A I S ARTIFICIAL INTELLIGENCE STUDIES

Diffusion-Based Anomaly Detection for Railway Track Fault Diagnosis

Canan Tastimur[*a]

[a,*] Erzincan Binali Yildirim University,
Engineering Architecture Faculty,
Dept. of Computer Engineering
24100 – Erzincan, Türkiye
ORCID: 0000-0002-3714-6826

[*] Corresponding author.
e-mail: ctastimur@erzincan.edu.tr

ABSTRACT

Early detection of railway rail surface defects is critical. Late or missed detection of these defects can result in loss of life and property. Monitoring the smart maintenance processes of railway rail lines, a critical infrastructure within the scope of Industry 4.0, offers a significant advantage. This study presents an innovative method for anomaly detection of potential rail surface defects on railway lines. The proposed approach is based on a diffusion model and accelerates and improves the maintenance process by detecting rail deterioration early. This method not only eliminates the need for data labeled as faulty or healthy, but also allows the model to predict fault types previously unseen using unsupervised learning on test data. Model training is completed using only healthy rail images. In the test phase, the model reconstructs the faulty images and calculates anomaly scores from pixel-level differences. Experimental studies were conducted on a publicly available rail fault dataset to evaluate the method's performance. The study also examined the distribution of anomaly scores, score histograms, and heat maps in detail.

Keywords: Anomaly detection, Diffusion model, Rail defect, Railway infrastructure monitoring

Anahtar Kelimeler: Anomali tespiti, Demiryolu altyapısı izleme, Ray arızası, Difüzyon modeli

Submitted: 05.10.2025
Revised: 15.12.2025
Accepted: 31.12.2025

doi:10.30855/ais.2025.08.02.02

Demiryolu Ray Arıza Teşhisi için Difüzyon Tabanlı Anomali Tespiti

ÖZ

Demiryolu ray yüzey arızalarının erken tespiti kritik öneme sahiptir. Bu arızaların geç veya gözden kaçması can ve mal kaybına neden olabilir. Endüstri 4.0 kapsamında kritik bir altyapı olan demiryolu ray hatlarının akıllı bakım süreçlerinin izlenmesi önemli bir avantaj sağlar. Bu çalışma, demiryolu hatlarındaki potansiyel ray yüzey kusurlarının anomali tespiti için yenilikçi bir yöntem sunmaktadır. Önerilen yaklaşım bir difüzyon modeline dayanmaktadır ve ray bozulmalarını erken tespit ederek bakım sürecini hızlandırır ve iyileştirir. Bu yöntem, hatalı veya sağlıklı olarak etiketlenen verilere olan ihtiyacı ortadan kaldırmanın yanı sıra, modelin test verileri üzerinde gözetimsiz öğrenme kullanarak daha önce görülmemiş hata türlerini tahmin etmesini de sağlar. Model eğitimi yalnızca sağlıklı ray görüntüleri kullanılarak tamamlanır. Test aşamasında, model hatalı görüntüleri yeniden oluşturur ve piksel düzeyindeki farklılıklardan anomali puanları hesaplar. Yöntemin performansını değerlendirmek için herkese açık bir ray arıza veri kümesi üzerinde deneysel çalışmalar yürütülmüştür. Çalışmada ayrıca anomali puanlarının dağılımı, puan histogramları ve ısı haritaları ayrıntılı olarak incelenmiştir.

To cite this article: C. Tastimur, “Diffusion-Based Anomaly Detection for Railway Track Fault Diagnosis,” *Artificial Intelligence Studies,* vol.8, no.2, pp. 116-124, 2025. doi:10.30855/ais.2025.08.02.02

1. Introduction

The issue of ensuring the safety of railway lines has been widely studied in recent years. Early detection of rail faults is essential for the safe and sustainable operation of railway infrastructure. To prevent both transportation disruptions and loss of life and property, intelligent predictive maintenance methods have been the subject of research. Traditionally, manual inspection of rail faults has been characterized by time-consuming, labor-intensive, and human error, leading to the adoption of artificial intelligence-supported predictive maintenance methods [1, 2]. Cracks, surface defects, and missing fasteners on railway rail surfaces endanger transportation safety. Early detection of these faults in railway infrastructure is a requirement of intelligent transportation systems within the scope of Industry 4.0 [3].

The development of intelligent transportation systems has led to the digitalization of railway infrastructures. Real-time monitoring of rails and automatic detection of maintenance requirements represent a part of this process. Small-scale faults on railway rail surfaces have the potential to cause major accidents in the future [4]. Thanks to the intelligent predictive maintenance process, periodic rail inspection can be completed and fully meet the needs of modern systems [4]. In recent years, image-based deep learning techniques have been developed for fault diagnosis, and these solutions have produced promising results in early fault diagnosis in intelligent transportation [5]. Supervised deep learning architectures, in particular, have achieved high performance in classifying faults in rail images. However, such approaches also require large amounts of labeled data. Obtaining this data is time-consuming and costly [6]. The role of advanced image-based systems is increasing, as predictive maintenance systems optimize maintenance scheduling and reduce human intervention thanks to the approaches in the literature [7, 8].

Mittal and Rao [9] presented a transfer learning-based approach for railway rail fault diagnosis. Zhank et al. [10] detected rail surface cracks using transfer learning-based InceptionV3 and YOLOv5 methods. Despite the need for labeled data, an accuracy of around 94% was achieved. In another transfer learning-based study [11], ResNet50v, VGG16, MobileNet, and YOLOv11 models were compared. They achieved 96.4% accuracy in rail surface fault diagnosis with YOLOv11. Ferdousi et al. [12] proposed transfer learning-based rail fault diagnosis using ResNet-50, VGG-19, and MobileNetV3 architectures. In another study, rail fault classification based on InceptionV3 and YOLOv5 was proposed [13]. An accuracy of 94% was achieved. In the hybrid model of YOLOv8-Segment and self-supervised U-Net, Anita et al. [14] They classified five different rail faults and achieved 95% mAP. Another study [15] used the YOLOv8 model and performed rail fault diagnosis with the GhostHGNetV2 architecture. The model size was reduced by 70%, while the mAP50 reached approximately 96%, making it ideal for real-time applications.

In [16], the Fast-RCNN architecture was trained to classify surface defects on the rail surface, and 95% accuracy was achieved. The method produced successful results despite the different lighting conditions and data diversity of the rail images, but it still requires a large amount of labeled data. In another study [17], three different methods, YOLOv5, Faster-R-CNN, and Efficient Net, were compared for the detection of rail defects. The highest recall accuracy was achieved with Faster-R-CNN with 93%. In this study, data preprocessing and data augmentation were also applied to the dataset. Sharma et al. [18] developed an FPGA-based edge AI for real-time rail defect detection. The developed CNN-based architecture for rail defects produced 88.9% accuracy. Kooban et al. [19] worked on the detection of rail components. They worked with GPR data, imagery and radar data to detect faults in components such as ballast and drainage. Wang et al. [20] also worked on the detection of foreign objects on the rail surface. They performed anomaly detection using reconstructed error maps. In this respect, the motivation for our work is similar.

In this study, an innovative deep learning-based predictive maintenance method is developed for the early detection of railway track faults. The diffusion model's advanced image processing capabilities and robust feature extraction allow us to identify anomalies on the rail surface. Diffusion models provide high-quality reconstructions by filtering out noise. While traditional CNN-based approaches have the disadvantages of requiring labeled data and limited generalization, the diffusion model does not require labeled data and produces successful results in detecting anomalies on the rail surface that the model has not previously seen. The model improves the learning process by balancing between the original and noisy images during training. In this study, anomaly regions are visualized using the reconstruction faults generated by the model. Experiments were conducted on data under different lighting conditions. The proposed approach provides an early warning mechanism for rail safety. Future work will focus on integrating the proposed propagation-based framework into real-time monitoring systems for railway infrastructure. Table 1 summarizes the methods and performance results used in recent studies on railway fault detection.

Table 1. Literature comparison for railway track fault detection

Reference	Method	Metrics	Disadvantages
[9]	CNN	Accuracy 94%	Labeled data requirement
[10]	InceptionV3, YOLOv5	Accuracy 94%	Labeled data requirement, complex model
[11]	ResNet50, VGG16, MobileNet, YOLOv11	Accuracy 96.4%	Very low accuracy for YOLOv11
[13]	InceptionV3 ve YOLOv5	Accuracy 94%	Labeled data requirement
[14]	YOLOv8-Segment + Self-supervised U-Net	mAP 95%	Complex model structure
[15]	YOLOv8 + GhostHGNetV2	mAP50 96%	Training difficulties
[16]	Fast-RCNN	Accuracy 95%	Large, labeled data requirement
[17]	YOLOv5, Faster-RCNN, EfficientNet	Recall 93%	Big data requirement
[18]	FPGA tabanlı CNN Edge AI	Accuracy 88.9%	Low accuracy compared to other models

2. The Proposed Approach

This study presents an artificial intelligence-assisted fault diagnosis method for ensuring railway safety. A new approach, a diffusion-based deep learning model, was developed. Because existing CNN-based approaches in literature require labeled data and have limited generalization capabilities, a diffusion-based model is used instead of traditional CNN. This model relearns structural information in images using noise filtering, resulting in reconstruction errors. These errors are then used to obtain an anomaly score. Faulty areas on the rail surface are learned using reconstruction error maps. The model was trained only on healthy rail images. In the testing phase, it was tested with faulty rail images.

The UNet2D architecture is also incorporated into the learning model, based on the Denoising Diffusion Probabilistic Model (DDPM). This architecture estimates the noise added to each noisy data point based on the time step. It attempts to gradually learn how to eliminate the noise added to healthy rail images. In the test step, when the model is fed defective images, the architecture attempts to denoise the healthy image, generating reconstruction faults. These defects are converted into anomaly scores that determine the anomalous level of the image. As shown in Figure 1, the overview of the proposed method is presented. The training parameters used in this study are as follows: the learning rate was set to 0.001, the Adam optimizer was employed, the batch size was 32, and the model was trained for 20 epochs on an NVIDIA RTX 3090 GPU.

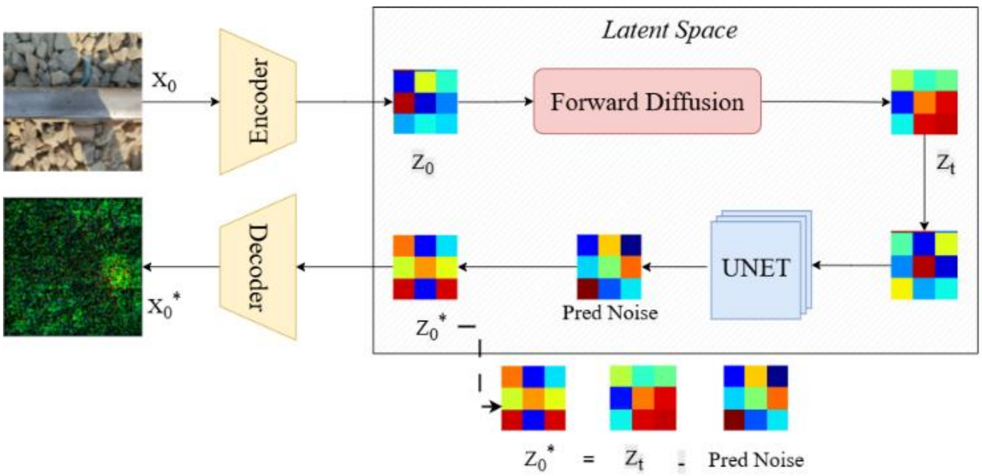


Figure 1. The overview of the proposed method

2.1. Denoising Diffusion Probabilistic Model

Classical deep learning models require large amounts of labeled data. This can make it difficult to diagnose rare fault conditions. Diffusion-based models have recently become quite popular for detecting anomalies in images. The operating principle of this model differs from the classical deep learning approach. The basic idea is that noise is first added to the intact image, approximating it to a random distribution. Then, the reverse process is applied, denoising is performed, and the noise is gradually removed to recover the original image. This feature offers a significant unsupervised learning capability for fault detection in medical imaging and industrial systems. When gradually adding noise to an image, processing should be performed on fault-free, or healthy, images; otherwise, the model may not be able to reproduce the original image during denoising. This degradation in reconstruction quality forms the basis of the system's anomaly recognition mechanism. This allows the model to identify previously unseen instances as faulty. When testing a faulty image, the model's reconstruction will be poor, producing a high anomaly score [22]. The diffusion process consists of two steps: forward diffusion and reverse diffusion. In forward diffusion, Gaussian noise is gradually added to an input image x_0 to obtain the distorted image x_t. This process is expressed as in (1).

$$q(x_t \mid x_0) = \mathcal{N}(x_t ; \sqrt{\alpha_t} x_0 , (1 - \alpha_t) I) \tag{1}$$

In (1), $\alpha_t$is the noise constant (a decreasing series) used at each time step. x_t represents the noisy image at step t, x_0 represents the original (healthy) image, and $\mathcal{N}$ represents the normal distribution. The model takes the x_t image as input for the noise level at the t value and attempts to estimate the input at this noise level. In the denoising phase, it gradually removes the noise from the x_t image to obtain x_0 . The reverse denoising process is expressed as (2). The main innovation in diffusion models lies in reversing this decay process [23]. The main goal of the model is to recover the denoised image (e.g., x_t) in forward steps and estimate its original state.

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1} ; \mu_0(x_t , t) , \textstyle\sum_\theta(t)) \tag{2}$$

The model parameters θ and the learned mean and variance are estimated in the (2). μ_0 and $\sum_\theta$ are the mean and variance functions learned by the model. The model uses a multilayered structure, such as the U-Net architecture, to make these predictions and takes both the degraded image. x_t and the time step t as inputs. For the model to be successful, it must successfully denoise towards the original structure at each time step. If a given test image does not conform to the training distribution, the model will not be able to correctly denoise it, and the resulting image will deviate significantly from the original [24]. This difference can be considered a direct anomaly. The main goal the model tries to learn during training is to minimize the difference between the predicted noise and the true noise. Therefore, the loss function used is usually the mean squared error (MSE). The model tries to estimate the noise for a given noise level. The loss function is calculated as in (3).

$$\mathcal{L} = \mathbb{E}_{x_0 \in \approx \mathcal{N}(0,1), t}\left[\left\|\epsilon - \epsilon_\theta(\sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t}\epsilon, t)\right\|^2\right] \tag{3}$$

In (3), the ϵ is the actual noise component the model is trying to learn, while $\epsilon_\theta$is the value the model predicts. In other words, the model directly learns how much and what type of noise is added at each time step. Thanks to this powerful learning strategy, the model not only learns to generate images but also to analyze them in detail [25]. After the model is trained, each image is reconstructed in the testing phase. This process compares the original image with the model's output. This difference, usually calculated on a pixel-by-pixel basis, is called the anomaly score and is calculated in (4).

$$AnomalyScore = \|x_{test} - \hat{x}_{recon}\|^2 \tag{4}$$

In (4), $\hat{x}_{recon}$ is the output the model produces from the test image. This difference is low in healthy (normal) images, while it is much higher in defective or faulty images. This allows the system to make decisions based solely on the reconstruction success, without needing labeled data. This method allows the model to automatically detect anomalies such as superficial cracks, breaks, or deformations on the rails.

3. Experimental Results

In this study, a public dataset consisting of 150 railway track images was used with two classes: Defective and Healthy. Only healthy images were used in model training, while evaluations were conducted on defective images during the testing phase. Representative samples from the dataset are presented in Figure 2. As illustrated in Figure 3, the training loss gradually decreased to 0.05, demonstrating the model's convergence. Anomaly scores obtained during the testing period were evaluated and formed using a histogram.

Figure 2. Example images of used dataset

The PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) metrics were used to evaluate the quality of the images produced as a result of the study. These metrics measure the similarity of the images to the original images after segmentation and reconstruction. The PSNR metric is used to measure the difference between an image and the reconstructed image and is expressed in decibels (dB) calculated using the MSE error as in (5). The higher this value, the closer the reconstructed image is to the original. PSNR typically ranges from 20 dB to 50 dB, with values below 30 dB indicating severe degradation and values above indicating high quality.

$$PSNR = 10.\log\left(\frac{max^2}{MSE}\right) \tag{5}$$

PSNR, because it calculates only pixel-by-pixel differences, cannot provide information about whether the structural integrity of the image is preserved. To address this issue, the SSIM metric is used, which provides information about the structural integrity of the image. This metric is calculated as in (6), taking into account the brightness $I(x,y)$, contrast $c(x,y)$, and structural similarity $s(x,y)$. The resulting value ranges from 0 to 1, with the value closer to 1 indicating a higher structural similarity in the reproduced image.

$$SSIM(x,y) = [I(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \tag{6}$$

By using these two metrics together, it examines the quality of images produced on both perceptual and pixel level images.

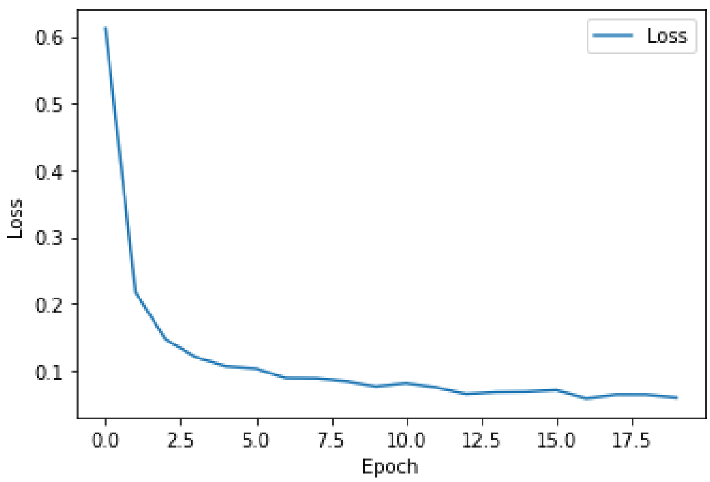


Figure 3. Loss function of the proposed method

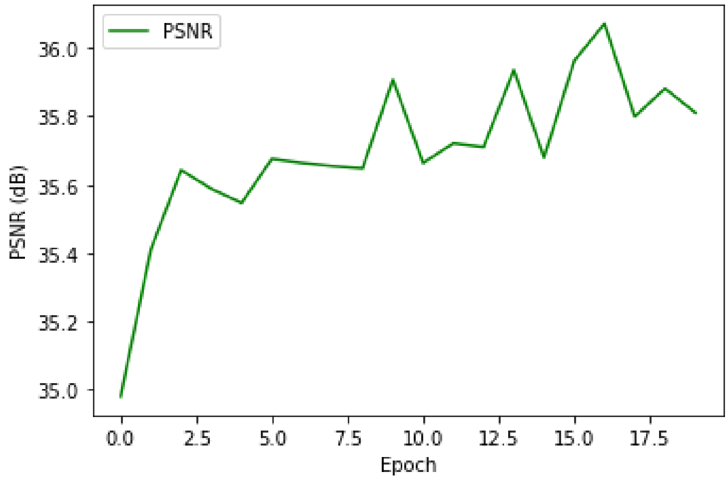


Figure 4. PRSN metric of the proposed method

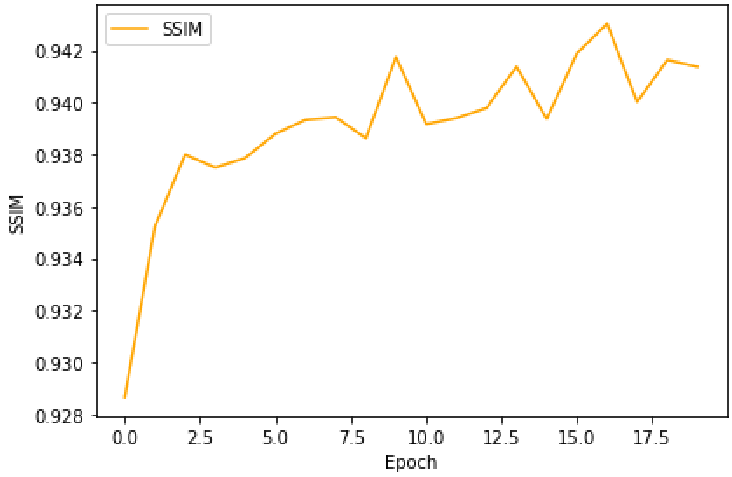


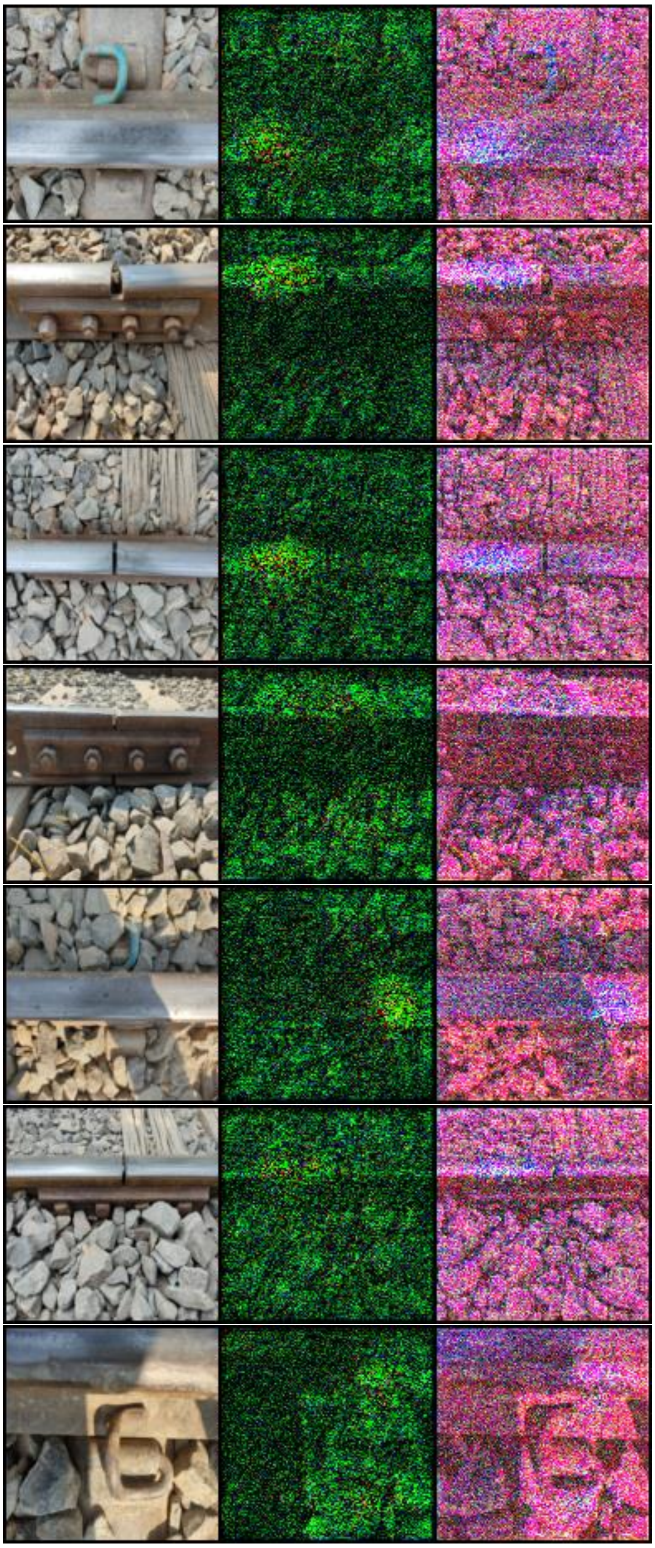
Figure 5. SSIM metric of the proposed method

The metric representing the pixel-level quality of the results produced by the proposed diffusion-based rail fault diagnosis method is PSNR, as shown in Figure 4. This value is based on the calculation of the mean square error between the original image and the synthesized image and represents the amount of noise in the image and the quantitative sensitivity on a pixel-by-pixel basis. In the literature, PSNR values are generally classified as poor quality when they fall below 20 dB, medium quality in the 20–30 dB range, and high quality at 30 dB and above. Specifically, the 30–40 dB range represents low distortion, at a level that is invisible to the human eye in most cases. The 35.7 dB value achieved in our study suggests that the proposed method can detect small defects on the rails and preserve textural integrity.

Although these results were obtained by running the proposed method for 20 epochs, it demonstrates that the diffusion model has strong generalization capacity and is prone to noise reduction. The high PSNR value allows for capturing the fine details in the detection of rail faults. This metric is important not only for visual quality but also for ensuring that the detection and classification algorithms used in subsequent stages are fed with cleaner, more information-rich input. The model was observed to converge between epochs 17 and 20, where both training and validation losses stabilized. Therefore, this interval was selected as the convergence point to prevent overfitting and ensure optimal generalization performance.

Figure 5 illustrates the SSIM value, which examines the structural integrity of the images presented. This value ranges from 0 to 1. A value of 1 indicates that the processed image is completely identical to the reference image in terms of structure, brightness, and contrast. In the literature, values of 0.90 and above correspond to situations where distortion is minimally perceptible to human perception. The

SSIM value of 0.941 obtained in the study demonstrates that the original images retain both global (general structure and form) and local (edges, texture, small defect areas) characteristics with extremely high accuracy. Diffusion-based enhancement carries the risk of losing structural details while reducing noise; however, the SSIM value of 0.941 demonstrates that the model successfully achieves this balance and produces data suitable for defect detection without compromising visual quality.



(a) Original (b) Reconstructed (c) Difference

Figure 6. Image-level anomaly representation using diffusion-based reconstruction

Figure 6 shows that the proposed method calculates the pixel differences in the original and reconstructed images, highlighting the faulty areas and visualizing the anomaly map.

4. Conclusions

The topic of anomaly detection on railway tracks has attracted considerable research attention in recent years. In this study, anomaly diagnosis of railway track faults was performed using a diffusion-based model. The proposed approach appears to have achieved significant success in anomaly detection. The proposed architecture was evaluated with loss values, PRSN, and SSIM metrics. The model's ability to perform reconstruction using only healthy images and perform anomaly testing on faulty images was evaluated. This advantage distinguishes the proposed model from other deep learning approaches by eliminating the need for large, labeled data. The model's automatic learning capabilities, the indirect reflection of attention mechanisms in anomaly score calculations, and the reconstruction-based detection mechanism offer a promising basis, particularly for real-time monitoring systems. Furthermore, providing transparent output to the user through visualization methods ensures the system's user-friendliness for field applications. This model could be a key component in the development of intelligent transportation systems.

Conflict of Interest Statement

No conflict of interest was declared by the author.

References

[1] S. Wei, B. Li, Q. Liu, and H. Zhang, "Rail surface defect detection based on deep learning," IEEE Access, vol. 8, pp. 163817–163826, 2020, doi: 10.1109/ACCESS.2020.3021964.

[2] X. Li, Y. Song, and J. Li, "A survey on railway track inspection technologies," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 4917–4932, Jun. 2022, doi: 10.1109/TITS.2021.3106117.

[3] J. Zhang, M. Wang, and C. Liu, "Automatic railway track defect detection using convolutional neural networks," Computers in Industry, vol. 129, p. 103447, 2021, doi: 10.1016/j.compind.2021.103447.

[4] Y. Sun, Y. Xu, and C. Wang, "Challenges in using deep learning for railway track maintenance," Expert Systems with Applications, vol. 159, p. 113599, 2020, doi: 10.1016/j.eswa.2020.113599.

[5] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, pp. 6840–6851, 2020.

[6] A. Wolleb, R. Sandkühler, P. K. Weyde, and P. C. Cattin, "Diffusion models for medical anomaly detection," in Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI), pp. 718–728, 2022, doi: 10.1007/978-3-031-16434-7_69.

[7] T. Berrada, Y. Gao, and A. El Saddik, "A predictive maintenance framework for railway systems based on machine learning," in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), pp. 382–387, Oct. 2019, doi: 10.1109/SMC.2019.8913924.

[8] R. Samek, T. Wiegand, and K. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," IEEE Signal Processing Magazine, vol. 36, no. 6, pp. 62–70, Nov. 2019, doi: 10.1109/MSP.2019.2919272.

[9] S. Mittal and D. Rao, "Vision based railway track monitoring using deep learning," arXiv preprint, arXiv:1711.10377, Nov. 2017.

[10] Z. Zhang, Y. Liu, X. Wang, and T. Huang, "Surface defect detection on rail tracks via deep learning," Int. J. Creative Research Thoughts (IJCRT), vol. 9, no. 6, pp. 123–130, 2021.

[11] C. Breitenberger, M. Köstler, and H. Klinke, "Automated visual inspection of railway tracks using CNNs," Int. J. Creative Research Thoughts (IJCRT), vol. 9, no. 6, pp. 215–221, 2021.

[12] R. Ferdousi, M. H. Ali, and S. Islam, "A reusable AI-enabled defect detection system for railway using ensembled CNN," in Proc. Int. Conf. Intell. Syst. Appl. (INTELLI), pp. 45–50, 2023.

[13] S. Mittal and D. Rao, "Vision based railway track monitoring using deep learning," arXiv preprint, arXiv:1711.10377, Nov. 2017.

[14] Anita, R. Venkatesh, and A. R. Ramya, "Railroad track defect detections using deep learning," Int. J. Eng. Manag. Res. (IJEMR), vol. 11, no. 1, pp. 98–102, 2021.

[15] Z. Zhang, X. Zhao, J. Chen, and Q. Yang, "Rail-5k: A real-world dataset for rail surface defects detection," arXiv preprint,

arXiv:2109.13209, Sep. 2021.

[16] J.-Y. Choi and J.-M. Han, "Deep learning (Fast R-CNN)-based evaluation of rail surface defects," Applied Sciences, vol. 14, no. 5, p. 1874, 2024, doi: 10.3390/app14051874.

[17] T. Li, Y. Fu, J. Zhang, and L. Wang, "An edge AI system based on FPGA platform for railway fault detection," arXiv preprint, arXiv:2408, Aug. 2024.

[18] A. Sharma, K. Patel, and M. R. Jones, "TrackSafe: A comparative study of data-driven techniques for railway defect detection using image datasets," Engineering Applications of Artificial Intelligence, vol. 125, p. 106487, 2023.

[19] F. Kooban, A. N. Dideban, and M. B. Mahdavifar, "Advanced technology in railway track monitoring using the GPR technique: A review," arXiv preprint, arXiv:2501, Jan. 2025.

[20] T. Wang, X. Liu, J. He, and Y. Zhou, "Intelligent railway foreign object detection: A semi-supervised convolutional autoencoder based method," arXiv preprint, arXiv:2108, Aug. 2021.

[21] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Adv. Neural Inf. Process. Syst., vol. 33, pp. 6840–6851, 2020.

[22] J. Wolleb, L. Buehler, R. Sandkühler, and P. C. Cattin, "Diffusion models for medical anomaly detection," in Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI), pp. 578–588, 2022.

[23] J. Choi, H. Seo, M. Lee, and H. Joo, "Adversarial training and diffusion models for anomaly detection," IEEE Access, vol. 10, pp. 98123–98134, 2022, doi: 10.1109/ACCESS.2022.3207678.

[24] G. Yang, X. Zhang, and G.-J. Qi, "Diffusion models for anomaly detection: A survey," arXiv preprint, arXiv:2303.15265, 2023.