Proceedings of the Seventh International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 15, Paper 11.7
Civil-Comp Press, Edinburgh, United Kingdom, 2026
ISSN: 2753-3239, doi: 10.4203/ccc.15.11.7

Development of SCMaglev Tire Anomaly Detection Model

T. Fujimasu[1], A. Hibino[1], M. Sakuma[1], Y. Himeki[2], G. Katayose[2] and T. Kokubo[2]

[1]Central Japan Railway Company, Japan
[2]ITOCHU Techno-Solutions Corporation, Japan

Abstract

In recent years, image analysis AI, including deep learning–based methods, has rapidly advanced. As a result, numerous cases have been reported in which operational labor savings and efficiency improvements have been achieved across various industries through the application of image analysis AI. At Central Japan Railway Company (JR Central), an in-house AI development environment has been established to detect tire surface flaws on SCMaglev vehicle tires, and efforts have been made to internalize AI-based image analysis technologies. In this study, we investigate the impact of upgrading camera resolution on the detection performance of an existing trained model and examine detection methods for identifying smaller-scale flaws. The results provide insights into how high-resolution imaging can be effectively utilized without retraining the detection model.

Keywords: SCMaglev, rolling stock maintenance, visual inspection, object detection, image recognition, multi-scale.

1 Introduction

In recent years, image analysis AI technologies centered on deep learning have rapidly evolved and enabled labor savings and efficiency improvements in many industries. In the railway sector as well, the application of automated inspection based on image data has been expanding. Supported by advances in high-performance computing environments and imaging technologies, research and development aimed at automating visual inspections have become increasingly active.

However, technical and operational challenges remain regarding how to maintain the performance of trained object detection models when imaging devices are updated or when differences in image resolution arise under new imaging conditions.

The SCMaglev (Superconducting Maglev) developed by JR Central proceeds using rubber tires mounted on bogies at speeds below 150 km/h and switches to magnetically levitated operation at speeds exceeding 150 km/h. The planned maximum operating speed for commercial service is 500 km/h. These rubber tires play a critical role in safely stopping the vehicle, and periodic inspections are conducted to confirm that there are no surface flaws. Currently, tire surface inspections are performed by inspectors through visual examination of each tire, which is time-consuming. Consequently, automating tire surface inspection using AI is a particularly important challenge.

We have been implementing the internal development of AI-based image analysis technologies and have established an integrated AI development process covering data acquisition, annotation, model training, and on-site validation using edge devices. In previous studies, high detection accuracy was achieved for rubber tire surface flaws using YOLO-based object detection models, demonstrating the effectiveness of AI-driven automation for visual inspection.

However, these results assume detection accuracy for flaws of sizes visible at the resolution used during training data creation. It remains unclear whether existing models can be directly applied when imaging devices are upgraded to higher resolution in order to capture smaller-scale flaws. Although high-resolution images generally contain more information and are expected to enable detection of finer defects, object detection models such as YOLOv5 internally resize large input images, potentially causing the loss of features associated with small flaws.

Therefore, from the perspective of operational cost in railway maintenance, it is critically important to determine whether existing models can be reused when imaging devices are updated and whether detection performance can be supplemented without retraining.

Based on this background, the objective of this study is to systematically evaluate the impact of higher camera resolution on existing models and to investigate whether small-scale flaws can be detected using existing models.

2 Methods

2.1 Overall System Configuration

To create the training models, an offline AI development environment, including GPU workstations, was established within a research center. Fixed cameras were installed along the track at the vehicle depot to capture images of tire surfaces, and a NAS

(Network Attached Storage) for storing the captured images, as well as edge devices for performing inference, were deployed within the depot.

This configuration enabled an end-to-end workflow encompassing image acquisition, annotation, model training, and model evaluation. An overview of the system configuration is shown in Figure 1.

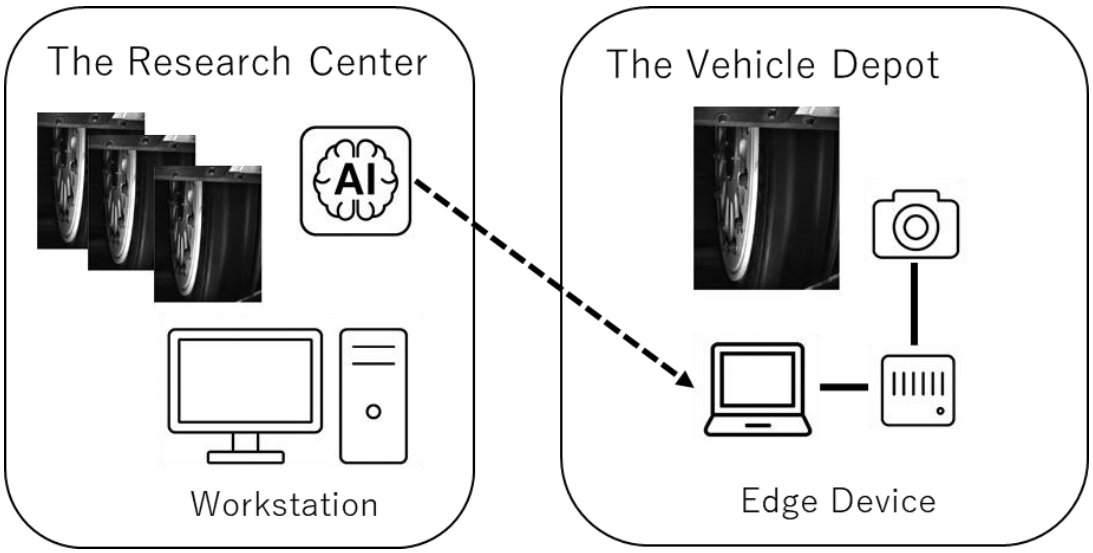

Figure 1: On-site validation using edge device and retraining to improve accuracy

2.2 Imaging Configuration and Camera Specifications

The imaging system used in this study continuously captures the surfaces of rubber tires on SCMaglev vehicles traveling at low speeds using multiple fixed cameras. The system is designed to acquire surface images covering the entire circumference of each tire without omission.

Previously, low-resolution cameras with a resolution of 512 × 480 pixels were used. To improve detection performance for smaller flaws, the imaging devices were upgraded to high-resolution cameras with a resolution of 2448 × 2048 pixels. Examples of tire images captured at conventional and high resolutions are shown in Figure 2. Note that the high-resolution images are displayed after brightness adjustment.

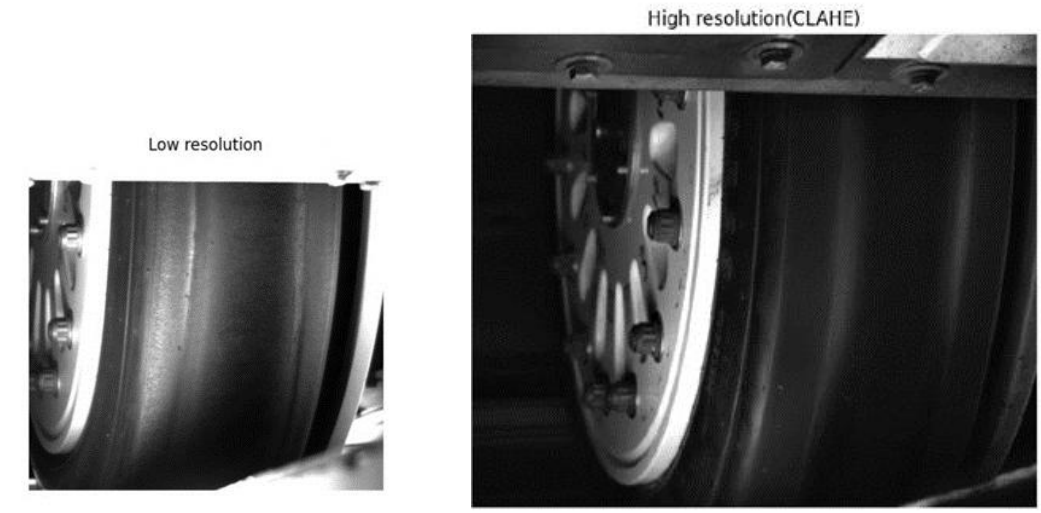

Figure 2: Conventional-Resolution Image (Left) and High-Resolution Image (Right)
Note: The high-resolution image is displayed with increased brightness.

2.3 Dataset Construction and Annotation Policy

The training models developed in previous studies, before the introduction of high-resolution cameras, were created using only conventional low-resolution images. Tire surface flaws visible in low-resolution images were defined as detection targets and annotated accordingly.

To ensure dataset diversity and robustness, data augmentation techniques such as ±10% brightness variation and horizontal flipping were applied, maintaining generalization performance against variations expected in actual operation, including lighting conditions and camera viewing angles.

A total of 1,795 annotated images were prepared, and after applying data augmentation, a dataset of 10,880 images was constructed for training the detection models [1–3].

2.4 Detection Model and Training Settings

The object detection model used in this study was YOLOv5, with a standard input resolution of 640 × 640 pixels. The conventional-resolution training dataset described in the previous section was used for training. Multiple models were trained using different combinations of hyperparameters, including batch size, number of epochs, and learning rate.

This approach allowed us to evaluate performance variability and select the model that demonstrated the highest detection accuracy on conventional-resolution images. The best-performing model was deployed to on-site edge devices, where inference stability and detection accuracy were revalidated.

2.5 Strategy for Applying High-Resolution Images

With the introduction of high-resolution cameras, the number of pixels per image increased significantly to 2448 × 2048 pixels, enabling the capture of smaller flaws that were previously unobservable. However, directly inputting high-resolution images into the existing model causes them to be resized to 640 × 640 pixels, potentially resulting in the loss of fine-grained features.

To address this issue, a preprocessing strategy based on image tiling was adopted. The original high-resolution image was divided into multiple smaller patches, and inference was performed on each patch. Patch sizes of 128, 256, 512, 640, and 1024 pixels were evaluated, and detection performance was compared across different patch sizes.

This approach aims to clarify optimal preprocessing conditions for maximizing the detection performance of existing models on high-resolution images without retraining.

3 Results

3.1 Performance on Conventional-Resolution Images

In previous studies, a training dataset consisting of 10,880 images was constructed from 1,795 low-resolution images (512 × 480 pixels) using data augmentation techniques such as horizontal flipping and brightness variation. The flaws targeted in this dataset were those that could be visually identified by inspectors.

Training and inference using YOLOv5 (input size: 640 × 640 pixels) achieved a detection rate of 100% for the target class with 0% false positives. These results confirmed stable detection performance for flaws of sizes observable with low-resolution cameras [3].

3.2 Application Results for High-Resolution Images

The existing trained model was applied to newly acquired high-resolution images to evaluate its ability to detect tire surface flaws. Four high-resolution images containing flaws classified as “small,” “mid-level,” and “large” were prepared as test data. Examples are shown in Figure 3.

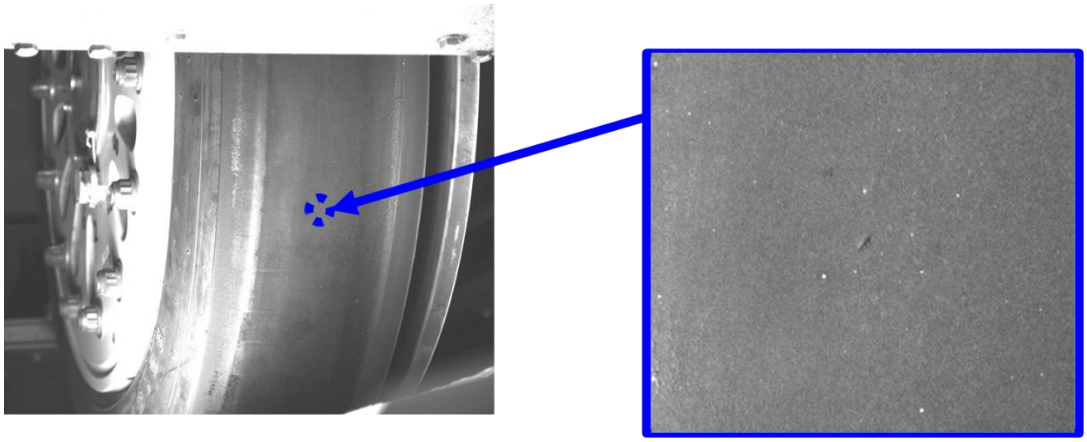

Figure 3: Examples of High-Resolution Test Images Containing Small Flaw Type

First, high-resolution images were directly input into the existing trained model, and inference was performed without any additional preprocessing. The detection results obtained under various confidence thresholds are summarized in Table 1.

ID	Classification of Flaw Size	Confidence Threshold								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	large	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	mid-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3	small	No	No	No	No	No	No	No	No	No
4	small	No	No	No	No	No	No	No	No	No

Table 1: Detection Performance of the Existing Model for High-Resolution Inputs.

When high-resolution images were directly input into the existing model, flaws classified as “large” and “mid-level ” were successfully detected across all confidence thresholds. In contrast, flaws classified as “small” were not detected at any confidence threshold.

Next, for the two images containing undetected “small” flaws, square patches centered on the target flaws were cropped and used for inference with the existing model. The patch sizes were set to 128, 256, 512, 640, and 1024 pixels, and inference was conducted for each size. The detection results for each patch size are shown in Table 2.

ID	Classification of Flaw Size	Patch Size	Confidence Threshold								
			0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3	small	128	No	No	No	No	No	No	No	No	No
		256	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
		512	Yes	Yes	Yes	No	No	No	No	No	No
		640	Yes	Yes	No	No	No	No	No	No	No
		1024	No	No	No	No	No	No	No	No	No
4	small	128	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
		256	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
		512	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
		640	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
		1024	Yes	Yes	Yes	Yes	No	No	No	No	No

Table 2: Re-Detection Results for Small Flaws Using Cropped Images of Various Sizes.

The results showed that small flaws could be detected when appropriate patch sizes were used.

These findings indicate that dividing high-resolution images into smaller regions prior to inference enables the detection of small-scale flaws. Moreover, detection performance depends on the relationship between flaw scale and patch size, suggesting the existence of an optimal patch size that is neither too large nor too small.

4 Discussion

When high-resolution images (2448 × 2048 pixels) capturing tire surface flaws were applied directly to a YOLOv5 model trained on conventional-resolution images (input size: 640 × 640 pixels), relatively large flaws were detected, whereas small flaws were not. This is attributed to the forced downscaling of high-resolution images during preprocessing, which significantly reduces pixel density and causes the features of small flaws to be lost. An illustration of this downscaling process is shown in Figure 4.

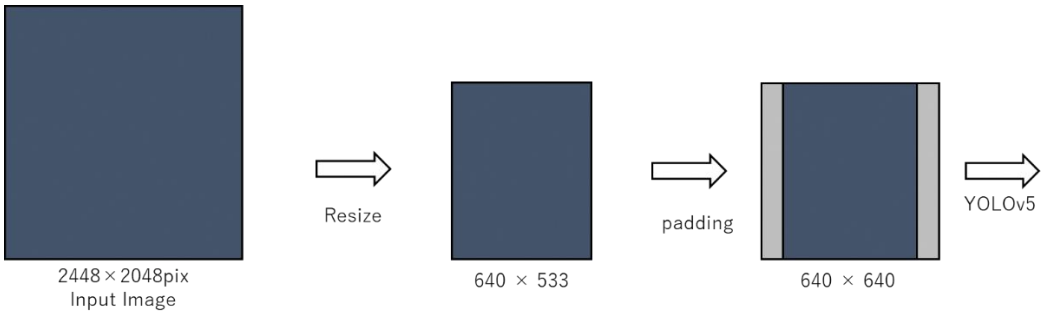

Figure 4: Illustration of How High-Resolution Images Are Downscaled to 640×640 Before YOLOv5 Inference, Causing Small Flaws to Become Undetectable

In contrast, when high-resolution images are cropped into regions smaller than the model input size (e.g., 128 × 128 pixels), the cropped patches are upscaled to 640 × 640 pixels during inference. As a result, flaw features are relatively enlarged, preserving discriminative information and enabling detection even for small flaws. An illustration of this processing flow is shown in Figure 5.

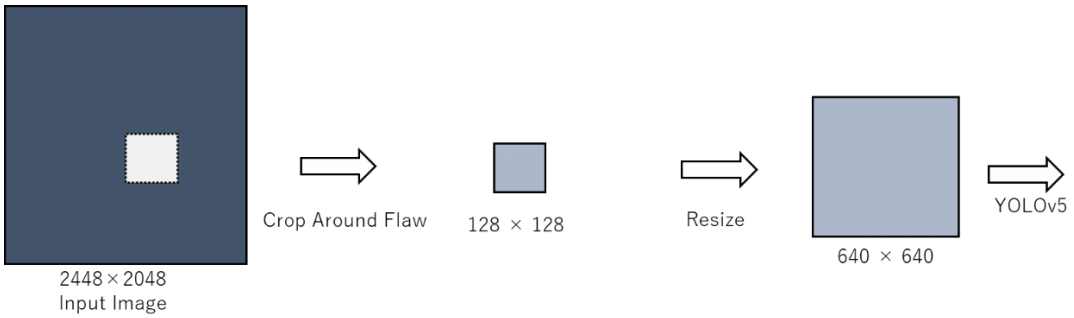

Figure 5: Processing Flow Showing How Small Cropped Patches Are Upscaled to 640×640, Enlarging Flaw Features and Enabling Detection by YOLOv5

These results suggest that an appropriate patch size exists depending on the expected flaw scale. Selecting patch sizes based on the anticipated size of flaws is a key factor in maximizing detection performance.

5 Conclusions and Contributions

This study systematically analyzed the impact of increasing imaging resolution on existing trained detection models. The results showed that directly inputting high-resolution images leads to the loss of small-flaw features during downscaling, resulting in difficulty in detection. However, by combining appropriate image tiling as a preprocessing step, small flaws can be detected even with existing models, without retraining.

The primary contribution of this study is demonstrating the feasibility of adapting existing models to high-resolution imaging systems through preprocessing design rather than model retraining.

In railway maintenance applications, there is a stringent requirement to avoid missed detections. To address this, a multi-scale inference pipeline that performs parallel inference using multiple patch sizes and integrates detection results could further reduce miss rates. Future work is expected to enhance robustness against variations in imaging conditions and contribute to the implementation of highly reliable automated visual inspection systems.

Acknowledgements

The authors would like to express their sincere gratitude to the maintenance personnel who contributed to the continuous accumulation of tire image data, the meticulous creation of annotation data, and the prompt evaluation of inference results. Their active participation in constructive discussions regarding the future of visual inspection greatly contributed to clarifying design guidelines and operational requirements for this study.

References

[1] T. Suzuki, A. Hibino, M. Adachi, “Research on AI image recognition technology for railway vehicle inspection”, Proceedings of STECH 2021 – The 10th International Symposium on Speed-up and Sustainable Technology for Railway and Maglev Systems, The Japan Society of Mechanical Engineers, Chiba, Japan, Nov. 23–25, 2021.

[2] T. Suzuki, A. Hibino, Y. Morishita, M. Adachi, “Research on application of AI diagnostic imaging technology to railway vehicle inspection”, Proceedings of WCRR 2022 – World Congress on Railway Research, Birmingham, United Kingdom.

[3] M. Yoshizawa, A. Hibino, M. Adachi, T. Suzuki, “Development of visual inspection method for railway vehicle underfloor equipment”, Proceedings of TRANSLOG 2022 – The 31st JSME Transportation and Logistics Conference, The Japan Society of Mechanical Engineers, Tokyo, Japan, Nov. 30 – Dec. 2, 2022, Paper TL7-4.