Railway Track Defect Detection: A ViT based method

Yi Yao Liu, Wen Xiao He

I. Introduction

Rail transport plays a pivotal role in global logistics and infrastructure. In 2023, rail freight services accounted for approximately 45.3% of global tonne-kilometres in China and 35.8% in Russia, together representing over 80% of worldwide rail freight activity [1]. Worldwide, rail freight comprised about 16–18% of all freight tonne-kilometres as of 2020 . In North America, rail moved nearly 40% of domestic cargo by ton-kilometre, making it a cost-effective and energy-efficient mode compared to road haulage [2]. Given the heavy reliance on railways worldwide, ensuring the safety and reliability of track infrastructure is of paramount importance. Railway infrastructure remains vulnerable to track defects—such as broken rails, buckling, and cracks—which can lead to derailments, safety hazards, and significant delays. In the United Kingdom during 2023, there were 938 separate track buckling events, disrupting over 100,000 train services, cancelling roughly 10,000 trains, causing 7,000 hours of delays, and costing the network an estimated £47 million [3]. In the United States, broken rails are one of the leading causes of derailments, which occur at a rate of about three per day on average, contributing to substantial costs in equipment damage, service interruptions, and liability.

To promptly identify defects and prevent the adverse consequences caused by track defects, we need to develop solutions that enable fast and effective detection. Therefore, it is essential to develop reliable and efficient methods for early detection of track defects, aiming to enhance operational safety and minimize economic loss. Industrial anomaly detection (IAD) serves as a critical pillar of intelligent manufacturing, enabling automated quality inspection to identify defects ranging from micro-scratches to structural failures. Its applications extend to critical scenarios such as aero-engine blade maintenance, where undetected defects could trigger catastrophic failures [4]. However, real-world industrial environments are inherently complex: factors like varying illumination, viewpoint shifts, and unaligned samples pose significant challenges to the generalization of existing models [5]. Traditional methods such as feature extraction using descriptors including SIFT, SURF, HOG heavily rely on manually designed features and classic machine-learning techniques. For example, P. Viola and M. Jones developed Viola-Jones detector which leverages Haar features with integral images for rapid face detection [6]. Other methods such as Deformable Part Models(DPM) break objects into parts and model their spatial relations [7]. These approaches are limited to low-level patterns and have poor generalization to lighting, viewpoint and occlusion [8].

Deep learning which learns features automatically via neural networks has become the most popular method for IAD recently for its simplicity and better generalization abilities. Convolution Neural Networks(CNNs) based detectors like R-CNN and YOLO series achieve higher accuracy and are robust to complex environmental variations [8, 9].

J. Huang and C. Li et,al. developed a Vision Transformer(ViT) [10] based model, achieving an accuracy of 99.8% on MVTec dataset [11]. Besides that, P.Mishira and D. Fornasier developed another ViT-based model preserving the spatial information and embedded patches [12]. Z. Zhang and Z.Zhao et, al published masked multi-scale reconstruction(MMR) model achieved sample-level anomaly detection performance (AUROC%) of 84.7% on AeBAD-S dataset [13].

Despite advancements in methods such as MMR, which enhances causal reasoning among image patches, limitations remain—particularly in handling diverse defect types, cross-domain adaptability, and robustness in dynamic scenarios.

Existing IAD datasets and methods exhibit critical gaps. Traditional datasets like MVTec AD and VisA [14, 15, 16] focus on aligned samples and consistent scales, overlooking the domain shifts prevalent in real inspections. Even the AeBAD dataset, designed to address domain shifts, is limited to aero-engine blades, restricting its ability to validate models across diverse industrial objects. Methods like MMR, while superior on AeBAD, struggle with generalization when faced with unseen defect types (e.g., blowholes in magnetic tiles or rail cracks) and dynamic video scenarios. Additionally, synthetic anomaly generation (e.g., NSA) often fails in unaligned samples, leading to poor generalization.

In order to solve these problems, the focus of this study is to improve the adaptability and robustness of MMR on various types of track defects through three key innovations. Firstly, we have integrated different industrial data sets to simulate cross-domain situations, especially magnetic tile defect dataset (for ceramic defects such as blowholes and cracks) [17] and RSDDs (for track defects, such as fine cracks and penetration) [18]. In order to deal with the specific changes of data sets, we adopt a domain adaptation strategy: for each data set, we preprocess the image with a uniform resolution (224×224), while retaining the original defect features, and use contrast learning to align the cross-domain feature distribution-this ensures that the model learns the defect pattern rather than the data set-specific deviation. We further divide these data sets into training (70%) and verification (30%) sets. The training data covers all object types, and the verification data includes invisible combinations of objects and defects (for example, rail cracks in low light conditions) to test the generalization ability.

Secondly we enhance synthetic anomaly generation by leveraging DefectSpectrum [19], a flexible defect-generation model, to supplement rare or underrepresented defects. For tiny defects (e.g., grooves in blades or rail micro-cracks), we adjust the model's noise intensity and spatial distribution parameters to generate defects with sub-10px dimensions, ensuring they mimic real-world subtlety. For large structural defects (e.g., blade fractures or tile breaks), we modify the generation pipeline to introduce irregular edges and depth variations, avoiding the overly smooth synthetic anomalies that plague methods like NSA. Generated anomalies are blended into normal samples using Poisson image editing to maintain contextual consistency, and we control the ratio of synthetic-to-real data (1:3) during training to prevent overfitting to synthetic patterns.

Thirdly we optimize MMR's architecture and loss function to enhance defect sensitivity. For the ViT backbone, we replace the vanilla multi-head self-attention with a hybrid attention mechanism: local attention (3×3 window) to capture fine-grained defect details and global attention (sparse sampling) to model long-range spatial dependencies between defect and normal regions—this addresses MMR's tendency to overlook tiny defects in large images. For the loss function, we extend the original multi-scale feature alignment loss by adding a weighted term: pixels corresponding to tiny defects (annotated via GT labels from datasets like Magnetic-tile-defect-datasets) are assigned 2× higher weight, forcing the model to prioritize these regions during reconstruction. We also introduce a contrastive loss between normal and abnormal feature embeddings, encouraging the model to amplify differences between them—this reduces false negatives in cases where defects are visually similar to normal patterns (e.g., rail rust vs. dark patches).

This work makes three main contributions. The first aspect is that it expands MMR's applicability beyond aero-engine blades, validating its performance across diverse industrial objects and defect types. Moreover it proposes a multi-source data augmentation strategy that combines real datasets and synthetic defects, mitigating the challenge of limited abnormal samples. Lastly it enhances MMR's robustness to domain shifts and tiny defects, providing a more practical solution for real-world industrial inspection.

By addressing these gaps, our research advances IAD from specialized scenarios to generalizable industrial applications, aligning with the need for reliable automated quality control.

References

[1] *Global volume of rail freight transport share by country (million tonne kilometers)*, ReportLinker dataset, China and Russia led with 45.26% and 35.79% global share, respectively, 2023. [Online]. Available: https://www.reportlinker.com/dataset/2a38fd5495530d77fe6f976a504b981503cb7794.

[2] Wikipedia contributors, *Rail freight transport*, Wikipedia, The Free Encyclopedia, North America: 2 863 billion t-km (2019); China: 4 389 billion t-km; EU: 400 billion t-km, 17.1 % freight share :contentReference[oaicite:1]index=1, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Rail_freight_transport.

[3] The Sun, "Rail services hit by broken, cracked or warped tracks hits five-year high," *The Sun*, Apr. 2024, In 2023 there were 938 track buckling incidents in the UK, causing £47 million in network costs and disrupting over 100,000 trains. [Online]. Available: https://www.thesun.co.uk/news/27446840/rail-services-broken-tracks-high/.

[4] X. Wang et al., "Towards more accurate industrial anomaly detection: A component-level feature-enhancement approach," *Electronics*, vol. 14, no. 8, 2025, ISSN: 2079-9292. DOI: 10.3390/electronics14081613. [Online]. Available: https://www.mdpi.com/2079-9292/14/8/1613.

[5] Z. Zhang, Z. Zhao, X. Zhang, C. Sun, and X. Chen, "Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction," *Computers in Industry*, vol. 151, p. 103 990, 2023, ISSN: 0166-3615. DOI: https://doi.org/10.1016/j.compind.2023.103990. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361523001409.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I. DOI: 10.1109/CVPR.2001.990517.

[7] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, “Feature extraction from faces using deformable templates,” *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, Aug. 1992, ISSN: 1573-1405. DOI: 10.1007/BF00127169. [Online]. Available: https://doi.org/10.1007/BF00127169.

[8] F. Neha, D. Bhati, D. K. Shukla, and M. Amiruzzaman, *From classical techniques to convolution-based models: A review of object detection algorithms*, 2024. arXiv: 2412.05252 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2412.05252.

[9] M. Trigka and E. Dritsas, “A comprehensive survey of machine learning techniques and models for object detection,” *Sensors*, vol. 25, no. 1, 2025, ISSN: 1424-8220. DOI: 10.3390/s25010214. [Online]. Available: https://www.mdpi.com/1424-8220/25/1/214.

[10] A. Dosovitskiy et al., *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021. arXiv: 2010.11929 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2010.11929.

[11] J. Huang et al., “Self-supervised visual anomaly detection with image patch generation and comparison networks,” in *Advanced Intelligent Computing Technology and Applications*. Springer Nature Singapore, 2024, pp. 96–113, ISBN: 9789819756094. DOI: 10.1007/978-981-97-5609-4_8. [Online]. Available: http://dx.doi.org/10.1007/978-981-97-5609-4_8.

[12] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, and G. L. Foresti, “Vt-adl: A vision transformer network for image anomaly detection and localization,” in *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*, IEEE, Jun. 2021, pp. 01–06. DOI: 10.1109/isie45552.2021.9576231. [Online]. Available: http://dx.doi.org/10.1109/ISIE45552.2021.9576231.

[13] Z. Zhang, Z. Zhao, X. Zhang, C. Sun, and X. Chen, *Industrial anomaly detection with domain shift: A real-world dataset and masked multi-scale reconstruction*, 2023. arXiv: 2304.02216 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2304.02216.

[14] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, “The mvtec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection,” *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1038–1059, Apr. 2021, ISSN: 1573-1405. DOI: 10.1007/s11263-020-01400-4. [Online]. Available: https://doi.org/10.1007/s11263-020-01400-4.

[15] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, “Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9584–9592. DOI: 10.1109/CVPR.2019.00982.

[16] T. Defard, A. Setkov, A. Loesch, and R. Audigier, “Padim: A patch distribution modeling framework for anomaly detection and localization,” in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo et al., Eds., Cham: Springer International Publishing, 2021, pp. 475–489, ISBN: 978-3-030-68799-1.

[17] Y. Huang, C. Qiu, Y. Guo, and K. Yuan, “Saliency of magnetic tile surface defects,” *The Visual Computer*, 2020, Dataset available at https://github.com/abin24/Magnetic-tile-defect-datasets.

[18] Q. Wu, *Rsdds*, 2024. DOI: 10.21227/qtv6-n081. [Online]. Available: https://dx.doi.org/10.21227/qtv6-n081.

[19] S. Yang et al., *Defect spectrum: A granular look of large-scale defect datasets with rich semantics*, 2024. arXiv: 2310.17316 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2310.17316.