Employing a hybrid model based on texture-biased convolutional neural networks and edge-biased vision transformers for anomaly detection of signal bonds

Takuro Hoshi[a,*] **Seiya Shibayama**[a] **and Xiaonan Jiang**[b]
[a]JR East Information Systems Company, Technical Research Center, AI Technology Group, Tokyo, Japan
[b]East Japan Railway Company, Technology Innovation Headquarters, Tokyo, Japan

Abstract. The railway system of Japan plays a vital role in the national transportation network. A key issue in public transport safety is anomaly detection in railways. Lately, developing robust algorithms and methods for anomaly detection has become the premier task in this field. Recently introduced approaches based on convolutional neural networks, generative adversarial networks, and vision transformers (ViTs) have remarkably improved the research in anomaly detection. Our work proposes a high-performance module for the anomaly detection of signal bonds. First, we present an overview of the proposed module; then, the object detection model and the proposed hybrid classification model based on texture-biased convolutional neural networks and edge-biased ViTs are introduced. Finally, the proposed anomaly detection module is evaluated for accuracy using the dataset from the East Japan Railway Company and the receiver operating characteristic curve. The results show that our proposed module achieves a better performance on real-life data than the two methods that were combined to generate it, raising hope toward possible usage in other areas as well. [DOI: 10.1117/1.JEI.32.2.023039]

Keywords: computer vision; convolutional neural networks; fine tuning; signal bond defects; vision transformers.

Paper 221156G received Oct. 25, 2022; revised manuscript received Apr. 6, 2023; accepted for publication Apr. 7, 2023; published online Apr. 20, 2023.

1 Introduction

Railway systems are the main transportation method in Japan.[1] Traveling via train is more affordable than via airplane for short distances, and it provides a more comfortable experience. The signal bonds, the key features of the railway track circuit, are the equipment for detecting the presence of trains, and they are configured by being welded directly to the rails. Due to the different speeds of deterioration, detecting the broken signal bonds by the human eyes is challenging, and even small scratches can delay the transportation. To avoid this undesired outcome, the signal bonds must be checked and maintained regularly, which increases the railways' maintenance costs. To reduce the maintenance cost, a commonly used method for detecting anomalies consists of overlaying rail images of the same area and detecting the eventual differences between them. Machine learning (ML)[2–4,4,5] techniques have been applied to this end; however, things are still in their incipient phase, which has already delivered some promising results that should be soon improved as ML is an active research area with steady improvements.

In this paper, we focus on the anomaly detection of signal bonds. Deep learning techniques have been employed in anomaly detection in various fields.[6–11] We developed an anomaly detection module that identifies signal bond defects from images captured using the railway monitoring system installed on commercial trains of the East Japan Railway Company (JR-EAST). Because the images captured by the railway monitoring system involve multiwires and odds, which could interfere with anomaly detection, the EfficientDet[12]-based object

*Address all correspondence to Takuro Hoshi, houshi@inet.jeis.co.jp

detection model was employed for extracting signal bond parts. We labeled the normal bonds and abnormal bonds to train the object detection model. Then, classification models based on convolutional neural networks (CNNs) and vision transformers (ViTs)[13] as well as the hybrids between them were created. The performance of these models in the anomaly detection of signal bonds was validated by evaluations of real-life data. In addition to the certification that the proposed hybrid model performs better than its two constituents, we found that the ViT model is biased toward recognizing shapes rather than textures as with CNN.

2 Previous Work

2.1 *Foundation Work*

One of the fundamental issues in ML is anomaly (or outlier) detection, which entails locating instances that significantly deviate from the rest of the data exhibiting "typical" behavior. The past several years have seen extensive research into this issue, leading to the development of numerous algorithms with varied degrees of efficacy. ML applications in anomaly detection may be found in fields such as fraud detection, network intrusion detection, medical diagnosis, product defect identification, and monitoring of healthcare systems. The lack of ground truth labeling in such applications has caused research in this area to get more attention in recent years.

To detect irregularities of insulator breakage in high-speed railway catenaries, Gong et al.[14] suggested a deep learning-based technique mixed with semantic segmentation technology. Although Nugraha et al.[15] suggested using ML to precisely identify railroad irregularities and forecast the state of vital components. Yet, these techniques may greatly benefit from advancement. To find brain tumors, Sivasangari et al.[16] employed ML and image classifiers. Using magnetic resonance imaging, it was possible to distinguish between diseased and healthy cells. Nevertheless, the suggested approach lacked clarity, and the employed algorithm lacked the necessary flexibility. Mohamad et al.[17] used deep learning for detecting and localizing abnormal or extreme events in the sea surface temperature of the Red Sea images. This method provided better performance in terms of sensitivity and specificity.

In electronics, Wachter et al.[18] concentrated on the internal measurements of field-programmable gate array boards exposed to gamma radiation using three different anomaly detection ML techniques to find abnormalities in the sensor readings. Adversarial autoencoder and Hotelling's T-squared distribution were suggested by Goto et al.[19] for use in solder joint anomaly detection on printed circuit boards as well as for inspection purposes. The approaches taken by Wachter et al.[18] are not robust. Nonetheless, the findings from the work of Goto et al.[19] demonstrate that all abnormalities may be found before the boards are completely inoperable. Bakumenko et al.[20] used methods built on ML to reduce sample risk and financial audit inefficiencies. However, the disadvantage of these strategies was that they would only be effective in the contexts of accounting and auditing, where the majority of errors are most likely to occur, and would inefficiently sample data to identify higher-risk journal entries.

In addition, it is worth mentioning fraud detection and prevention. The usage of credit cards is tracked through ML anomaly detection. Thankfully, the technology may also instantly stop a dubious charge. In cybersecurity, ML algorithms enable systems to quickly evaluate various data points and decide whether to grant or refuse access. When there is not an abnormality, the system can automatically provide access; when there is, an alarm is sent to the system administrator. In business, in 2019, Amazon used ML combined with the cloud-native business intelligence (BI) service "Quick Sight" to enable anomaly detection, forecasting, and autonarrative capabilities of the BI products.

To sum up, there is still much to be done, and there is a need to lower the associated expenses even if the current approaches have consistently strived to enhance accuracy in defect detection.

2.2 *Related Work*

Kudo et al.[4] studied the abnormality detection for signal bonds using ML based on image recognition to reduce the number of periodic inspections. Kudo et al.[21] found that signal bond

abnormalities can be detected using image processing technology. They proposed an algorithm for judging whether signal bonds are normal or abnormal by monitoring images. Kasai et al.,[1] in their research work in the field of safety-critical studies tailored to railway engineering, described the development of a track facility monitoring device for all electrified sections in its operating area. Moreover, they proposed the development of a condition-based maintenance (CBM) support tool that utilizes a large volume of monitoring data and optimizes the maintenance work.

This work came to announce an JR-EAST's working plan on technical development to deploy the device to more sections, including the sections not yet electrified. Suzuki,[22] in his chapter entitled "Aiming to Realize a Smart Society and Seamless Mobility with ICT: JR-EAST's Challenge for Business Innovation," presented the research made by the JR-EAST. The aim is to detect abnormalities in a signal bond, which is a part of train detection facilities called track circuits. The methodology of the abnormality detection process comprises (1) pattern recognition by investigating images of signal bonds and (2) abnormality detection by comparing a target image with a previous image. To improve the detection ability, JR-EAST installs many images into the system for learning.

3 Signal Bonds and Overview of Anomaly Detection Modules

Signal bonds are important equipment for allowing railway circuit current flow onto the rails. The bonds are usually welded to the rails, but as the bonds may detach from these or the wiring may break due to deterioration over time, periodic inspections are required. An image of a signal bond taken by the line monitoring device (hereinafter, "wide-angle image") is shown in Fig. 1. After applying an object detection technique, the considered dataset contains over 3000 regular signal bond images and only 51 abnormal signal bond images. We randomly chose 10 abnormal images and 90 regular images as the test dataset, leaving the remaining 41 abnormal signal bond images as the training dataset. The extremely small number of abnormal signal bond images in the training dataset is a big challenge in creating the anomaly detection module.

As illustrated in Fig. 2, we initially extract the signal bonds from the railway images by utilizing the objection detection model EfficientDet because the railway monitoring system also captures the other parts of the railway, which may interfere with the anomaly detection methods, and hence they need to be eliminated before employing the training classification model or conducting the anomaly detection method. Then, the extracted signal bond images are classified by the classification model to distinguish between normal signal bonds and defective ones. The details of the object detection model and classification model are introduced below.

Fig. 1 Wide-angle image.

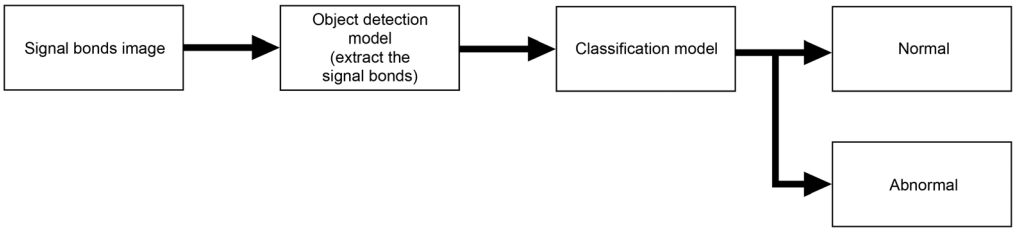


Fig. 2 Overview of the anomaly detection module.

4 Proposed Method

4.1 *Object Detection Model (Extraction of Examined Object)*

In the following, we introduce the details of the object detection model. To resolve the issue of distinguishing useful features from the background or other interference, we implemented an object detection model that detects the examined signal bonds in the images. The object detection model, based on the EfficientDet-$D1$ model, performs transfer learning using the labeled signal bond images. The result was that we were successful in extracting the signal bond parts from the wide-angle image. An example of object detection results is shown in Fig. 3.

4.2 *Classification Model of CNN and ViT*

CNNs include three layers: convolutional, pooling, and a fully linked layer, as demonstrated in their fundamental architecture in Fig. 4. The foundational component of CNN that is primarily responsible for computation is the convolution layer. In a procedure known as feature extraction, it isolates and identifies the distinct aspects of the image for study. There are several pairs of convolutional or pooling layers in the feature extraction network. The biggest feature is the convolution layer, and the second feature is the pooling layer, which is also called subsampling.

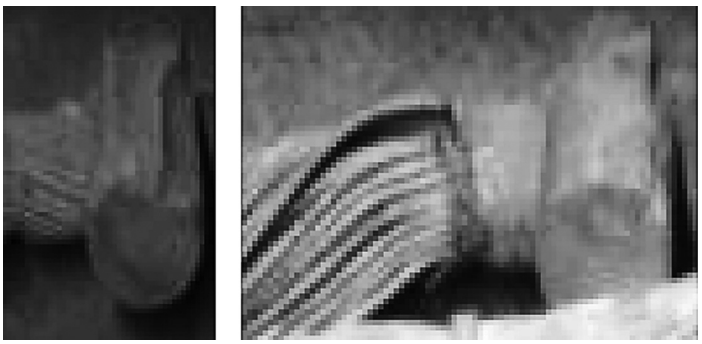


Fig. 3 Object detection results.

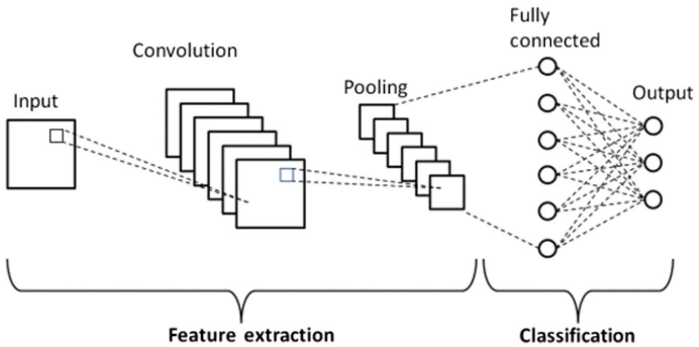


Fig. 4 Basic architecture of CNN[23].

Using the output from the convolution process and the characteristics acquired in earlier stages, a fully connected layer predicts the class of the picture.

Although ViT is an architecture that processes pictures using selfattentional techniques, it is made up of several transformer blocks. Each one has two sublayers: a feed-forward layer and a multihead self-attention layer. The feed-forward layer transforms the output of the self-attention layer nonlinearly, whereas the selfattention layer determines the attention weights for each pixel in the picture based on its connection with all other pixels.

CNNs are able to achieve impressive performances on complex tasks, such as object detection or image classification. In the latter field, CNNs can classify images by learning the representations of objects in the images. Nevertheless, ImageNet[24]-trained CNNs are strongly biased toward recognizing textures rather than shapes, which is different from human behavioral evidence.[25,26] Therefore, we assumed that the CNN-based classification models could be adapted for signal bond classification because the texture in normal and abnormal single bonds is a little different in some cases. Moreover, the abnormal signal bonds include some different shapes, such as the broken wire in the signal bond or the broken wire out of the signal bonds. We assumed that the performance of CNN-based classification models may not be good enough because of the mentioned textures bias of CNNs. Conversely, ViTs achieve remarkable performances in image classification compared with CNN and use fewer computational resources for pretraining. In a ViT-based classification model, the input image is represented as a series of image patches and the method correctly embeds each of them. Then, the positional embedding becomes the input for the transformer encoder. The ViT-based models also learn to encode the relative locations of the image patches to reconstruct the whole image. Because it encodes the relative location of the image patches, we assume that ViT models are more biased toward recognizing shapes rather than textures as with CNNs.

To verify our assumption, we create two models, a ViT-based classification one and a CNN-based classification one, for distinguishing the normal signal bonds and anomalies. We checked the gradient-weighted class activation mapping (Grad-CAM)[27–29] of the two models using the same test dataset to determine on which portion each method was focused. The CNN-based classification model carries out transfer learning based on EfficientNet.[30] Transfer learning is a method of replacing the final output layer and relearning the weights of the neural network for an existing learned model that has been trained in another domain. For the final output layer activation function, a sigmoid function, which could be set to a threshold in which scratches would not be ignored, was adopted. It is commonly believed that, in deep learning, ∼100 images per anomaly are required to achieve high accuracy in anomaly image classification.[31] Because our abnormal signal bonds dataset is not sufficient for training, we applied augmentation including shear and rotation on the dataset. The test data were used to examine our assumptions on the CNN- and ViT-based classification models on unknown data (generalization performance), so they were not included in the training part. Based on the test result of the two models, we found that the CNN model is biased toward textures, whereas the ViT one is toward edges, which validated our assumptions. The details of the result are shown in Sec. 4.1. Given this, our next question was whether we could combine these two models to focus on both textures and edges to lose the bias and achieve a better performance than the single models.

4.3 *Proposed CNN–ViT Hybrid Model*

We introduce the proposed CNN–ViT hybrid model. Its structure is shown in Fig. 5. The input image is passed to each of the predefined CNN and ViT models. Each model produces a flattened output; then it performs batch normalization and integrates the output of two models. The integrated output is processed by the three-layer dense structure, and then, it is passed to the sigmoid activation function to obtain the output.

The transformers lack some of the inductive biases inherent in CNNs, and the performance of the ViT may not be good enough when trained on insufficient datasets. Our training dataset contained only ∼3000 images, so it cannot be regarded as a large-scale one. Therefore, we trained the proposed model in the CNN and ViT models in parallel, expecting that the CNN model would work faster. Should this have happened, the output on the CNN model would have been biased, but its output on the ViT model would be minimized. Therefore, we assumed that performing prelearning on the ViT model is appropriate. The data ratio between the pretraining

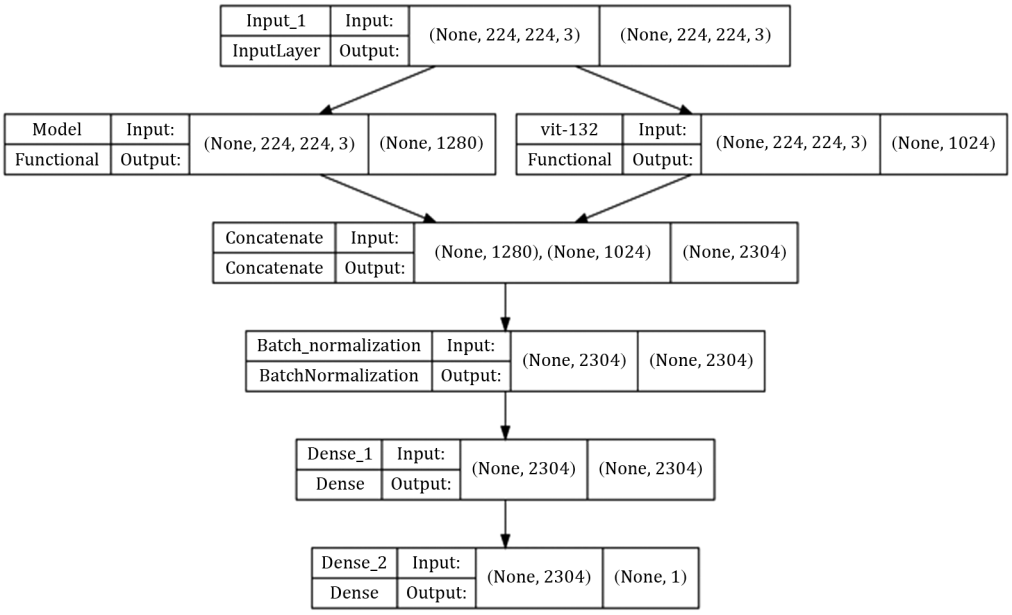


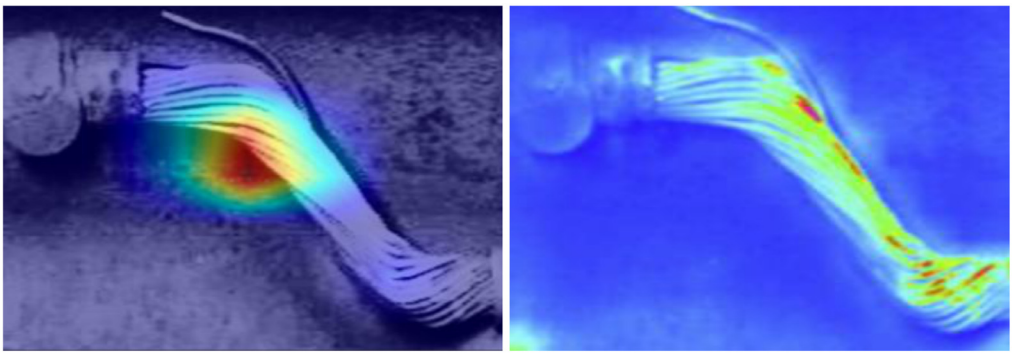
Fig. 5 Structure of the proposed CNN–ViT hybrid model.

dataset and the original training one is a very important parameter for the overall optimization by performing transfer learning. Naturally, if only the ViT part is optimized in advance, a reversal phenomenon occurs: only the ViT model converges and the training of optimizing the CNN part may not proceed. To avoid such a situation, we proceeded step by step. Specifically, both models were prelearned and converged; then, the weight of each model was frozen and incorporated, and finally, the output was combined in the dense layer. The combined output was fully recombined by the dense layer, and only the neural network of that part was optimized by the original training dataset. Because the pretrained model was frozen this time, the performance of each model after training could also be verified with test data. We compared the performance of the three models (CNN, ViT, and hybrid), and the results are discussed in Sec. 5.2.

5 Results

5.1 *Results for the Single Model of CNN and ViT*

We checked the Grad-CAM graph of each test image. One can immediately remark that the CNN model is focused on the abnormal portion of texture, whereas the attention map of ViT focuses on the edge part. An example of a Grad-CAM graph is shown in Fig. 6. Both the CNN and ViT models can achieve a relatively good performance on signal bond anomaly detection.



(a) (b)

Fig. 6 Result of Grad-CAM: (a) CNN and (b) ViT.

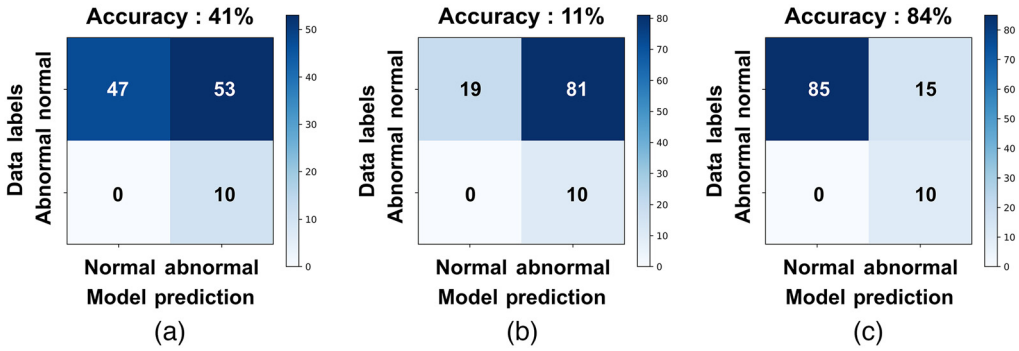


Fig. 7 Comparison of confusion matrixes: (a) CNN, (b) ViT, and (c) CNN–ViT models.

5.2 *Results for the CNN–ViT Hybrid Model*

In this section, we compared the results delivered by the CNN–ViT hybrid model, the CNN model, and the ViT model. The test was performed on the premise that no abnormal images are missed, and the calculation reflects the correct answer rate of the normal images. We calculated the corresponding confusion matrixes to check the accuracy of each model (Fig. 7) and created the receiver operating characteristic (ROC) curve[32] for each model (Fig. 8). We set the limit for classifying the abnormal signal bond correctly. The accuracy of the normal signal bond of the CNN–ViT model showed the best performance, namely, 84%. The Grad-CAM graph of the CNN–ViT hybrid model produces a flattened output and then performs batch normalization and integrates the output of each of the two models. This leads to losing the position information of the target and cannot be checked by the Grad-CAM; hence, we did not rely on this visualization method in our work.

We also checked the distribution of each model because the results of normal and abnormal images are expected to have different distributions, a feature that could provide a method for separating them. Based on the box plots in Fig. 9, we were able to find that the result of the CNN–ViT hybrid model is improved with respect to the ones of the ViT and CNN models. As for the distribution, the one of a normal image converges to 0 considerably, which makes it possible to exclusively determine abnormal images. The most important reason for this is that the number of abnormal images is small, so it is expected that the method can be further improved if fed with a larger number of abnormal images.

6 Discussion

JR-EAST, as part of its response to the growing labor shortage, is planning to promote smart maintenance. This is encapsulated in the shift from time-based maintenance to CBM.[33] In recent years, the image recognition algorithms conventionally used for detecting abnormalities in

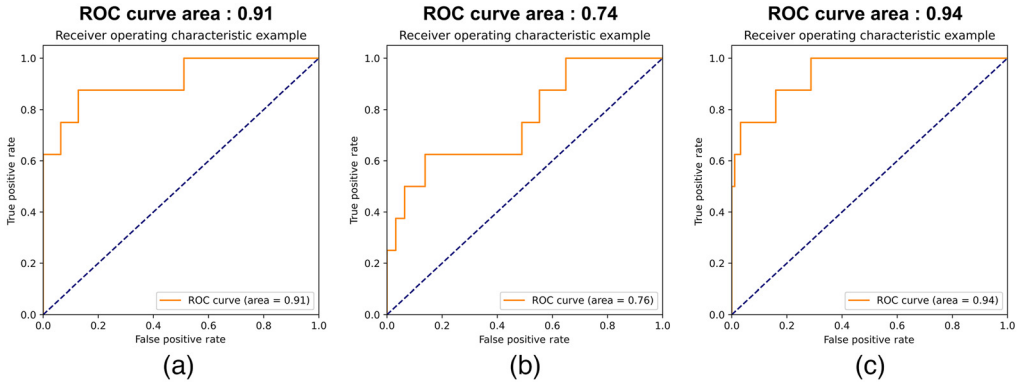


Fig. 8 Comparison of ROC curves: (a) CNN, (b) ViT, and (c) CNN–ViT models.

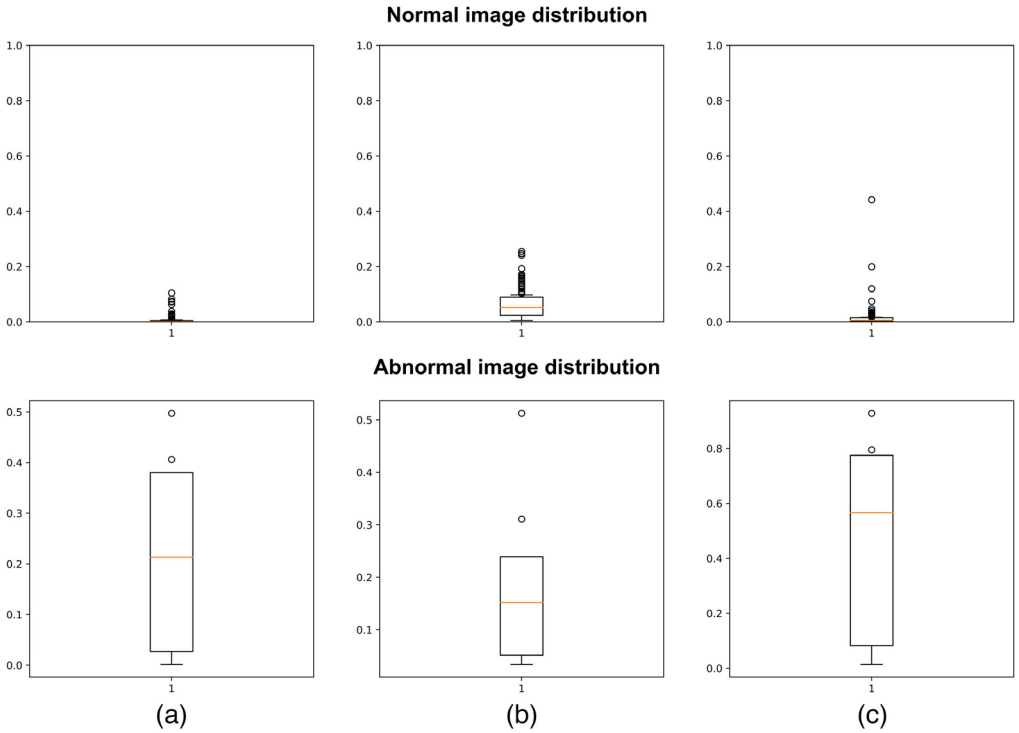


Fig. 9 Comparison of the box plot: (a) CNN, (b) ViT, and (c) CNN–ViT models.

equipment used for CBM have been replaced with ML-based algorithms. However, the cases of successful detection of anomalies in inspection targets with diverse shapes, such as signal bonds, are still few. In this contribution, we have shown, by combining multiple methods, that it is possible to detect anomalies in diverse equipment with a high degree of accuracy using ML. To achieve a good performance that allows the approach to be employed in real-life operations, it is necessary to organize an on-site proof of concept cycle to improve the outcome together with the customer who provided the dataset. The application of hybrid models is expected to go beyond railway companies and to contribute in a major way to the social implementation of CBM in fields such as food manufacturing and medicine. An effective method for anomaly detection, which reacts to both texture and edge abnormalities with high sensitivity, can contribute to distinguishing not only the defective parts in the manufacturing industry but also the inspection of food ingredients in food manufacturing. In the medical field, we believe that this approach can contribute to image analysis with a higher sensitivity than the existing techniques in diagnostic imaging support. Furthermore, to improve the visualization of the considered methods, new techniques, such as Grad-CAM++,[28,29] could be employed.

Past research has been conducted to develop robust algorithms and methods for anomaly detection in railroads. They are making steady, continuous improvements. Recently introduced approaches based on CNNs, generative adversarial networks, or ViTs have remarkably improved the research in anomaly detection. Nevertheless, there has been no attempt, so far, to use a hybrid model based on texture-biased CNNs and edge-biased ViTs for anomaly detection of signal bonds. Consequently, this study is an attempt to fill a gap in the literature. The proposed method is a hybrid model that combines two promoting architectures—texture-biased CNNs and edge-biased ViT models—to overcome some important limitations of each approach on its own. By leveraging both techniques, this ViT-based model can outperform existing architectures, especially in the low-data regime, while achieving a similar performance in large datasets.

When it comes to the study's shortcomings, ML models are still vulnerable to biases and other problems that might generate moral dilemmas. Signal fusion models for detecting anomalies are linked to certain hazards and mitigation strategies. When the definition of "unexceptional" is independently discovered and used without checks, it could unintentionally display harmful prejudices.

7 Conclusion

We proposed a method for the anomaly detection of signal bonds by combining two reasonably successful methods. Compared with previous studies of anomaly detection for railroad tracks, we found, by applying object detection (inspection target extraction) and classification, that the new method detects anomalies with high accuracy even for complex facilities with diverse shapes of inspection targets. We also found that the ViT model mainly focuses on edges in comparison with the CNN model. We proposed a CNN–ViT hybrid classification model that delivers better results than its two components, thus proving our hypothesis that the ViT model, which had already been highly accurate, could be reinforced with the output of the CNN model to provide a more effective method for anomaly detection.

Acknowledgments

We wish to thank East Japan Railway for supplying the sample data. We would also like to thank colleagues in Technical Research Center in JR East Information Systems for useful discussions. The author has no financial relationship related to this manuscript, and there are no other conflicts of interest that must be disclosed.

Code, Data, and Materials Availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to the safe operation of railways.

References

1. R. Kasai et al., "Development of a track facility monitoring device and future prospects for the device," *JR East Tech. Rev.* **34**, 21–24 (2016).
2. D. Ferreño et al., "Prediction of mechanical properties of rail pads under in-service conditions through machine learning algorithms," *Adv. Eng. Software* **151**, 102927 (2021).
3. T. Lane and C. E. Brodley, "An application of machine learning to anomaly detection," in *Proc. 20th Natl. Inf. Syst. Secur. Con.*, Baltimore, pp. 366–380 (1997).
4. Y. Kudo et al., "Fundamental development of monitoring for signal bonds," *JR East Tech. Rev.* **39**, 53–56 (2020).
5. G. Pang et al., "Deep learning for anomaly detection: a review," *ACM Comput. Surv.* **54**, 1–38 (2022).
6. Z. Li et al., "A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment," *Int. J. Adv. Manuf. Technol.* **103**, 499–510 (2019).
7. B. Wang et al., "Early event detection in a deep-learning driven quality prediction model for ultrasonic welding," *J. Manuf. Syst.* **60**, 325–336 (2021).
8. M. G. Mohammadi, D. Mahmoud, and M. Elbestawi, "On the application of machine learning for defect detection in L-PBF additive manufacturing," *Opt. Laser Technol.* **143**, 107338 (2021).
9. M. Munir et al., "DeepAnT: a deep learning approach for unsupervised anomaly detection in time series," *IEEE Access* **7**, 1991–2005 (2019).
10. S. Xu, H. Wu, and R. Bie, "CXNet-m1: anomaly detection on chest X-rays with image-based deep learning," *IEEE Access* **7**, 4466–4477 (2019).
11. S. Naseer et al., "Enhanced network anomaly detection based on deep neural networks," *IEEE Access* **6**, 48231–48246 (2018).
12. M. Tan et al., "EfficientDet: scalable and efficient object detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis., Pattern Recognit.* (2020).
13. A. Alexey et al. "An image is worth 16x16 words: transformers for image recognition at scale," https://arxiv.org/abs/2010.11929 (2020).
14. Y. Gong et al., "Anomaly detection of high-speed railway catenary damage," *IETE J. Res.* (2022).

15. A. Nugraha et al., "Detection of railroad anomalies using machine learning approach," in *Int. Conf. ICT for Smart Soc. (ICISS)*, IEEE (2021).
16. A. Sivasangari et al., "Detection of abnormalities in brain using machine learning in medical image analysis," in *Int. Conf. Sustain. Comput. and Data Commun. Syst. (ICSCDS)*, IEEE (2022).
17. H. Mohamad et al., "Abnormal events detection using deep neural networks: application to extreme sea surface temperature detection in the Red Sea," *J. Electron. Imaging* **28**(2), 021012 (2019).
18. E. Wachter et al., "Using machine learning for anomaly detection on a system-on-chip under gamma radiation," *Nucl. Eng. Technol.* **54**(11), 3985–3995 (2022).
19. K. Goto et al., "Adversarial autoencoder for detecting anomalies in soldered joints on printed circuit boards," *J. Electron Imaging* **29**(4), 041013 (2020).
20. A. Bakumenko et al., "Detecting anomalies in financial data using machine learning algorithms," *Systems* **10**(5), 130. (2022).
21. Y. Kudo et al., "Study about abnormal detection of a signal bond for track circuits," in *Proc. Railway Eng.*, p. 0018 (2017).
22. M. Suzuki, "Aiming to realize a smart society and seamless mobility with ICT: JR-EAST's challenge for business innovation," in *Business Innovation with New ICT in the Asia-Pacific: Case Studies (chapter 10)*, M. Kosaka et al., Eds., Springer Nature Singapore Pte Ltd. (2021).
23. J. Shim et al., "Anomaly detection method in railway using signal processing and deep learning," *Appl. Sci. (MDP)* **12**, 12901 (2022).
24. A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst. 25* (2012).
25. W. Brendel and M. Bethge, "Approximating CNNs with bag-of-local-features models works surprisingly well on imagenet," in *Int. Conf. Learn. Represent.* (2019).
26. R. Geirhos et al., "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," in *Int. Conf. Learn. Represent., ICLR* (2019).
27. R. R. Selvaraju et al., "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *IEEE Int. Conf. Comput. Vision (ICCV)*, pp. 618–626 (2017).
28. A. Chattopadhay et al., "Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks," in *IEEE Winter Conf. Appl. of Comput. Vision (WACV)*, Vol. 12, pp. 839–847 (2018).
29. A. Chattopadhay et al. "Grad-CAM++: improved visual explanations for deep convolutional networks," https://arxiv.org/ abs/1710.11063 (2018).
30. M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Int. Conf. Machinability Learn.* (2019).
31. I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.* (2018).
32. A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.* **30**, 1145–1159 (1997).
33. R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Comput. Ind. Eng.* **63**, 135–149 (2012).

Takuro Hoshi received his BS degree from Toho University, Japan, in 2008. He currently works at JR East Information Systems, a group of the East Japan Railway Company. His research interests include the detection and monitoring of railway images and data analysis.

Seiya Shibayama received his BA degree from Seijo University, Japan, in 2020. He currently works at JR East Information Systems, a group of the East Japan Railway Company. His research interests include image recognition and anomaly detection.

Xiaonan Jiang received his BE degree from Hubei University, China, in 2014, and his MS degree in information science and technology from Hokkaido University, Japan, in 2018. He currently works at the East Japan Railway Company. His research interests include anomaly detection and data analysis.