Imbalanced Anomaly Detection Using Augmented Deeper FCDDs for Wooden Sleeper Deterioration Prognostics

Takato Yasuno[1], Masahiro Okano[1], and Junichiro Fujii[1]

[1] *Yachiyo Engineering, Co.,Ltd., Tokyo, Taito-ku, 111-8648, Japan*
tk-yasuno@yachiyo-eng.co.jp
: ms-okano@yachiyo-eng.co.jp, jn-fujii@yachiyo-eng.co.jp

ABSTRACT

Maintaining high standards for user safety during daily railway operations is crucial for railway managers. To aid in this endeavor, top- or side-view cameras and GPS positioning systems have facilitated progress toward automating periodic inspections of defective features and assessing the deteriorating status of railway components. However, collecting data on deteriorated status can be time-consuming and requires repeated data acquisition because of the extreme temporal *occurrence imbalance*. In supervised learning, thousands of paired data sets containing defective raw images and annotated labels are required. Concretely, the one-class classification approach offers the advantage of requiring quite a few anomalous images to optimize parameters for training large normal images. The deeper fully-convolutional data descriptions (FCDDs) were applicable to several damage data sets of concrete/steel components in structures, and fallen tree, and wooden building collapse in disasters. However, it is not yet known to feasible to railway components. In this study, we devised a prognostic discriminator pipeline to automate one-class classification using the augmented deeper FCDDs for defective railway components. We also performed sensitivity analysis of the mixture and erasing augmentations, and the deeper backbone rather than the shallow baseline of convolutional neural network (CNN) with 27 layers. Furthermore, we visualized defective railway features by using transposed Gaussian upsampling. We demonstrated our application to railway inspection using a video acquisition dataset that contains wooden sleeper deterioration. Finally, we examined the usability of our approach for prognostic monitoring and future work on railway component inspection.

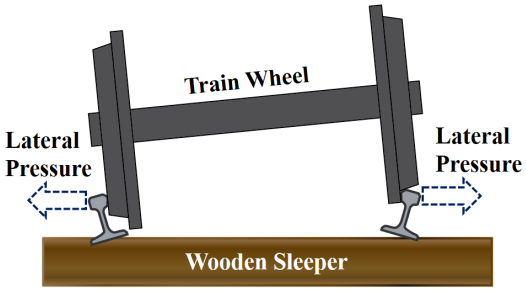

Figure 1. Illustration of derailment inside track

1. INTRODUCTION

1.1. Wooden Sleeper, Derailment Risk in Rural Railway

Derailment accident has been divided two types : derailment inside track, and riding up outside rail. As shown in Fig. 1, that is illustrated with reference to (Miwa, 2019), one of cause of derailment inside track is why decayed wooden sleeper makes strengthen the lateral pressure, and the distance between the parallel rails could expand. While analyzing fatal train accidents on Europe's mainline railways from 1980 to 2019, Evans discovered that train collisions, derailments, and railroad crossings were the primary causes of accidents (Evans, 2011)(Evans, 2020). In Japan, according to the traffic safety committee's 1987-2018 derailment statistics, there were 173 cases, with three main causes: 65 natural disasters, 47 railroad crossings, and 33 railway infrastructures (Oyama & Miwa, 2022). To prevent the potential accident of derailment, condition-based and risk-based maintenance in railway infrastructures have been one of key activities for daily operations. In detail, derailment accidents have been categorized into two types based on the railway tracks: derailment among inter-rails and riding on the outside rail. Frequently, inter-rail derailments occur because decayed wood sleepers weaken the supporting load of the rail, causing the distance between the parallel rails to expand. Rural railway managers typically operate on a small-scale and have a weak financial status, where inter-rail derailment can negatively impact the profit per day

kilometers because of the short length of running operations. Given that rural rail tracks predominantly use wood sleepers, derailment among the inter-rails is more likely to occur. Therefore, in rural railway maintenance, monitoring wooden sleeper deterioration is critical to reduce the risk of derailment. However, repairing or renewing a decayed sleeper typically requires significant human labor. To improve the performance of monthly inspections of rural railway tracks, deep learning-based visual inspection techniques can be employed.

1.2. Existing Deep Learning for Visual Track Inspection

Tang et al. reviewed articles on artificial intelligence applications in railway systems from 2010 to 2020, categorizing them into five subdomains: maintenance and inspection, traffic planning and management, safety and security, autonomous driving and control, and passenger mobility (Tang et al., 2022). The most popular research field was found to be maintenance and inspection, with over 81 papers (57%). Ji et al. reviewed existing deep learning applications for rail-track condition monitoring from 2013 to 2021, specifically focusing on supervised deep learning and recent adoption by rail industries (Ji, Woo, Wong, & Quek, 2021). The number of papers on this topic surged in 2018, and 14 regions worldwide were represented in rail-related research. This indicates that the rail industry has shown growing interest in adopting deep learning methods. Of the studies, 70% used raw image-type data for deep learning models, while the remaining types included acoustic emission signals, defectograms, maintenance records, and synthetic data from generative models. The purpose of these studies was to detect, classify, and/or localize rail surface defects, including various components such as rails, insulators, valves, fasteners, and switches.

Over 38 deep learning models have been adopted by researchers, with CNNs being popular for feature extraction and RNN/LSTM being used for sequential data types. Researchers followed a consistent process flow for deep learning applications for rail-track condition monitoring, with the first subsystem being image acquisition using cameras installed on rail maintenance vehicles to capture input data. The second subsystem involved optional prerocessing, where images were resized, enhanced, noise-removed, or cropped for target areas using image-processing techniques. The input data were then prepared for training and testing deep learning models. Finally, the trained model was produced using the parameters for real-world applications. Given the criticality of rail-track condition monitoring, inspections by human operators could double the accuracy of the system. However, the distribution of rail-track image data is often uneven and extremely disproportional, resulting in class imbalance problems. This study proposes an unsupervised learning method using a one-class classification algorithm as a novel application in rail-track condition monitoring.

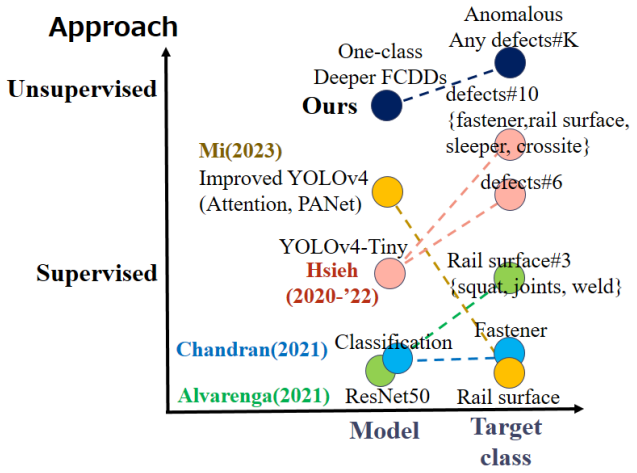

Figure 2. Related models and targets for railway prognostic inspection and our approach.

1.3. Anomaly Detection for Imbalanced Deterioration

Fig. 2 depicts several railway inspection applications that utilize deep learning models for the detection of defective classes of railway components. These applications are often based on supervised deep learning approaches such as classification (Alvarenga et al., 2021; Chandran, Asber, Thiery, Odelius, & Rantatalo, 2021) and object detection (Mi, Chen, & Zhao, 2023; Hsieh et al., 2020; Hsieh, Hsu, & Huang, 2022). In supervised learning studies, the authors assigned class labels by annotating whole images and bounding boxes that enclosed defective regions. The railway components targeted for defect detection include rail tracks, fasteners, and sleepers. Railway defects are inherently uncertain events, and the number of anomalous images is often imbalanced toward the normal class. Defect classes are not yet completely defined in railway inspections. For instance, Hsieh et al. (Hsieh et al., 2020) defined six normal classes and four defective classes by focusing on clips on wooden/concrete sleepers, spikes, fishplate, slide-bed plates, and guard rail plates. The authors collected a limited number of real images, which resulted in seven classes having fewer than 100 defective images and three classes having two or more hundred images.

It is not easy to reconstruct synthetic images to represent the health condition of railway components amidst a complex background consisting of trees, grass, and ballast stone. Furthermore, generating defective features as annotation data that can contribute to architectural performance is challenging. Collecting defective status data to build a railway inspection application always requires significant time investment, given that the temporal occurrence of defects is extremely imbalanced. To achieve stable and high performance, a supervised learning approach demands thousands of paired datasets consisting of defective real images and annotated labels or bounding boxes.

In contrast, few studies have been published on the unsupervised deep learning approach to automate visual inspection in the field of railway track. Especially, the one-class classification based anomaly detection has the advantage of requiring quite a few anomalous images to optimize the parameters for training relatively large normal images. Additionally, the visual heat map explanation enables us to discriminate between localized defective features. The authors (Yasuno, Okano, & Fujii, 2023) found that the deeper fully-convolutional data descriptions (FCDDs) has been applicable to several damage data sets of concrete/steel components in structures: pavement, bridge, and dam, and fallen tree, and wooden building collapse in disasters: typhoon, earthquake. However, it is not yet known to feasible to railway components that includes deterioration of wooden sleepers. In this study, we propose a prognostic discriminator pipeline to automate the one-class classification of defective railway components.

2. AUGMENTED ANOMALY DETECTION

2.1. Augmented Deeper FCDDs and Risk-weighted Score

The authors (Yasuno et al., 2023) have already reported the deeper FCDDs and found the applicability to damage data sets of bridge, dam, and building. However, as an unsupervised deep anomaly detection approach, the deeper FCDDs has been not yet known to feasible to video frame images in railway track that contains ballast stones, rail, spike, fastener, and concrete/wooden sleepers. For risk-based maintenance of railways, visualizing hazard-mark heatmaps and computing risk-weighted anomaly scores is crucial for rural railway inspection and prognostic support for effective repairs under usable resources: time, labors, and budget.

Let F_i be the i-th frame of an image with a size of $h \times w$. Let A_i be the augmented image using a preprocess $A_i = g(F_i)$ mapping from the frame of raw image F_i. We consider the number of training images and the weight W of the fully convolutional network (FCN). Let the $\Phi^B_W(A_i)$ denote a mapping of the deeper CNN to backbone B based on the input frame image. The one-class classification model was formulated using the cross-entropy loss function as follows:

$$\begin{aligned}\mathcal{L}_{DeepSVDD} = &-\frac{1}{n}\sum_{i=1}^{n}(1-z_i)\log \ell(\Phi^B_W(A_i)) \\ &+ z_i \log[1-\ell(\Phi^B_W(A_i))],\end{aligned} \tag{1}$$

where $z_i = 1$ denotes the anomalous label of the i-th frame of the image and $z_i = 0$ denotes the normal label of the i-th frame of the image. A pseudo-Huber loss function is introduced to obtain a more robust loss formulation (Ruff, Vandermeulen, Franks, Müller, & Kloft, 2021) in Equation (2). Let $\ell(u)$ be the loss function and define the pseudo-Huber loss as follows:

$$\ell(u) = \exp(-H(u)),\ H(u) = \sqrt{\|u\|^2+1} - 1. \tag{2}$$

Then, a deeper FCDD loss function can be formulated :

$$\begin{aligned}\mathcal{L}_{deeperFCDD} = &\frac{1}{n}\sum_{i=1}^{n}\frac{(1-z_i)}{uv}\sum_{x,y}H_{x,y}(\Phi^B_W(A_i)) \\ &- z_i \log\left[1-\exp\left\{\frac{-1}{uv}\sum_{x,y}H_{x,y}(\Phi^B_W(A_i))\right\}\right],\end{aligned} \tag{3}$$

where $H_{x,y}(u)$ are the elements (x, y) of the receptive field of size $u \times v$ under a deeper FCDD. In the equation (3), if we set an unsupervised learning, the positive second term are canceled out. If we use an imbalanced data that includes fewer anomalous images and relatively large normal images, a deeper FCDD loss function (3) is less influenced by the positive second term. The risk-weighted anomaly score S^{rw}_i of the i-th image is expressed as the sum of all the elements of the receptive field as follows:

$$S^{rw}_i(B) = r_i \sum_{x,y} H_{x,y}(\Phi^B_W(A_i)),\ i = 1, \cdots, n. \tag{4}$$

Here, r_i is the weight of the derailment risk caused by the wooden sleeper deterioration. For example, for larger ratios of the curve, the weight r_i can be set higher. Specifically, we can provide i-th ratio of the curve to match the position by the global navigation satellite system (GNSS). In contrast, when the deteriorated wooden sleeper and the precast concrete (PC) sleeper are adjacent, the weight r_i may be set lower. In the present study, we set the risk neutral environment that $r_i = 1$.

We herein present the construction of a baseline FCDD (Yasuno et al., 2023) with an initial backbone $B = 0$ and performed CNN27 mapping $\Phi^0_W(A_i)$ from the augmented input image A_i in the dataset. We also present deeper FCDDs focusing on elaborate backbones $B \in \{$VGG16, ResNet101, Inceptionv3$\}$ with a mapping operation $\Phi^B_W(A_i)$ to achieve more elaborately performance. In this paper, we present ablation studies on a rural railway dataset for detection towards decayed wooden sleeper.

2.2. Mixture and Erasing Augmentations

The past decade has seen renewed importance of augmented image data for deep learning (C. Shorten, 2019). Image data augmentation is categorized into two i.e., basic image manipulation and deep learning approaches. The former augmentation includes 1) kernel filters, 2) geometric transformations, 3) random erasing, 4) mixing images, and 5) color space transformations. The latter consists of three components: 1) adversarial training, 2) neural style transfer, and 3) generative learning model-based transformation.

To address the challenges of imbalanced fewer anomalous data and complex rail track scene, we think that augmentation preprocessing such as mixup, and random erasing can be effective for one-class classification models. In the present

study, we proposes the preprocess of image manipulations using mixing and random erasing techniques as follows. Using multiple raw images in the imbalanced fewer anomalous data and relatively large normal one, the mixture image augmentation creates diversified images, such as the Random Erasing (Zhong, Zheng, Kang, Li, & Yang, 2020), mixup (Zhang, Cisse, Dauphin, & Lopez-Paz, 2018)(Tango, Ohkawa, Furuta, & Sato, 2022), Cutout (Devries & Taylor, 2017), and RICAP (Takahashi, Matsubara, & Uehara, 2018). To detangle the ballast and grassy feature in railway track, the random erasing technique is straightforward and efficient as the regional dropout in the training images. We also believe that the augmentation of mixing images diversifies the deteriorated region of interest on the wooden/precast sleeper and ballast background on the imbalanced data set in railway track.

2.3. Hazard-mark Heatmap Upsampling Receptive Field

Convolutional neural network (CNN) architectures, comprising millions of common parameters, have exhibited remarkable performance for visual inspection, but the underlying reasons for this superiority remain unclear. Heatmap visualization techniques for detecting and localizing anomalous features are typically categorized as masked sampling and activation map approaches. The former includes methods such as occlusion sensitivity (Zeiler & Fergus, 2013) and local interpretable model-agnostic explanations (Ribeiro, Singh, & Guestrin, 2016). The latter category includes activation maps such as class activation maps (CAMs) (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2015) and gradient-based extensions (Grad-CAM) (Selvaraju et al., 2017). Nonetheless, aforementioned methods of disadvantage is its requirement for parallel computation resources and iterative computation time for local partitioning, masked sampling, and for generating a gradient-based heatmap.

In this study for railway inspection applications, we adopt the receptive field upsampling approach (Liznerski et al., 2021) to visualize anomalous features using an upsampling-based activation map with Gaussian upsampling from the receptive field of the FCN. The primary advantages of the upsampling approach are the reduced computational resource requirements and shorter computation times. The proposed upsampling algorithm generates a full-resolution anomaly heatmap from the input of a low-resolution receptive field $u \times v$.

Let $H \in R^{u \times v}$ be a low-resolution receptive field (input), and let $H' \in R^{h \times w}$ be a full-resolution of hazard-mark heatmap (output). We define the 2D Gaussian distribution $G_2(m_1, m_2, \sigma)$ as follows:

$$[G_2(m_1, m_2, \sigma)]_{x,y} \equiv \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-m_1)^2 + (y-m_2)^2}{2\sigma^2}\right). \quad (5)$$

The Gaussian upsampling algorithm from the receptive field is implemented as follows:

1. $H' \leftarrow 0 \in R^{h \times w}$
2. for all output pixels d in $H \leftarrow 0 \in R^{u \times v}$
3. $\quad u(d) \leftarrow$ is upsampled from a receptive field of d
4. $\quad (c_1(u), c_2(u)) \leftarrow$ is the center of the field $u(d)$
5. $\quad H' \leftarrow H' + d \cdot G_2(c_1, c_2, \sigma)$
6. end for
7. return H'

After conducting experiments with various datasets, we determined that a receptive field size of 28×28 is a practical value. When generating a hazardous heatmap, unlike a revealed damage mark, we need to unify the display range that corresponds to the anomaly scores, which range from the minimum to the maximum value. In order to strengthen the defective regions and highlight the hazard marks, we define a display range of [min, max/4], where the quartile parameter is 0.25. This results in the histogram of anomaly scores having a long-tailed shape. If we were to include the complete anomaly score range, the colors would weaken to blue or yellow on the maximum side.

Table 1. Imbalanced training dataset of wooden sleeper deterioration.(In both class, the calibration 150 and test 200 images were fixed respectively.)

Positive ratio	Anomalous	Normal
1/1(supervised)	650	650
1/2	325	650
1/4	163	650
1/8	81	650
1/16	41	650
1/32	20	650
1/64	10	650
1/650(unsupervised)	1	650

3. Applied Results

3.1. Field Data Acquisition and Dataset Preparedness

As presented in Table 1, we have demonstrated a railway-related application through an experimental study on a rural railway track. In the present study, we prepared 2000 images that contains anomalous 1000 and normal 1000. Herein, the whole 1000 images in each class were partitioned at a ratio of 65:15:20 for the training 650, calibration 150, and test 200. In both class, the calibration and test images were fixed respectively.

We collected the dataset by recording videos using a camera mounted on a train traveling along a single track with a length of approximately 80 km in Japan. It iterated 2 days with sunny weather (March 9, 2023) and cloudy weather (April 26, 2023). The videos were recorded at a rate of 30 frames per second, which provided too much information to be used directly for learning an anomaly detection model. Therefore,

we used every fourth frame to generate 79 thousand images, which were then overlapped to represent the railway track in its entirety.

In the study, we created a dataset for demonstration to automate anomaly detection toward the wooden sleeper deterioration. We used transfer learning based on ResNet18 and ResNet101 to build two classification models to prepare the input data from the cropped images, summarized in follows.

1. Cropping top far and horizontal edges to reduce outside the region of interest.
   - To minimize background noise, we cropped each 4K frame to a size of 1200×1920 from the original size 2160×3840.
   - Crop the top far view of upper region with the height of 880 pixels(40.7%)where getting rough pixel data.
   - Crop the horizontal edges 1280 pixels(33.3%) to sum up the left-side of 640 pixels and right-side of 640 pixels outside the rail track.
2. Shadow/dark/without classification.
   - Since many locations contained large shadows or dark conditions, we built a shadow/dark/without shadow, classifier using ResNet18 with three classes: shadows in sunny weather, whole darkness in tunnel, and clear track without shadow.
   - We randomly selected 3000 images from the cropped images and labeled them as 1458 shadows, 152 whole darkness, and 1390 without shadows.
   - We trained the model using mini-batch 32 and 15 epochs, iterated using Adam, which resulted in a test accuracy of 96.7%.
   - We predicted the 79 thousand cropped images using the shadow/dark/without classifier and categorized them as 15006(19.0%) shadow in sunny weather, 8894(11.2%) whole darkness in tunnel, and 55180(69.8%) clear track without shadow.
3. Grassy/decayed wooden sleeper/normal classification.
   - There were conditionally grassy spots on the ballast track, wooden sleepers, and outside the track in each frame image.
   - Therefore, we built a grassy/decayed wooden sleeper/normal classifier using ResNet101 with three classes: grassy, decayed wooden sleeper, and normal without grass.
   - We randomly selected 1500 images from the clear track of cropped images and labeled them as 270 grassy, 435 decayed wooden sleepers, and 547 normal without grass.
   - We trained the model using a mini-batch of 128 and 15 epochs, iterated using Adam, which resulted in a test accuracy of 66.4%.
   - We predicted the 55180 clear track of cropped images without shaddows, using the grassy/decayed wooden sleeper/normal classifier and categorized them as 6488(27.0%) grassy, 5659(23.6%) decayed wooden sleeper, and 11853(49.4%) normal class.
4. Finally, for the present imbalanced studies, we annotated 1000 decayed wooden sleepers and 1000 normal images that includes neither shaddow nor grass.

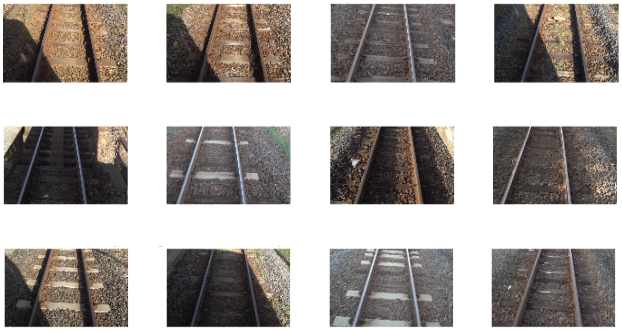

Figure 3. Cropped images of predicted results by first trained the ResNet18 classifier with 3 classes: shadow/dark/without.

As depicted in Fig. 3, the test images illustrate examples predicted by our initial classifier for shadow noise reduction, where we developed a classification model utilizing ResNet18 with three distinct classes : shadows in sunny weather, complete darkness in tunnel, and clear track without shadows.

3.2. Backbone Studies of Supervised Anomaly Detector

In the balanced data situation as shown in Table 1, the training data contains anomalous 650 images and normal 650 images. Herein, We implemented deeper backbone studies using the VGG16, ResNet101, Inceptionv3. During the training of the anomaly detector, we fixed the input size to 224^2. To train the model, we set the mini-batch size to 32 and ran 60 epochs. We used the Adam optimizer with a learning rate of 0.0001, a gradient decay factor of 0.9, and a squared gradient decay factor of 0.99. The training images were partitioned at a ratio of 65:15:20 for the training, calibration, and testing images. As shown in Table 2, our deeper FCDDs based on VGG16 (abbreviated as *deeperFCDD-VGG16*) outperformed the baseline CNN27 and other backbone-based deeper FCDDs in this railway dataset for detecting decayed wooden sleeper.

Table 2. Backbone ablation studies on defective detection using our proposed deeper FCDDs for Wooden sleeper.

Backbone	AUC	F_1	Precision	Recall
CNN27	0.8624	0.7688	0.7088	0.8400
VGG16	**0.9425**	**0.8475**	**0.8770**	**0.8200**
ResNet101	0.9304	0.8108	0.8823	0.7500
Inceptionv3	0.9412	0.8041	0.8415	0.7700

3.3. Imbalanced and Unsupervised Training Studies

As shown in Table 3, we implemented ablation studies on imbalanced data that contains fewer anomalous images and relatively large normal images. Herein, we applied our deeper FCDD-VGG16 achieved high performance in the aforementioned supervised results. Compared with the balanced case of positive ratio 1/1, we found that there was applicable range from the balanced ratio 1/2 to the ratio 1/16 where the accuracy were consistently high performance. However, extremely imbalanced range from 1/32 to 1/650, that accuracy were inferior to those of aforementioned applicable range. The rare positive ratio 1/32 represent the imbalanced data that contains quite a few 20 anomalous images and relatively large 650 normal images. In the situation, much more anomalous images should be acquired and added to the initial dataset.

Table 3. Imbalanced data studies using our deeper FCDD-VGG16 for wooden sleeper deterioration.

Positive ratio	AUC	F_1	Precision	Recall
1/1(superv.)	**0.9425**	**0.8475**	**0.8770**	**0.8200**
1/2(ano.325)	0.9467	0.8031	0.8579	0.7550
1/4(ano.163)	0.9468	0.8536	0.8472	0.8600
1/8(ano.81)	0.9469	0.8366	0.8284	0.8450
1/16(ano.41)	0.9354	0.8516	0.8165	0.8900
1/32(ano.20)	**0.9255**	**0.8131**	**0.8214**	**0.8050**
1/64(ano.10)	**0.8942**	**0.8213**	**0.7662**	**0.8850**
1/650(unsuperv.)	**0.6279**	**0.6250**	**0.6845**	**0.5750**

3.4. Augmention Studies of Supervised Anomaly Detector

As shown in Table 4, we demonstrated the mixture and random erasing preprocess to our deeper FCDD-VGG16 on the balanced data that the number of normal and anomalous class are same value respectively. Compared to never augmentation case in the first row, the Gaussian noise-based Random Erasing augmentation improved accuracy in terms of the F_1 and recall rather than others such as the mixup($\alpha = 0.2$), RICAP, and Cutout(dropout ratio 0.5). To automate visual inspection, the recall is critical for minimizing false positive error. We found that the Random Erasing augmentation improved recall accuracy without large loss of precision in the balanced data set.

Table 4. Augmentation studies using our deeper FCDD-VGG16 for wooden sleeper deterioration.

Augmentation	AUC	F_1	Precision	Recall
never Aug.	0.9425	0.8475	0.8770	0.8200
mixup	0.9401	0.8467	0.8810	0.8150
RICAP	0.9392	0.8377	0.8791	0.8000
Cutout	0.9380	0.8238	0.8994	0.7600
RandomErasing	**0.9383**	**0.8719**	**0.8592**	**0.8850**

3.5. Hazard-mark Heatmaps for Railway Prognostics

We visualized he damage features by using Gaussian upsampling in our Random Erasing deeper FCDD-VGG16 network. Additionally, we generated a histogram of the anomaly scores of the test images in the balanced case. In Fig. 4, a hazard-mark explanation is presented. The red region in the heatmap represents the decayed wooden sleepers. There is few region of false negative error over the ballast stones, precast concrete sleeper, and grass outside rail track. Fig. 5 illustrates that several overlapping bins exist in the horizontal anomaly scores. Therefore, for detecting decayed wooden sleepers, the score range was moderately separated.

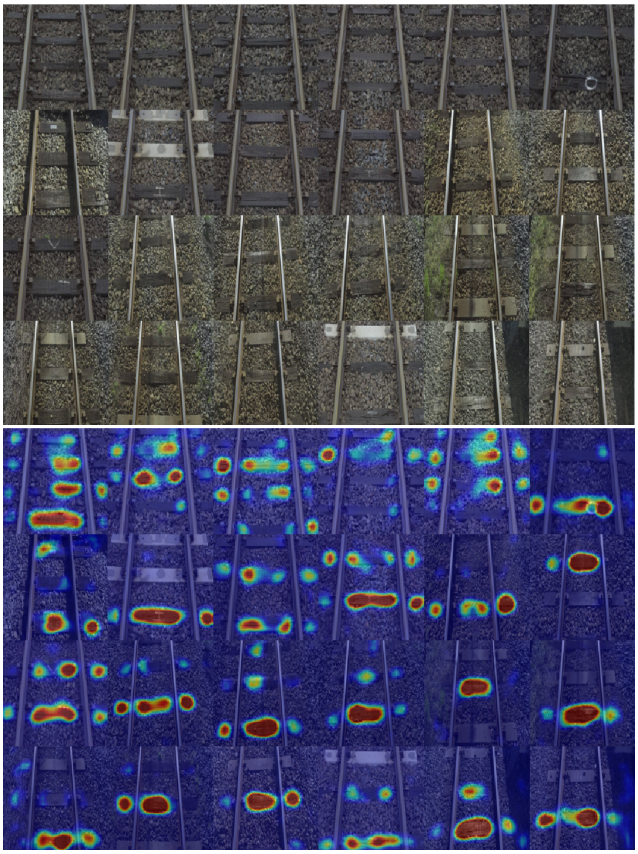

Figure 4. Input raw images (top) and hazard-mark heatmaps (bottom) of decayed wooden sleeper using our Random Erasing deeper FCDD-VGG16.

4. Concluding Remarks

4.1. Railway Inspection Application for Imbalanced Data

We developed a railway inspection application to automate one-class anomaly detection. To ensure feasibility of a railway application, we assessed an unsupervised and imbalanced anomaly detection approach for deeper FCDDs with pretrained backbones of VGG16, ResNet101, and Inceptionv3. To support decision to repair the wooden sleeper and

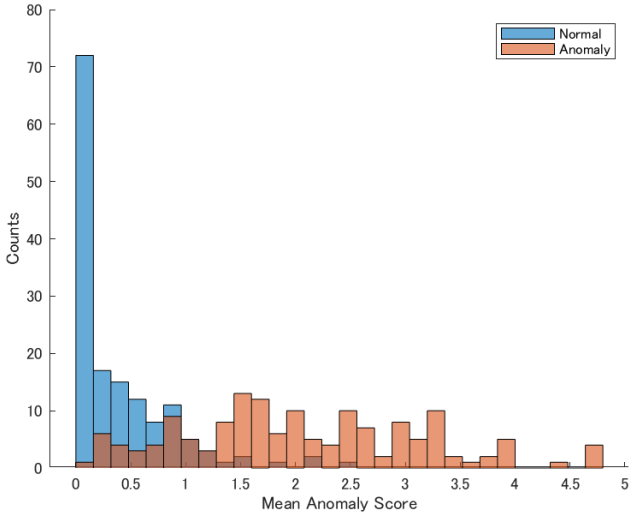

Figure 5. Histogram of decayed wooden sleeper scores corresponding to our Random Erasing deeper FCDD-VGG16.

minimize the possibility of derailment for safe railway operations, we represented an index of the risk-weighted anomaly score. Additionally, we visualized hazard-mark heatmaps using direct Gaussian upsampling of the receptive field of the FCN. In the present study, we evaluated our augmented deeper FCDDs on the targets of decayed wooden sleepers.

Our experiments produced a high accuracy over 87% in terms of F_1 and recall using our Random Erasing deeper FCDD-VGG16 architecture. That created the hazard-mark heatmap for visual explanation, even without annotating the decayed wooden sleeper at the localized regions. In our imbalanced studies, compared with the balanced case of positive ratio 1/1, we found that there was applicable range from the balanced ratio 1/2 to 1/16, where the accuracy were consistently high. However, extremely imbalanced range from 1/32 to 1/650, whose accuracy were inferior to those of aforementioned applicable range. Thus, our work presented a novel solution for augmented deeper FCDDs that offers a imbalanced deterioration detection tool for visual railway inspections and hazard-explainability.

4.2. Limitations and Future Works

This study discovered the feasibility of imbalanced anomaly detection highlighting wooden sleeper in railway track. However, this is limited on the target of wooden sleeper deterioration. The video images had few variations collected at spring season. Winter season may influence the unseen noise: shadow from west sunshine, decayed grass, iced, and snowy. The Random Erasing deeper FCDD-VGG16 achieved over 87% in terms of F_1 and recall. For more accurate applications, we have an anomalous data mining opportunity to train the remained dataset of over 5000 images that we have classified into 2000 decayed wooden sleeper, and 3000 normal images. Several promising directions exist for future research to improve the usability of visual inspection applications. To address the challenges of imbalanced data, this issue remains for infrequent defects such as spikes out of rail from wooden sleepers, cracks of concrete sleepers, and holes on ballast tracks. To overcome this challenge, the anomaly score generated by our deeper FCDDs can be used in edge devices for effective data acquisition of rare classes. By collecting only the frames that have hazard marks with significantly higher anomaly scores than a predefined threshold, the data acquisition process can be made more efficient. For risk-based maintenance to incorporate the potential hazard of derailment in each track, we are able to add the label of curve ratio (Oyama & Miwa, 2022). Then, we can utilize the risk-weighted anomaly score for supporting to make a decision of priority to repair for sustainable and safety operations.

ACKNOWLEDGMENT

We gratefully acknowledge the conductive comments of the anonymous referees. The authors wish to thank MathWorks and Computer Vision Toolbox Team for providing helpful MATLAB resources for Automated Visual Inspection. We also thank Nakasha Creative Co.,Ltd., for providing the opportunity to study rural railways.

REFERENCES

Alvarenga, T. A., Carvalho, A. L., Honorio, L. M., Cerqueira, A. S., Filho, L. M. A., & Nobrega, R. A. (2021). Detection and classification system for rail surface defects based on eddy current. *Sensors*, *21*(23).

Chandran, P., Asber, J., Thiery, F., Odelius, J., & Rantatalo, M. (2021). An investigation of railway fastener detection using image processing and augmented deep learning. *Sustainability*, *13*(21).

C. Shorten, T. K. (2019). A survey on image data augmentation for deep learning. In *Proceedings of the 35th international conference on machine learning* (Vol. 6:60).

Devries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *ArXiv*, *abs/1708.04552*.

Evans, A. W. (2011). Fatal train accidents on Europe's railways: 1980–2009. *Accident Analysis and Prevention*, *43*(1), 391-401.

Evans, A. W. (2020). *Fatal train accidents on Europe's railways: 1980–2019* (Tech. Rep.). Centre for Transport Studies, Imperial College London.

Hsieh, C.-C., Hsu, T.-Y., & Huang, W.-H. (2022). An online rail track fastener classification system based on YOLO models. *Sensors*, *22*(24).

Hsieh, C.-C., Lin, Y.-W., Tsai, L.-H., Huang, W.-H., Hsieh, S.-L., & Hung, W.-H. (2020). Offline deep-learning-based defective track fastener detection and inspection

system. *Sensors and Materials*, *32*(10), 3429.

Ji, A., Woo, W. L., Wong, E. W. L., & Quek, Y. T. (2021). Rail track condition monitoring: A review on deep learning approaches. *Intelligence and Robotics*, *1*(2), 151–175.

Liznerski, P., Ruff, L., Vandermeulen, R. A., Franks, B. J., Kloft, M., & Müller, K.-R. (2021). Explainable deep one-class classification. In *The international conference on learning representations(ICLR).*

Mi, Z., Chen, R., & Zhao, S. (2023). Research on steel rail surface defects detection based on improved YOLOv4 network. *Frontiers in Neurorobotics*, *17*.

Miwa, M. (2019). Railway maintenance transformed by Numerical Optimizer. In *Mathematical systems user conference 2019.*

Oyama, T., & Miwa, M. (2022). Applying probabilistic mathematical modeling approach and ai technique to investigate serious train accidents in japan. *Sustainability Analytics and Modeling*, *2*, 2667-2596.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (p. 1135–1144). Association for Computing Machinery.

Ruff, L., Vandermeulen, R. A., Franks, B. J., Müller, K.-R., & Kloft, M. (2021). Rethinking assumptions in deep anomaly detection. In *The international conference on machine learning (ICML).*

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE international conference on computer vision (ICCV)* (p. 618-626).

Takahashi, R., Matsubara, T., & Uehara, K. (2018). Ricap: Random image cropping and patching data augmentation for deep cnns. In *Proceedings of the 10th asian conference on machine learning* (Vol. 95, pp. 786–798). PMLR.

Tang, R., De Donato, L., Besinovic, N., Flammini, F., Goverde, R. M., Lin, Z., ... Wang, Z. (2022). A literature review of artificial intelligence applications in railway systems. *Transportation Research Part C: Emerging Technologies*, *140*, 103679.

Tango, K., Ohkawa, T., Furuta, R., & Sato, Y. (2022). *Background mixup data augmentation for hand and object-in-contact detection.*

Yasuno, T., Okano, M., & Fujii, J. (2023). One-class damage detector using deeper fully convolutional data descriptions for civil application. *Advances in Artificial Intelligence and Machine Learning*, *3(2)*, 996-1011.

Zeiler, M. D., & Fergus, R. (2013). *Visualizing and understanding convolutional networks.*

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond empirical risk minimization. In *International conference on learning representations.*

Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020). Random erasing data augmentation. In *The 34th aaai conference on artificial intelligence (AAAI-20).*

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2015). *Learning deep features for discriminative localization.*

BIOGRAPHIES

Takato Yasuno (ORCID: 0000-0002-4796-518X) He received D.E. degree from Tottori University. As a civil engineer in asset management planning, he has 23 years of experience in 128 consulting projects. Since 2017, he has been working at the Research Institute for Infrastructure Paradigm Shift (RIIPS) as a senior researcher at Yachiyo Engineering Co., Ltd. (YEC). He is also a member of the Japanese Society for Artificial Intelligence (JSAI). He has published over 20 articles on machine-learning methodologies for applications in civil engineering and natural disaster. His research interests include data mining, machine learning, automated damage inspection for diagnosis and prognostics.

Masahiro Okano He graduated from the Japan Electronics College in data science and development and operations (DevOps) for AI systems. He has experience in building deep neural network architectures and deploying civil and environmental applications. He works in the RIIPS as a researcher at YEC. He is a member of the JSAI. His research interests include prompt design and few-shot predictions using foundation models in civil engineering.

Junichiro Fujii He received B.E degree from Kyoto University and M.A.S. (Interdisciplinary Information Studies) degree from University of Tokyo. He had more than 20 years of experience in information systems development. He works in the RIIPS as the chief of the AI analysis Team at YEC. He is a member of the JSAI. His research interests include the application of artificial intelligence in civil engineering.