Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 4.7
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.4.7

Computer Vision-Based Anomaly Detection on Pantograph Carbon Strip

J. Song[1], K. Xue[1], K. C. Fung[1,2], K. H. Lin[1], V. T. Y. Ng[1] and K. M. Lam[1,2]

[1]Centre for Advances in Reliability and Safety, AIR@InnoHK Research Cluster, Hong Kong, China
[2]Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

This paper presents a computer vision-based AI system designed for detecting defects on the surface of carbon strips. Recognizing the limitations arising from inadequate representation of defective classes, we proposed a data augmentation approach that combines generative models and image processing, incorporating a semi-automated image selection process. Additionally, we have adopted a model ensemble technique to enhance identification accuracy. Through experimental validation, we demonstrated the effectiveness of our data augmentation methodology and model ensemble, resulting in an improved defect recall rate of 91.7%, false alarm recall rate of 93.9%, and accuracy of 92.9%.

Keywords: anomaly detection, pantograph-catenary system, generative model, image processing, model ensemble, computer vision.

1 Introduction

The pantograph-catenary system is one of the most important components of the electrical locomotive, which is the only way to effectively provide power to the electrical locomotive [1]. Failure to detect such anomalies can result in damage to the overhead line system and severe service disruptions on railway lines. To pre-empt such incidents, it is imperative to develop and apply an AI-based model for the detection of carbon strip anomalies.

However, the current monitoring systems are incompetent to be effectively and promptly utilized in practical applications. The decision tree utilized in this paper [2] relies on periodic inspection of textual records, which introduces delays and fails to provide timely alarms in the event of an accident. The paper [3] proposed a visual model based on an improved Faster R-CNN algorithm, although it focuses on the approximate localization of carbon strips rather than precise detection of defects. The paper [4] utilizes a laboratory-based scene setup to detect artificial defects, which limits its ability to account for false alarms caused by environmental factors such as rainwater, dirt, oil/grease, and positional variations of moving trains. These false alarms can affect the accuracy of the system, impairing its ability to optimize maintenance efficiency and the maintenance cost of the railway system.

In recent years, the emergence of deep learning technology has witnessed significant advancements in the field of surface anomaly detection. This paper aims to apply state-of-the-art deep learning techniques to enhance the reliability and safety of the intelligent transportation system. The contributions encompass:

1. Utilize generative models to enhance the insufficient representation of defect classes (e.g. chip, crack) on the carbon strips of pantograph.
2. Apply image processing techniques to simulate the lighting conditions typically encountered in railway tunnels.
3. Employ ensembled AI models to improve the defect identification performance.

2 Methods

The dataset contains 266 carbon strip images obtained from three distinct periods. Within this dataset, there are 173 false alarm images and 93 defective images. Our dataset primarily focuses on defective classes, specifically 'chip' and 'crack', rather than including a large number of normal images. Figure-1 provides an illustration showcasing these classes. All the images have a dimension of 2800 in width and 600 in height, with a single channel.

Figure 1: Illustration of defective classes. (Left: vertical crack; right: edge chip.)

2.1 Data Augmentation

Given the limited representation of defect classes and extremely large image size, we employ styleGAN-XL [5] as the generative models to synthesize defect images. To effectively harness the potential of the generative model for data enhancement, a

two-step semi-automatic selection approach was employed. Firstly, the VGG16 network [6] was utilized to automatically filter out images based on their Euclidean distance with the following steps: (1) The last average pooling layer of VGG16 served as the embedding layer; (2) The centroid embedding of real images was computed; (3) Generated images were selected based on their Euclidean distance from the centroid with a threshold. Subsequently, images exhibiting clear defects without any blurring were manually chosen for annotation. The workflow of automatic selection is shown in Figure-2.

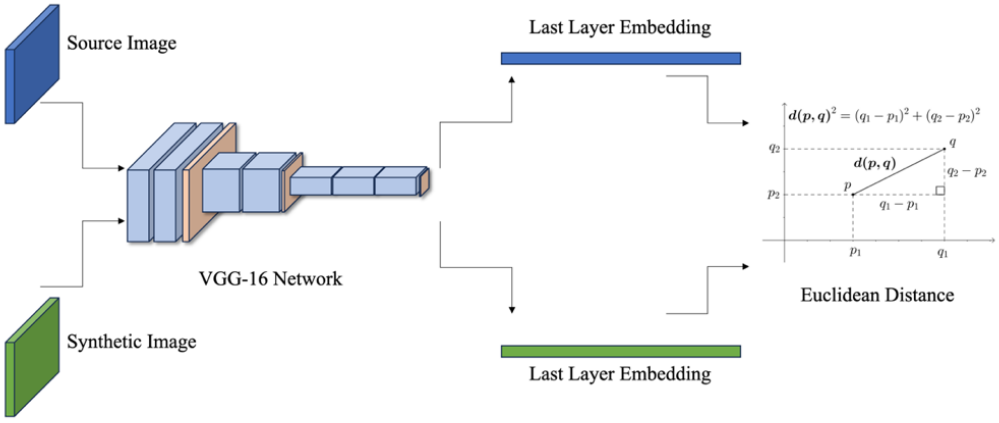

Figure 2: Automatic selection flow for generative images.

Adjusting the brightness, contrast, and hue of an image is a widely adopted image augmentation technique in image processing. In the case of simulating complex lighting conditions found in real railway tunnels, we introduce the concept of adding bright spots to the image [7]. We consider two types of light sources: spot-light and parallel light, and accordingly, we have two workflows: (1) Spot-light flow: randomly generate a value mask using Gaussian blur and apply it to the HSV image on the value channel. (2) Parallel-light flow: employ a line static mode with a random decay rate, initial position, and rotation, and then incorporate the value mask onto the HSV image on the value channel.

2.3 Model Ensemble

We use YOLOv8 model family as the base model for model ensemble [8]. YOLOv8 is the newest state-of-the-art YOLO model that can be used for object detection, image classification, and instance segmentation tasks [9]. We employ a lightweight model as the primary model, complemented by a more powerful and larger model as a supplementary component. The architecture of the ensembled model is shown in Figure-3. We feed the datasets into two models simultaneously and use two tuners to tune the confidential threshold for defect classes. This method will

produce a lot of redundant results. We use the following three methods to improve it: (1) Perform an image-level AND operation to remove false alarms; (2) Perform a non-maximum suppression to remove the duplicated bounding boxes after merge two model's result; and (3) Combine segmentation prototype from supplementary model to get the final foreground segmentation.

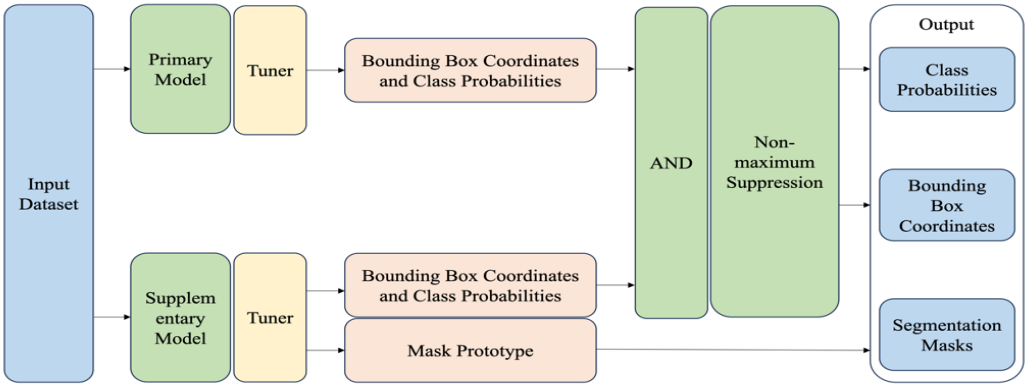

Figure 3: Architecture of the ensembled model.

3 Results

3.1 Image Generation

Initially, we generated 300 synthetic images using StyleGAN-XL. Then, the automatic selection network filtered out 216 images with Euclidean distance larger than 54. Finally, we manually selected 28 synthetic images with 19 chip instances and 10 crack instances for data augmentation. The synthetic results can be seen in Figure-4.

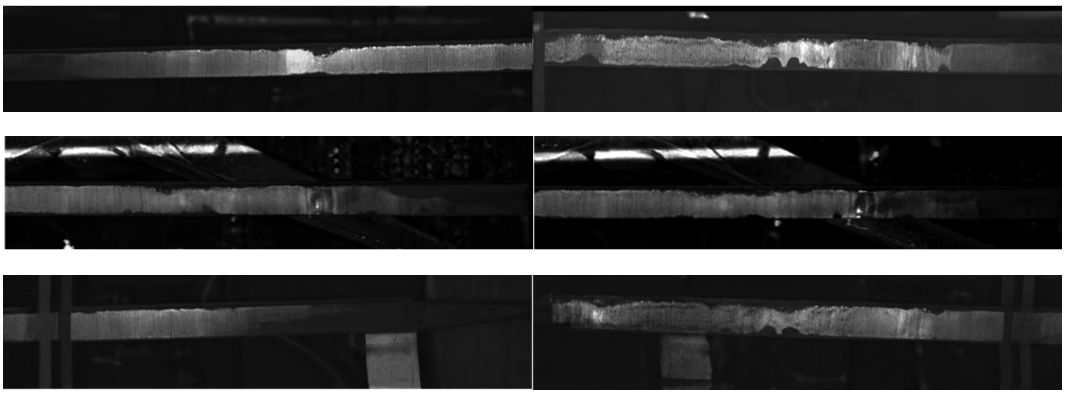

Figure 4: Results of synthetic images (Upper two: selected images; middle: images with large Euclidean distance; lower left: image without clear surface defect; lower right: blur image.)

We utilize OpenCV to facilitate lighting adjustment. Three distinct methods were employed: (1) Spot-light with large size, (2) Spot-light with small size, and (3) Parallel light. Figure-5 showcases samples of the synthesized images resulting from these methods.

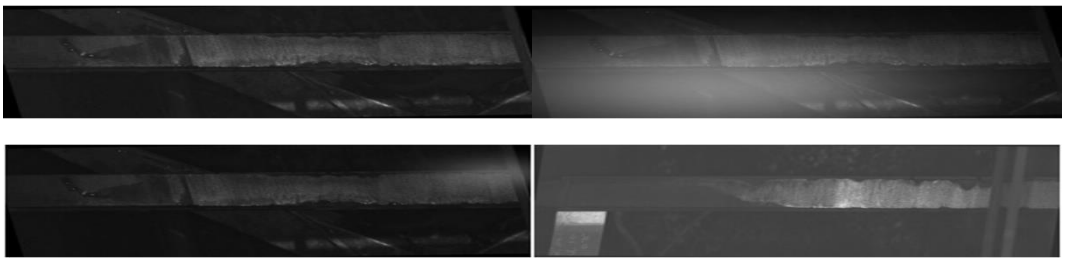

Figure 5: Results of lighting-adjusted images. (Upper left: source image; upper right: spot-light with large size; lower left: spot-light with small size; lower right: parallel light.)

3.2 Result of Data Augmentation Method

For defect identification, our dataset comprises 266 real images and an additional 28 synthetic images. To leverage the advantage of pre-trained models, all the images are stretched into 640×640. The training set undergoes data augmentation, including horizontal and vertical flips, rotation (±10°), and lighting adjustment as an extra augmentation method. After these operations, the training set consists of 1544 images, while the testing set contains 85 images.

Given the practical scenario, visual inspections are conducted by workers on images containing defects, regardless of the specific type or number of defects. To evaluate the performance, we employ both instance-level and image-level Defect Recall Rate (TPR) and False Alarm Recall Rate (TNR) as the metrics, which measure the accuracy and completeness of anomaly detection. Additionally, we utilize Accuracy (ACC) as a metric to assess the correspondence between positive decisions and the actual presence of defects.

$$TPR = \frac{TP}{TP+FP} \tag{1}$$

$$TNR = \frac{TP}{TP+FN} \tag{2}$$

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Where TP refers to true positives, TN refers to true negatives, FP refers to false positives, and FN refers to false negatives.

All the experiments were conducted on a Nvidia A-100 supercomputer, utilizing 100 epochs, a batch size of 16, and the SGD optimizer. The initial learning rate was set to 0.01, with a momentum of 0.937 and weight decay of 0.0005. The Mosaic technique was enabled [10], and we close the Mosaic in the last 10 epochs to improve

the model training [11]. The anomaly detection results are presented in Table-1 and Table-2.

Model	Results without Augmentation		Results with Augmentation	
	mAP50: bbox	**mAP50: mask**	**mAP50: bbox**	**mAP50: mask**
YOLOv8-n	0.726	0.664	0.756	0.711
YOLOv8-s	0.734	0.694	0.759	0.710
YOLOv8-m	0.752	0.684	0.766	0.766
YOLOv8-l	0.716	0.656	0.796	0.711
YOLOv8-x	0.721	0.655	0.746	0.709

Table 1: Instance-level comparison of data augmentation on a single model.

Model	Results without Augmentation			Results with Augmentation		
	TPR	**TNR**	**ACC**	**TPR**	**TNR**	**ACC**
YOLOv8-n	0.833	0.848	0.842	0.861	0.891	**0.878 (+0.036)**
YOLOv8-s	0.972	0.783	0.863	0.944	0.816	0.870 (+0.007)
YOLOv8-m	0.861	0.935	0.904	0.916	0.918	0.917 (+0.013)
YOLOv8-l	0.861	0.870	0.866	0.889	0.891	0.890 (+0.024)
YOLOv8-x	0.833	0.878	0.859	0.944	0.837	0.882 (+0.023)

Table 2: Image-level comparison of data augmentation on a single model.

Table-1 showcases an enhancement in instance-level performance, as indicated by the mAP50-bbox and mAP50-mask metrics. On the other hand, Table-2 displays a significant increase in the accuracy (ACC) of all YOLOv8 variants, attributed to the utilization of data augmentation techniques. Although certain YOLOv8 variants may display slight drops in performance, the majority of the variants exhibit improvements in TPR or TNR metrics.

3.3 Results of Model Ensemble Techniques

To harness the strengths of various model variants, we thoroughly evaluated all 10 possible combinations on both augmented and non-augmented datasets. The experimental results, presented in Table-3, provide a comprehensive analysis of the performance achieved by each combination.

Model 1	Model 2	Results			Increment		
		TPR	**TNR**	**ACC**	**TPR**	**TNR**	**ACC**
YOLOv8-n	--	0.861	0.891	0.878	--	--	--
YOLOv8-n	YOLOv8-s	0.944	0.878	0.906	+0.083	-0.013	+0.028
YOLOv8-n	YOLOv8-m	0.944	0.898	0.918	+0.083	+0.007	+0.040
YOLOv8-n	YOLOv8-l	0.944	0.898	0.918	+0.083	+0.017	+0.040
YOLOv8-n	YOLOv8-x	0.972	0.837	0.894	+0.111	-0.053	+0.016
YOLOv8-s	--	0.944	0.816	0.870	--	--	--

YOLOv8-s	YOLOv8-m	0.944	0.898	0.918	0	+0.082	+0.048
YOLOv8-s	YOLOv8-l	0.917	0.939	0.929	-0.027	+0.123	**+0.059**
YOLOv8-s	YOLOv8-x	0.944	0.898	0.918	0	+0.082	+0.048
YOLOv8-m	--	0.916	0.918	0.917	--	--	--
YOLOv8-m	YOLOv8-l	0.889	0.960	0.929	-0.027	+0.042	+0.012
YOLOv8-m	YOLOv8-x	0.917	0.918	0.918	+0.001	0	+0.001
YOLOv8-l	--	0.889	0.891	0.890	--	--	--
YOLOv8-l	YOLOv8-x	0.972	0.857	0.906	+0.083	-0.034	+0.016

Table 3: Model ensemble experiment results.

It is evident that all the individual models experience improvements in ACC as a result of model ensemble, and most of ensemble model have improvement in false alarm recall rate and defect recall rate. To strike a balance between computational cost and model performance, we carefully select the combination of YOLOv8-n and YOLOv8-m, achieving an impressive 91.8% ACC, along with 94.4% TPR and 89.8% TNR.

4 Conclusions & Contributions

This paper presents a computer vision-based system for detecting surface defects on carbon strips. Considering the limited dataset, an innovative data augmentation approach is introduced, comprising two key aspects: (1) Utilizing generative models for synthesizing images with semi-automatic selection, and (2) Simulating spot-light and parallel light through lighting adjustment. Additionally, model ensemble is employed to enhance identification accuracy. Experimental results demonstrate the effectiveness of these methods, showcasing a maximum 3.6% improvement in ACC resulting from data augmentation and maximum 5.9% increment in ACC from model ensemble, respectively.

Acknowledgements

The work presented in this article is supported by Centre for Advances in Reliability and Safety (CAiRS) admitted under AIR@InnoHK Research Cluster.

References

[1] Y. Yao, Z. Yang, J. Wang, H. Wang, M. Mu, M. Xing, W. Zhang, "Numerical Simulation and Line Measurement Analysis of Pantograph and Catenary Interaction on Overlap Span of High-Speed Railway", in J. Pombo, (Editor), "Proceedings of the Fifth International Conference on Railway Technology: Research, Development and Maintenance", Civil-Comp Press, Edinburgh, UK, Online volume: CCC 1, Paper 4.21, 2022, doi:10.4203/ccc.1.4.21.

[2] Kuźnar, Małgorzata, Augustyn Lorenc, and Grzegorz Kaczor. "Pantograph Sliding Strips Failure—Reliability Assessment and Damage Reduction Method Based on Decision Tree Model." Materials 14.19 (2021): 5743. https://doi.org/10.3390/ma14195743.

[3] Jiang, Siyang, Xiukun Wei, and Ziming Yang. "Defect detection of pantograph slider based on improved Faster R-CNN." 2019 Chinese Control And Decision Conference (CCDC). IEEE, 2019. doi: 10.1109/CCDC.2019.8832401.
[4] Na, Kyung-Min, et al. "Detecting deformation on pantograph contact strip of railway vehicle on image processing and deep learning." Applied Sciences 10.23 (2020): 8509. https://doi.org/10.3390/app10238509.
[5] Sauer, Axel, Katja Schwarz, and Andreas Geiger. "Stylegan-xl: Scaling stylegan to large diverse datasets." ACM SIGGRAPH 2022 conference proceedings. 2022. https://doi.org/10.1145/3528233.3530738.
[6] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). https://doi.org/10.48550/arXiv.1409.1556.
[7] Crispell, Daniel, et al. "Dataset augmentation for pose and lighting invariant face recognition." arXiv preprint arXiv:1704.04326 (2017). https://doi.org/10.48550/arXiv.1704.04326.
[8] Ganaie, Mudasir A., et al. "Ensemble deep learning: A review." Engineering Applications of Artificial Intelligence 115 (2022): 105151. https://www.sciencedirect.com/science/article/abs/pii/S095219762200269X?via%3Dihub.
[9] Zhao, Xu, et al. "Fast Segment Anything." arXiv preprint arXiv:2306.12156 (2023). https://doi.org/10.48550/arXiv.2306.12156.
[10] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020). https://doi.org/10.48550/arXiv.2004.10934
[11] Ge, Zheng, et al. "Yolox: Exceeding yolo series in 2021." arXiv preprint arXiv:2107.08430 (2021). https://doi.org/10.48550/arXiv.2107.08430.