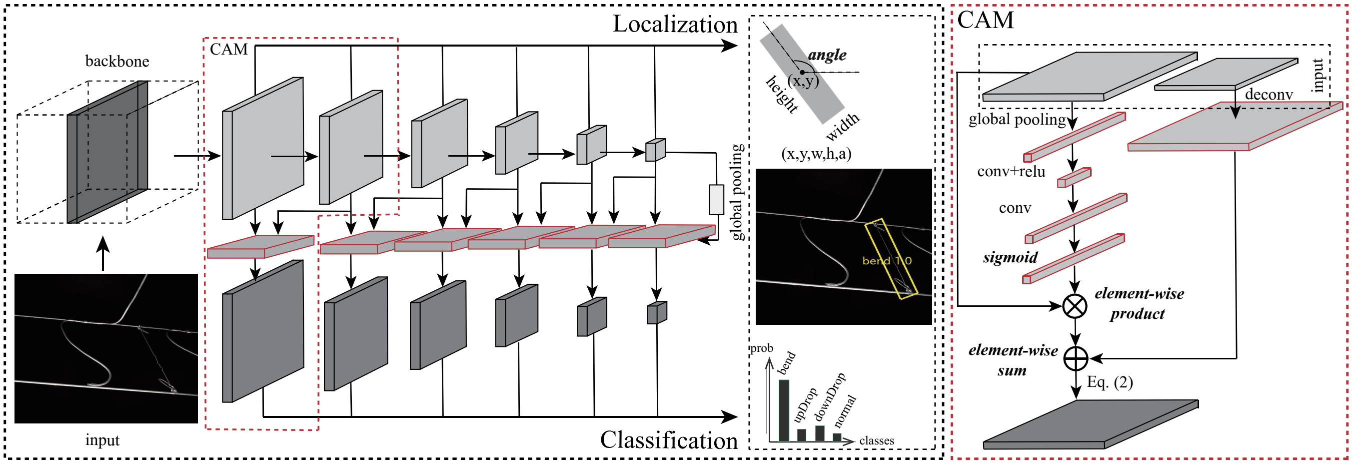


Fig. 2. The overall framework of D3-Net. It is composed of two streams for classification and localization respectively. In the localization stream, it regresses the directional-aware rectangle with an angle parameter. In the classification stream, we introduce channel attention mechanism for obtaining more discriminative feature to better handle the little difference of classes.

is multi-stage, which greatly expends computer memory and holds time. On the other hand, *the inherent directional qualities of man-made objects have not been taken into account*, which results in the unexpected missing when the objects are densely distributed or shielded, as shown in Fig. 1.

To overcome these limitations, we propose an end-to-end single-stage directional-aware defect detection network (D3-Net) in high-speed catenary system, which is composed of two streams for predicting object bounding boxes and classes scores respectively. Specifically, we take full use of the inherent directionality of the line pattern dropper and additionally introduce the prior of angle in the localization stream. Compared with the traditional horizontal detection methods, the inclined rectangle is more tighter and can greatly reduce the redundancy region in predicted box. We also design a channel attention module (CAM) in the classification stream to capture the subtle feature discrepancy between the defective and normal dropper. The main contributions of this paper are summarised as follows:

- We propose a single-stage defect detection network to achieve fault diagnose in catenary system. To the best of our knowledge, this is the first work that detect the defective objects in catenary system in one-stage network, which can greatly save space and time.
- In the localization stream, we additionally take the directional property of the objects into consideration to capture the directional-aware features. We also present the CAM to enhance the feature representation of the fine-grained difference in the classification stream.
- We introduce a dropper defect dataset with inclined rectangle annotation and benchmark it. The code and datasets will be available at the homepage of the authors[1].

[1]http://www.escience.cn/people/changyi/index.html

2. DIRECTIONAL-AWARE DEFECT DETECTION

As shown in Fig. 2, D3-Net is composed of two streams, a fully convolutional stream for directional-aware object bounding box regression and a stream based on channel attention for defect classification. Specifically, the localization stream embeds in auxiliary prior to model the directionality of objects, which regresses the center (x, y), width w, height h and angle a of objects. The channel attention defect classification stream is trained to obtain more discriminative features by embedding in CAM at prediction layers.

2.1. Directional-aware Localization

Different from general objects, the line pattern dropper is long striped shape and exhibits strong directionality, as shown in Fig. 3. We suggest that the conventional SSD [11] is no longer suitable for droppers defect detection. For example, when the droppers are close to each other, the horizontal rectangle tends to regard the two droppers as one object, while the inclined rectangle could tightly surround each dropper with higher confidence, as shown in Fig. 1 (a). Regressing inclined bounding box can greatly reduce the redundancy region of the predicted box and the learned features in the interested region are more discriminative. Therefore, we construct a directional-aware localization stream, which can directly regress the angle of the object. Concretely, we design a new multi-angle anchor mechanism. At each location, the prior box rotates at a series of angles to generate multi-angle predictions. Moreover, we concern that the aspect ratio for the droppers (long and narrow) is highly unbalanced. Consequently, we pre-define one prior width and one prior height for default boxes[2], which decreases the total number of prior

[2]In our work, we empirically set height and width ratio as 5:1.

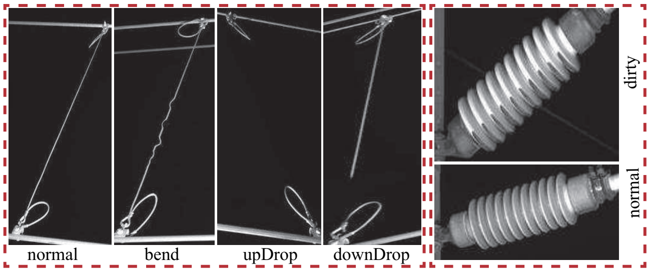


Fig. 3. Categories of the droppers and insulators.

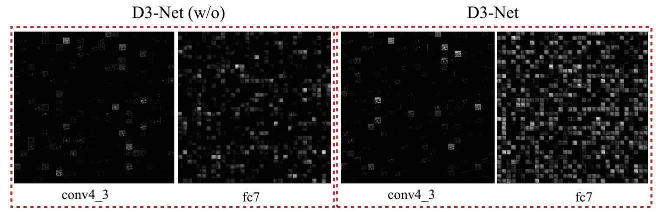


Fig. 4. The effectiveness of CAM. The left and right side show the feature maps ("conv4_3" and "fc7") of D3-Net without CAM, named as D3-Net (w/o), and D3-Net, respectively.

boxes. For regressing angle of the object, we improve the localization loss as:

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{j} \sum_{m} x_{ij} ||\hat{l}_i^m - \hat{g}_j^m||_1, \tag{1}$$

where $x_{ij} \in (1, 0)$ is an indicator for matching the i-th prior box $\hat{l}$ to the j-th ground truth $\hat{g}$. m includes the five parameters (x, y, w, h, a), where (x, y) is the center of the rectangle, w, h, a is the width, height and angle of the inclined rectangle.

2.2. Channel Attention Classification

We consider the fault diagnose as a classification problem and regard each defect as a category. Different from the natural object classification task, the dropper with different defects belongs to the same main category and the discrepancy among them is small. As shown in Fig. 3, the normal and defective droppers have little difference, especially the bend dropper and normal dropper, which makes the classification more difficult. This motivates us to learn more discriminative features so as to capture the subtle discrepancy. In the classification stream, we design a channel attention module (CAM) and embed it across each scale, which could re-calibrate the features with higher activation for better representation, in Fig. 2.

The input of CAM is a feature map $L(x)$ after several convolution operations and then an average global pooling is performed to significantly increase the receptive field. After two convolution layers and one relu activation layer, a sigmoid layer normalizes the output range to $[0, 1]$, which are the weights α of channels. Applied the weights α on original input feature map $L(x)$, we obtain more discriminative feature $A(x)$ by channel-wise multiplication between the weights α and the source feature map $L(x)$. Furthermore, we sum the high-stage feature map $H(x)$ to $L(x)$ for keeping the consistency of different scale. The CAM can be formulated as:

$$F(x) = \alpha L(x) + H(x). \tag{2}$$

To verify the effectiveness of the CAM, we extract the feature maps from "conv4_3" and "fc7", as shown in Fig. 4. Compared the feature maps after channel attention in Fig. 4 (right) with the original one in Fig. 4 (left), we can observe that there are more channels activated. That is to say the CAM enables us to learn more abundant features for better classification. Moreover, from the training perspective, these active channels would benefit for the gradient backward.

2.3. Implementation Detail

D3-Net is an end-to-end trainable, fully convolution neural network whose architecture is inspired by SSD [11]. It uses VGG16 [18] as backbone and implemented on top of the open source code of SSD [11], which is based on the Caffe platform [19]. It is trained with a GTX1080 GPU and an i7-7700K CPU on Ubuntu 16.04 64-bit operating system. For all experiments, the input images are resized to 300×300 for saving compute memory. We train the model for around 20K iterations from the pretrained VGG model using ImageNet datasets. The model is optimized by SGD with a mini-batch size of 16. We adopt a weight decay of 0.0001 with a momentum of 0.9 and set the initial learning rate to 0.001. The learning rate is divided by 10 at 4k, 8k and 12k iterations.

3. EXPERIMENTS

3.1. Experiments Setting

We conduct a series of experiments to verify the effectiveness of D3-Net. We evaluate D3-Net on the Dropper and Insulator datasets respectively, and compare both mAP and AP of every class between the original SSD [11]. It is worth noting that we have not compared our method with others because there is no other method to do the defect detection of dropper. The datasets, Dropper and Insulator, are manually labeled and the ground truth are composed of two parts, including an inclined rectangle marked by (x, y, w, h, a) and a class label shows the defective type. The source images of these two datasets are captured by HD cameras mounted on the top of the inspection vehicle for analysis.

Dropper The droppers are captured from an approximately 2294km line along the whole section of the jing-guang high-speed railway, in which 5519 images and 5750 droppers exist. The size of images are both 5120×3840. These images are composed of four types of droppers. As shown in Fig. 3, "bend" means the dropper presents loose state. When the

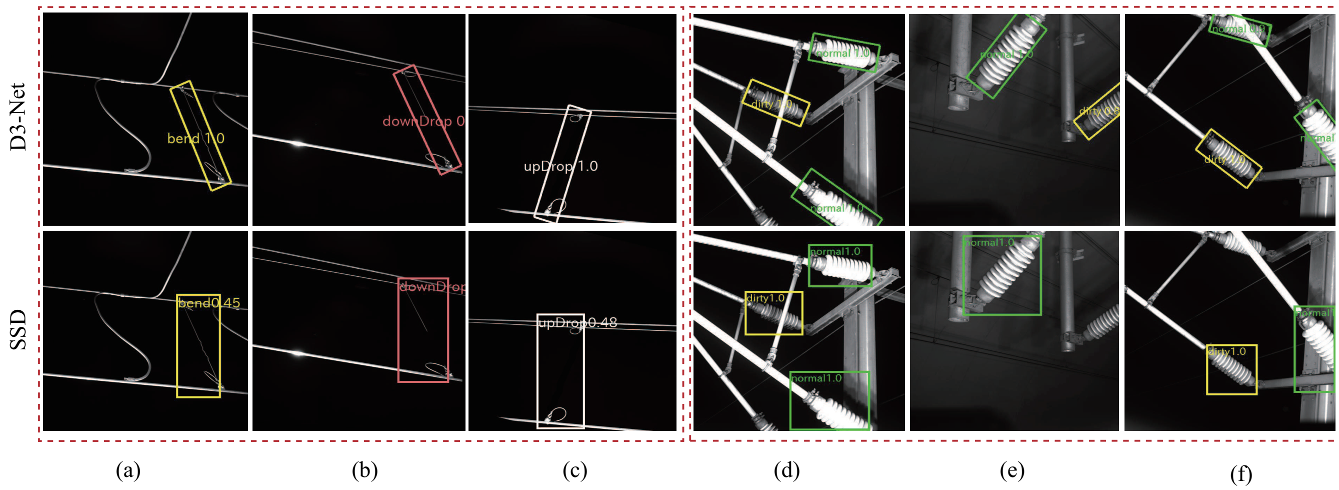


Fig. 5. Defect detection results of D3-Net and original SSD for individual examples from Dropper (a, b, c) and Insulator (d, e, f). The first row is performed by D3-Net and the second row is detected by SSD.

Table 1. Defect detection results of Dropper dataset.

methods	mAP	AP			
		bend	normal	downDrop	upDrop
SSD	89.22	87.87	88.52	90.87	89.64
D3-Net (w/o)	90.48	90.75	89.42	90.88	90.88
D3-Net	**90.83**	**90.84**	**90.66**	**90.90**	**90.91**

dropper falls off, we label it as "upDrop" or "downDrop" that depends on the way it falls off. If the dropper falls off from top, we call it "upDrop" and when it falls off from bottom, we call it "downDrop". Except these defective droppers, the normal droppers are also regarded as a category named as "normal". We uses 4415 images for training and 1104 images for testing. Our proposed method predict more tighter box around object and little redundancy background in Fig. 5. More importantly, the proposed D3-Net could guarantee higher confidence over the original SSD. For the quantitative index in Table 1, we also have an obvious improvement.

Insulator Insulator dataset contains 976 images of insulators and come from the same high-speed line as Dropper, in which 878 images are used as training and 98 images are used as test. The size of images are both 6600×4400. All of them are divided into two types, "dirty" and "normal". Specifically, the "dirty" insulators show the most common defect when the upper surface or bottom surface is smudged due to the long term exposure in wild. The defect detection results of Insulator are shown in Table 2 and Fig. 5. It can also be observed that the predicted box using D3-Net contains less background pixels, so classification between object and background is easier. Especially the (d) and (f) in second row, the confidence is low and predicted box has a large deviation when the object on the edge of the image or part of it is invisible.

Table 2. Defect detection results of Insulator dataset.

methods	mAP	AP	
		dirty	normal
SSD	81.55	88.42	74.69
D3-Net	**83.77**	**86.12**	**81.43**

3.2. Ablation Study

To further verify the effectiveness of CAM, we compare the final model D3-Net with the model D3-Net (w/o) without the CAM at classification stream. The results in Table 2 illustrate that the D3-Net has better performance than D3-Net (w/o) in terms of mAP as well as AP of every class. The extracted feature maps in Fig. 4 (right) show that multiple channels are allowed to be emphasised.

4. CONCLUSION

We propose a directional-aware defect detection network (D3-Net) which is composed of two streams, including a localization stream to regress an inclined rectangle with an angle parameter and a classification stream to classify the defective and normal object. It is worth noting that we are the first to detect the defects of dropper and insulator in catenary system using an end-to-end single-stage neural network. Concretely, we improve the traditional horizontal detection to a directional-aware detection for modeling the directionality of man-made object. Meanwhile, we introduce channel attention mechanism at classification stream to obtain more discriminative features and better cope with the little difference between defective and normal object. The performance of our model outperms the traditional horizontal detection.

5. REFERENCES

[1] N. Zhou and W. Zhang, "Investigation on dynamic performance and parameter optimization design of pantograph and catenary system," *Finite Elements in Analysis and Design*, vol. 47, no. 3, pp. 288–295, 2011.

[2] Aydin I., Karaköse M., and Akin E., "A new contactless fault diagnosis approach for pantograph-catenary system," in *International Symposium MECHATRONIKA*, 2012, pp. 1–6.

[3] A. K. Dubey and Z. A. Jaffery, "Maximally stable extremal region marking-based railway track surface defect sensing," *IEEE Sensors Journal*, vol. 16, no. 24, pp. 9047–9052, Dec 2016.

[4] S. Liao and J. An, "A robust insulator detection algorithm based on local features and spatial orders for aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 963–967, 2015.

[5] Z. Han, Z. Liu, H. Yang, and Y. Han, "Insulator fault detection based on curvelet coefficients morphology and zonal energy method," *Journal of the China Railway Society*, vol. 35, no. 3, pp. 36–40, 2013.

[6] Markus Oberweger, Andreas Wendel, and Horst Bischof, "Visual recognition and fault detection for power line insulators," 2014.

[7] S. Barmada, M. Raugi, M. Tucci, and F. Romano, "Arc detection in pantograph-catenary systems by the use of support vector machines-based classification," *IET Electrical Systems in Transportation*, vol. 4, no. 2, pp. 45–52, 2014.

[8] A. Landi, L. Menconi, and L. Sani, "Hough transform and thermo-vision for monitoring pantograph-catenary system," *Proceedings of the Institution of Mechanical Engineers*, vol. 220, no. 4, pp. 435–447, 2006.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015, pp. 91–99.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. CVPR*, 2016, pp. 779–788.

[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "Ssd: Single shot multibox detector," in *Proc. ECCV*, 2016, pp. 21–37.

[12] Z. Liu, J. Zhong, Yang Lyu, K. Liu, Y. Han, L. Wang, and W. Liu, "Location and fault detection of catenary support components based on deep learning," in *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2018, pp. 1–6.

[13] 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, 2018.

[14] 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.*, 2018.

[15] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Transactions on Systems, Man, and Cybernetics*, 2018.

[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. Industrial Informatics*, vol. 14, no. 3, pp. 1067–1077, 2018.

[17] 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.*, 2018.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *ACM Int. Conf. Multimedia*, 2014, pp. 675–678.