

Virtual Reality and Convolutional Neural Networks for Railway Catenary Support Components Monitoring

Liu, Wenqiang; Liu, Zhigang; Nunez, Alfredo

DOI

[10.1109/ITSC.2019.8917061](https://doi.org/10.1109/ITSC.2019.8917061)

Publication date

2019

Document Version

Final published version

Published in

2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019

Citation (APA)

Liu, W., Liu, Z., & Nunez, A. (2019). Virtual Reality and Convolutional Neural Networks for Railway Catenary Support Components Monitoring. In *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019* (pp. 2183-2188). Article 8917061 Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/ITSC.2019.8917061>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' – Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Virtual Reality and Convolutional Neural Networks for Railway Catenary Support Components Monitoring

Wenqiang Liu, *Student Member, IEEE*, Zhigang Liu, *Senior Member, IEEE*, and Alfredo Núñez, *Senior Member, IEEE*

Abstract— The development of algorithms for detecting failures in railway catenary support components has, among others, one major challenge: data about healthy components are much more abundant than data about defective components. In this paper, virtual reality technology is employed to control the learning environment of convolutional neural networks (CNNs) for the automatic multicamera-based monitoring of catenary support components. First, 3D image data based on drawings and real-life video images are developed. Then, a virtual reality environment for monitoring the catenary support system is created, emulating real-life conditions such as measurement noise and a multicamera train simulation to resemble state-of-the-art monitoring systems. Then, CNNs are used to extract and fuse the features of multicamera images. Experiments are conducted for monitoring the cantilever support connection, both down (CSC-D) and up (CSC-U), and registration arm support connection, both down (RASC-D) and up (RASC-U). Experimental results show that the CNNs trained in the virtual reality environment can capture the most relevant spatial information of the catenary support components. Multicamera image detection based on CNNs detects screw loss for all four components. For CSC-D and RASC-U, normal and pin-loss images are also fully detected. A challenge remains in increasing the pin-loss detection for both CSC-U and RASC-D.

I. INTRODUCTION

The higher demand for railway services and the opening of new railway lines have intensified the required maintenance tasks of the railway infrastructure. In particular, for a reliable power supply, the condition of the whole railway catenary system and its components must be monitored periodically [1, 2]. Short available periods and large numbers of components make monitoring tasks very difficult to perform by traditional methods, such as human inspection. Thus, developing effective detection technology and methods is crucial for guaranteeing both availability and safety in current railway catenary systems.

Currently, monitoring railway components is conducted using different technologies and methods [3, 4]. Among all

sensor technologies available, video imaging has become more popular because of the latest advances in higher resolution cameras and big data collection and processing methods. Most of the research conducted to date using video image detection on railway applications are mainly focused on 2D image detection [5-9]. Research based on 2D images has also been conducted in the field of monitoring catenary support devices. Chen *et al.* [10] presented a three-level cascade deep convolutional neural network to detect and diagnose the condition of fasteners in catenary support devices. Liu *et al.* [11] presented a fault detection method for loose strands of the isoelectric line using an improved faster region-based convolutional network and a Markov random field model. Karakose *et al.* [12] proposed an approach for diagnosing the fault of the interaction between the pantograph and catenary based on an image-based model. Cho and Ko [13] presented a method based on the scale-invariant feature transform for measuring the dynamic stagger of contact wires and evaluated the reliability of railway overhead contact wires. In Liu *et al.* [14], a laser was employed to estimate catenary geometry parameters. The light spot from the laser in the contact wire was registered via 2D images. Using a genetic particle filter and a Kalman filter, the parameters were calculated and corrected. The use of 2D images has proven to be sufficient for a wide variety of problems. However, with 2D images, the spatial information of objects cannot always be recovered. With more detailed spatial information, for instance, using 3D image monitoring (see Fig. 1), it would be possible to address many problems of object detection in real-life environments, such as object occlusion and object orientation.

In the literature, new 3D image-based technology and algorithms have been proposed [15-17]. In the case of monitoring catenary support components, multicamera-based methods are already in practice and can obtain an almost 3D image resolution. However, because 3D data acquisition systems have not yet been widely used, very few 3D open source databases are available for training, such as automatic detection methodologies. In this paper, 3D computer-aided design and virtual reality technology are applied to create 3D model databases and simulate realistic conditions of the catenary support system. Then, a methodology is proposed based on convolutional neural networks (CNNs) using multicameras for defect detection of

Wenqiang Liu is with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China (e-mail: Liuwq_2009@126.com).

Zhigang Liu is with the School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China (e-mail: liuzg_cd@126.com).

Alfredo Núñez is with the Section of Railway Engineering, Department of Engineering Structures, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands (phone: +31 15 2789355; e-mail: A.A.NunezVicencio@tudelft.nl).

railway catenary support components. The simulated data from 3D virtual reality are used to train automatic detection methods that require a large quantity of data. Furthermore, experiments are conducted to evaluate the efficacy of multicamera measurements in the detection of screw loss and pin loss in both cantilever support connections and registration arm support connections.

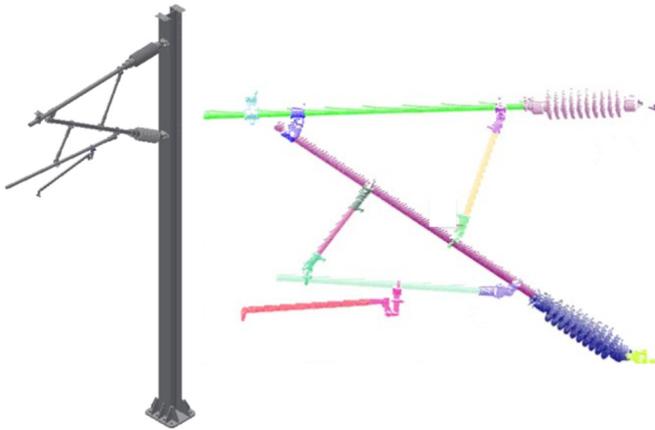


Fig. 1. 3D simulated image of the catenary support components.

The paper is structured as follows. In Section II, the methodology is discussed. Section III describes the experiments and results. Finally, in Section IV, conclusions and further research are discussed.

II. METHODOLOGY

A. Detection System

The first step in the proposed methodology is to build the catenary virtual detection environment, as shown in Fig. 2.

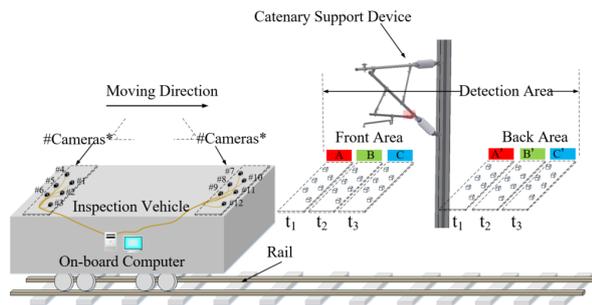


Fig. 2. Catenary virtual detection environment.

Then, the detection scheme mainly consists of three parts: the positioning system, image acquisition system, and data processing system. First, through the positioning system, it is determined whether the train is entering the detection area (see Fig. 2, when cameras reach area A). Then, when the train runs into the detection area at times t_1 , t_2 , and t_3 along the moving direction, the acquisition system captures the images. Twelve cameras installed on the roof of the inspection vehicle are considered, six cameras are used to

collect 3D images in the front area, and six cameras are used to obtain 3D images in the back area. A total of three sets of data are captured per catenary support, corresponding to (A, A'), (B, B'), and (C, C') drawn in red, green and blue, respectively, in Fig. 2. The 3D data are then used by the image processing system to detect the defects in the catenary support components.

B. Input Data

The 3D data of catenary support components are acquired from the virtual cameras. The components are presegmented and labeled according to the parameters of the catenary construction drawings.

1) 3D Projection

3D data contain the spatial information of objects, which is more useful for their automatic detection. In the literature, most advanced object detection methods are based on 2D planar images, from which many methods can be adapted to detect catenary support component defects. To make use of the advantages of both 3D and 2D imaging, a 3D projection method into 2D planar images is applied in this study. 3D projections mainly include orthographic projections and perspective projections. In this paper, we use the perspective projection, as shown in Fig. 3(a). This type of projection emulates measurement systems based on multiple 2D camera monitoring.

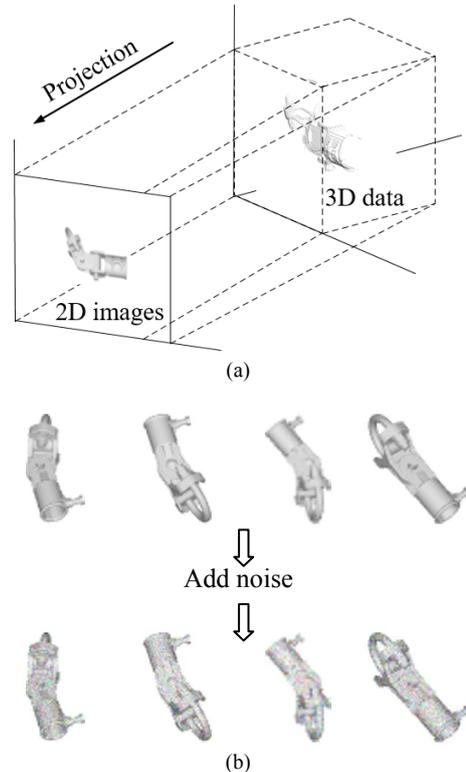


Fig. 3. (a) Projection simulated 3D data into 2D images. (b) Inclusion of noise in the image.

2) Adding Noise

In the field, due to different sources of stochasticities, such as measurement errors and environment, the quality or resolution of the collected data will experience certain impacts. To ensure that the collected data are as consistent as possible with the actual field environment, Gaussian white noise with a variance of 0.005 is added to the data, as shown in Fig. 3(b). The 3D model dataset with added Gaussian noise is referred to as 3D-fModelDs.

C. Image Preprocessing

As shown in Fig. 2, the image data are acquired by twelve different cameras located on the roof of the catenary inspection vehicle. The sizes of the projected images vary. The proposed detection method requires fixed-size input data, so the input data are resampled and resized. In image processing, there are many methods for resampling images. In this paper, bicubic interpolation is employed to resample catenary component images.

D. Detecting Defects

Good features are essential for identifying objects in images. In the literature, there are many excellent feature extraction algorithms [18-21]. However, with the rapid development of artificial intelligence technology, traditional feature methods are being replaced by convolutional neural networks (CNNs). CNNs have been widely used for solving object detection of complex problems.

In this paper, the AlexNet [22] network structure is used to extract image features and classify objects. Based on this network structure, CNNs for defect detection of catenary support components are proposed, as shown in Fig. 4. First, the image data (resampled, size $224 \times 224 \times 3$ per image) from twelve virtual cameras are sent to the feature extraction networks (five convolutional layers), generating twelve different feature maps, `feature_map` (size $6 \times 6 \times 256$ per map) and concatenating them. To make full use of the spatial information of objects, the maximum element value of the same spatial position in the twelve feature maps is fused into a new feature map, `feature_fus`, which is expressed as follows:

$$\text{feature_fus}_{x,y,z} = \sum_{x=1}^6 \sum_{y=1}^6 \sum_{z=1}^{256} \max(\text{feature_map}_{x,y,z}^i) \quad (1)$$

where x , y , z and i are the dimensions of the concatenated feature map. The new fused feature map is sent into the classification network (three fully connected layers). According to the number of catenary support component states K , a $1 \times K$ vector, s , is produced. As shown in Table 1, the total number of catenary support component states is 16. The softmax function is used to establish the object classification function and the loss function. The model is optimized by the Adam algorithm [23]. The core functions are described below.

$$P(Y = j | X = x_i) = \frac{e^{s_j}}{\sum_{k=1}^K e^{s_k}} \quad \text{for } j = 1, \dots, K \quad (2)$$

$$C_i = \text{argmax} \left(P(Y = y_i | X = x_i) \right) \quad (3)$$

$$L_i = -\log P(Y = y_i | X = x_i) \quad (4)$$

where P denotes the softmax function, y_i is the label of sample x_i , $s = f(x_i, W)$ is a learner that can be regarded as a composition of K linear functions, C_i denotes the classification, and L_i denotes the loss function.

TABLE I. TYPES AND STATES OF THE CATENARY SUPPORT COMPONENTS

| Type | CSC-D | CSC-U | RASC-D | RASC-U |
|-------|------------|----------------------|----------------------|----------------------|
| | 1 | 2 | 3 | 4 |
| State | 1 (1.1) | Normal (1.2) | Normal (1.3) | Normal (1.4) |
| | 2 (2.1) | Pin loss (2.2) | Pin loss (2.3) | Pin loss (2.4) |
| | 3 (3.1) | Screw loss (3.2) | Screw loss (3.3) | Screw loss (3.4) |
| | 4 (4.1) | Screw loose (4.2) | Screw loose (4.3) | Screw loose (4.4) |

Note: cantilever support connection-down (CSC-D), cantilever support connection-up (CSC-U), registration arm support connection-down (RASC-D), registration arm support connection-up (RASC-U).

III. EXPERIMENTS AND RESULTS

To evaluate the detection performance of the proposed approach, evaluation indexes are defined. The experiments were conducted in a system with Ubuntu 17.10 platform configured with 32 GB RAM, a CPU clocked at 3.7 GHz \times 12 and two GeForce GTX 1080Ti GPUs with 11 GB memory. TensorFlow [24], a machine learning open source framework, was chosen to implement and validate the proposed approach.

A. Datasets and Parameter Settings

Twelve cameras captured the image of catenary support components. Under normal circumstances, there are only three sets of data to be obtained in a catenary support device area. However, the errors caused by the image acquisition system and the onboard positioning system can inevitably affect the position of the images and then affect the spatial geometric features of input objects. To avoid this, this experiment took five groups of data at five random locations in each area. Finally, fifteen sets of data were generated in each catenary support device area. To train, validate and test the proposed CNN architecture, there were a total of 50,760 data points, among which the training dataset was 33,840, the validation dataset was 8,460, and the test dataset was 8,460. The experimental parameter settings were as follows. The learning rate as 0.0001, and the parameters of the proposed CNNs were initialized with the parameters pretrained by the ImageNet1K [25] dataset.

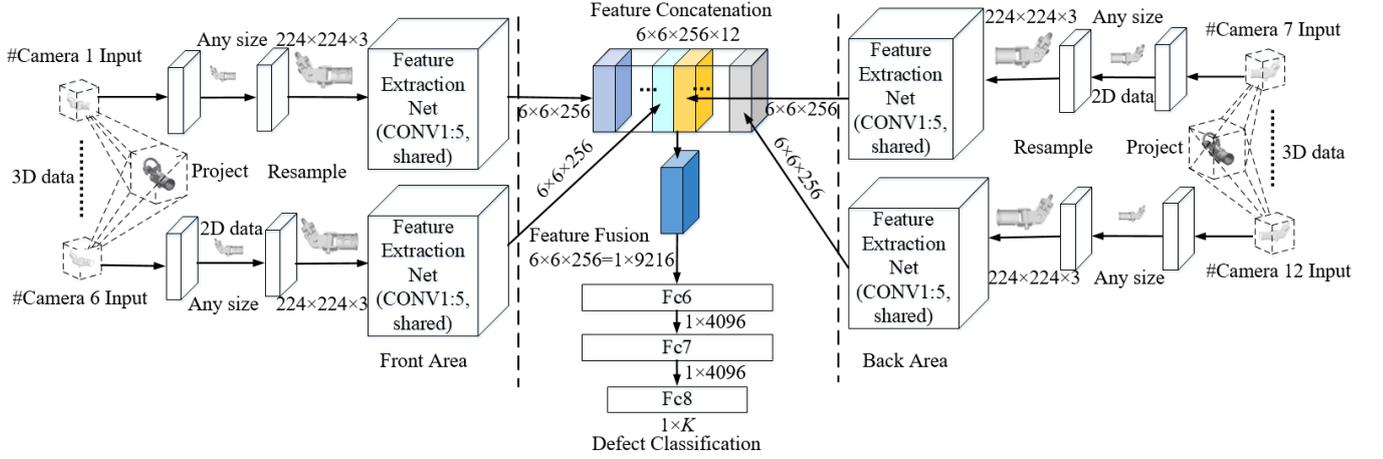


Fig. 4. Proposed CNNs for defect detection of catenary support components based on multicamera fusion.

B. Evaluation Indexes

The evaluation indexes include the precision, recall rate and F_1 score for each component state and the detection time consumption of input data for each group.

$$\text{precision} = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$\text{recall} = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

where TP is true positive, FN is false negative, FP is false positive, and TN is true negative.

C. Experiments and Results

In this paper, the experiment consisted of three key steps, including fine-tuning, validating and detecting.

1) Fine-Tuning

The proposed CNN architecture was initialized with the model parameters pretrained with the ImageNet1K dataset and then fine-tuned 20,000 times with the training dataset of 3D-fModelDs.

2) Validating

In general, in the process of training and fine-tuning a deep learning network model, to verify the accuracy of the model, the validation dataset is evaluated periodically. In this study, the period was set to 100. As shown in Fig. 5, the curves of validation accuracy of the models based on different numbers of cameras were drawn. To accurately determine the trend of the accuracy, the results were smoothed using the double exponential smoothing (DES) method.

3) Detecting

After the proposed CNNs were trained, the test dataset from 3D-fModelDs was used to comprehensively test and

evaluate the performance of the model. The classification accuracy of the model and the time consumed are recounted in Table 2. The precision and recall rate are in the confusion matrixes shown in Fig. 6 and Fig. 7 for detection based on one camera and detection based on twelve cameras, respectively. The diagonal values represent the precision of each object, and the last row of the graph represents the recall rate of each object. To simplify the graph, the matrix indexes of the component states in Table 1 were applied to plot the confusion matrixes, e.g., 1.1 \rightarrow (CSC-D, Normal), 2.1 \rightarrow (CSC-U, Normal).

TABLE 2. DETECTION RESULTS OF THE CATENARY COMPONENT STATES

| Method | Training Configuration | | | F_1 score | Time |
|--------|------------------------|-----------|----|-------------|---------|
| | Pretrain | Fine-tune | C# | | |
| CNN | ImageNet1K | 3D-fModel | 1 | 78.14% | 0.54 ms |
| CNN-12 | ImageNet1K | 3D-fModel | 12 | 87.03% | 6.57 ms |

D. Results Analysis

For fine-tuning, as the number of cameras increased, the accuracy of the proposed CNN model fine-tuned with 3D-fModelDs gradually increased. The more cameras that were used, the more spatial information that was available. However, the optimal number of cameras is part of further research. It could be the case that although more cameras are incorporated to obtain more spatial information, after a certain number of cameras, the possibility of data interference also increases due to noise and other stochasticities.

Next, the performance in the classification of the proposed CNN architecture is analyzed. Comparing Fig. 6(a) and Fig. 7(a), from the confusion matrixes with different numbers of cameras, with the increase in the number of cameras, the subtler changes in space were captured, and the precision of different states within each type of component improved.

From Table 2, the number of cameras affected the classification accuracy. However, the processing time consumed increased with the number of cameras, which will affect the detection speed of the system. A trade-off between time consumed, number of cameras and precision should be further considered. Furthermore, the millisecond time the system requires meets the requirements of the normal running speed of the inspection vehicle.

According to the confusion matrixes, the defect “screw loss” was detected 100% when 12 cameras were used. However, for the defect “pin loss”, more cameras did not substantially improve the detection results because the pin is not symmetrical and on only one side of the cameras. Thus, when using the information from both sides of the cameras, there was no additional useful information for its detection.

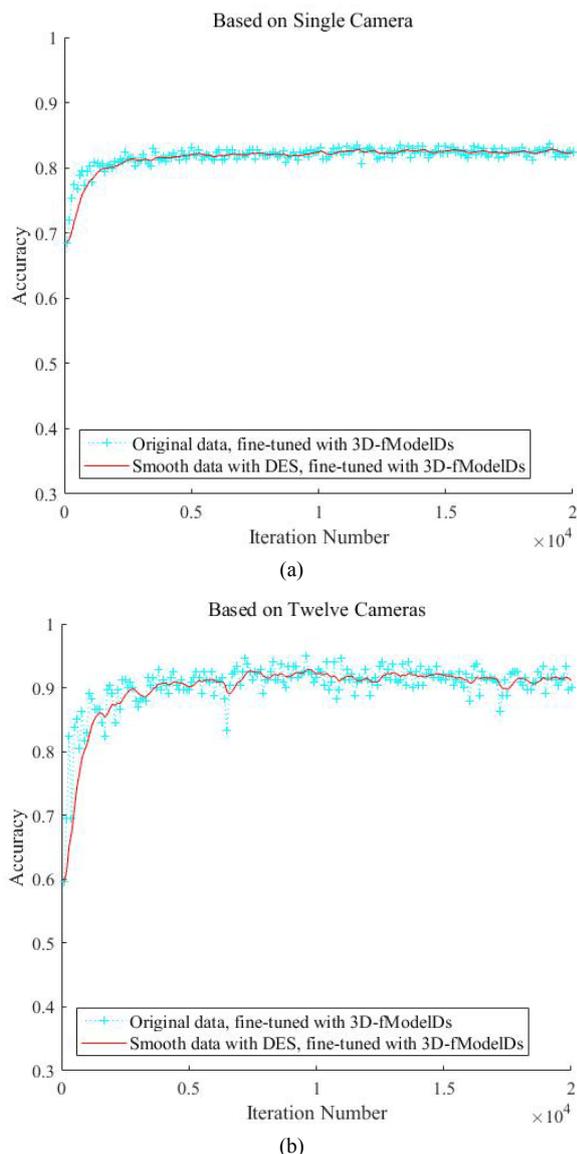


Fig. 5. Validation accuracy based on different numbers of cameras. (a) Single camera (b) Twelve cameras

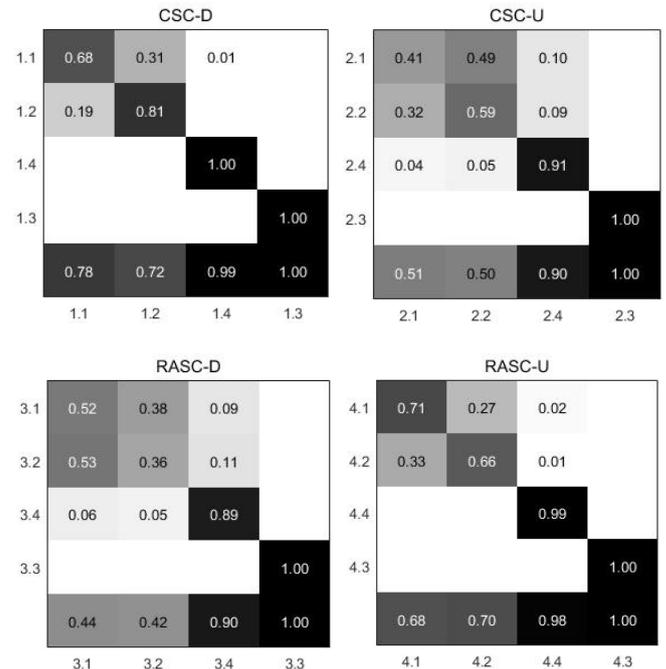


Fig. 6. Confusion matrix based on a single camera. Cantilever support connection-down (CSC-D), cantilever support connection-up (CSC-U), registration arm support connection-down (RASC-D), registration arm support connection-up (RASC-U). State 1: Normal, State 2: Pin loss, State 3: Screw loss, State 4: Screw loss.

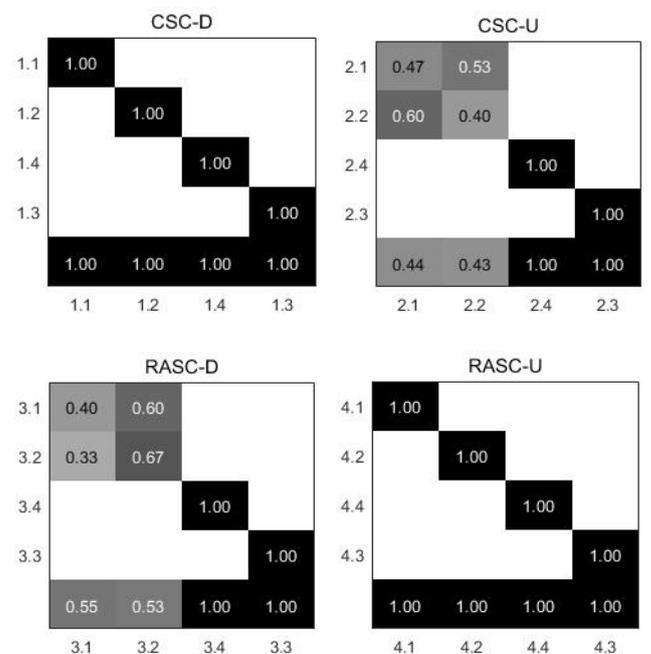


Fig. 7. Confusion matrix based on twelve cameras. Cantilever support connection-down (CSC-D), cantilever support connection-up (CSC-U), registration arm support connection-down (RASC-D), registration arm support connection-up (RASC-U). State 1: Normal, State 2: Pin loss, State 3: Screw loss, State 4: Screw loss.

IV. CONCLUSIONS

With the new developments of image monitoring technology, state-of-the-art methods are being used to solve complex problems with value in different industries. The methodology proposed in this paper is used for defect detection in railway catenary support components using multicamera image monitoring.

The methodology can make full use of the spatial information of catenary components. A virtual reality environment is proposed as a learning environment of methodologies, such as CNNs that requires a large quantity of training data. The CNN architecture proposed in this paper detects screw loss for all the components that were studied. For CSC-D and RASC-U, both normal and pin-loss images are also fully detected using twelve cameras.

To increase the detection of pin loss for both CSC-U and RASC-D, other CNN architectures will be tested. Further research will study other catenary support components and include the analysis of different CNN architectures. Closing the loop between monitoring information and maintenance activities is also a challenge [26-27]. The development of maintenance strategies for large-scale catenary support systems is also a topic for further research.

REFERENCES

- [1] S. Gao, Z. Liu, and L. Yu, "Detection and monitoring system of the pantograph-catenary in high-speed railway (6C)," in *Power Electronics Systems and Applications-Smart Mobility, Power Transfer & Security (PESA), 2017 7th International Conference on*, 2017, pp. 1-7: IEEE.
- [2] Z. Liu, "Detection and estimation research of high-speed railway catenary". *Springer Singapore*, 2017.
- [3] H. Wang, A. Núñez, Z. Liu, Y. Song, F. Duan, and R. Dollevoet, "Analysis of the evolvement of contact wire wear irregularity in railway catenary based on historical data". *Vehicle System Dynamics*, Volume 56, Issue 8, Pages: 1207-1232, 2018.
- [4] Z. Liu, H. Wang, R. Dollevoet, Y. Song, A. Núñez, and J. Zhang, "Ensemble EMD-based automatic extraction of the catenary structure wavelength from the pantograph-catenary contact force". *IEEE Transactions on Instrumentation and Measurement*, Volume 65, Issue 10, Pages: 2272-2283, 2016.
- [5] A. Jamshidi, S. Faghieh-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li, and B. De Schutter, "A big data analysis approach for rail failure risk assessment," *Risk analysis*, Volume 37, Issue 8, Pages: 1495-1507, 2017.
- [6] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Transactions on Intelligent Transportation Systems*, Volume 18, Issue 1, Pages: 153-164, 2017.
- [7] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel defect detection with machine learning," *IEEE Transactions on Intelligent Transportation Systems*, Volume 19, Issue 4, Pages: 1176-1187, 2017.
- [8] M. Torabi, M. Mousavi, and D. Younesian, "A high accuracy imaging and measurement system for wheel diameter inspection of railroad vehicles," *IEEE Transactions on Industrial Electronics*, Volume 65, Issue 10, Pages: 8239-8249, 2018.
- [9] A. Jamshidi, S. Hajizadeh, Z. Su, M. Naeimi, A. Núñez, R. Dollevoet, B. De Schutter and Z. Li, "A decision support approach for condition-based maintenance of rails based on big data analysis". *Transportation Research Part C: Emerging Technologies*, Volume 95, Pages: 185-206, 2018.
- [10] 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 Transactions on Instrumentation and Measurement*, Volume 67, Issue 2, Pages: 257-269, 2018.
- [11] Z. Liu, L. Wang, C. Li, and Z. Han, "A high-precision loose strands diagnosis approach for isoelectric line in high-speed railway," *IEEE Transactions on Industrial Informatics*, Volume 14, Issue 3, Pages: 1067-1077, 2018.
- [12] 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 Transactions on Industrial Informatics*, Volume 13, Issue 2, Pages: 635-643, 2017.
- [13] C. J. Cho and H. Ko, "Video-based dynamic stagger measurement of railway overhead power lines using rotation-invariant feature matching," *IEEE Transactions on Intelligent Transportation Systems*, Volume 16, Issue 3, Pages: 1294-1304, 2015.
- [14] Z. Liu, W. Liu, and Z. Han, "A high-precision detection approach for catenary geometry parameters of electrical railway," *IEEE Transactions on Instrumentation and Measurement*, Volume 66, Issue 7, Pages: 1798-1808, 2017.
- [15] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, Pages: 945-953.
- [16] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Pages: 1912-1920, 2015.
- [17] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, Volume 1, Issue 2, 2017.
- [18] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, Volume 60, Issue 2, Pages: 91-110, 2004.
- [19] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, 2006, Pages: 404-417: Springer.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, Volume 1, pp. 886-893: IEEE.
- [21] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, Pages: 1-8. IEEE.
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, Lake Tahoe, Nevada, 2012.
- [23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [24] S. S. Girija, Tensorflow: Large-scale machine learning on heterogeneous distributed systems. Software available from tensorflow.org, 2016.
- [25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, Pages: 248-255: IEEE.
- [26] A. Núñez, A. Jamshidi, and H. Wang, "Pareto-based maintenance decisions for regional railways with uncertain weld conditions using the Hilbert spectrum of axle box acceleration". *IEEE Transactions on Industrial Informatics*, Volume 15, Issue 3, Pages: 1496-1507, 2019.
- [27] Z. Su, A. Jamshidi, A. Núñez, S. Baldi, and B. De Schutter, "Multi-level condition-based maintenance planning for railway infrastructures – A scenario-based chance-constrained approach". *Transportation Research Part C: Emerging technologies*, Volume 84, Pages: 92-123, 2017.