CETRA^{2024} 15-17 May 2024, Cavtat, Croatia
8th International Conferenceon Road and Rail Infrastructure

ENHANCING RAIL DEFECT DETECTION: COMPARATIVE ANALYSIS OF TWO-LAYER AND FIVE-LAYER CNN NETWORKS

Saeed Mohammadkhah[1], Joaquim Tinoco[1], Jesus Frias[2], José Campos e Matos[1]
[1] University of Minho, ISISE, ARISE, Department of Civil Engineering, Guimarães, Portugal
[2] University of Vigo, CINTECX, GeoTECH, Vigo, Spain

Abstract

Effective railway infrastructure maintenance is vital for ensuring the safety and reliability of transportation systems. This paper investigates the efficacy of Convolutional Neural Networks (CNNs) in detecting railway rail defects without and with data preprocessing. Utilizing a database comprising 556 images of both defective and non-defective rails, we deployed two CNN configurations: (a) a two-layer CNN; and (b) a more complex five-layer CNN. Our study encompasses the evaluation of these networks before and after data balancing and implementation of different augmentation techniques. The results consistently demonstrated superior performance from the simpler two-layer CNN across all scenarios. This observation underscores the significance of network architecture over complexity, highlighting the two-layer CNN as an effective and efficient solution for rail defect detection. The model exhibits impressive performance metrics, with an accuracy of 91.58% and an F1 score of 83.49%. The findings provide valuable insights for optimizing defect detection systems in railway maintenance applications.

Keywords: railway safety, rail defects, convolutional neural network (CNN), image processing, data balancing, pre-processing, augmentation techniques

1 Introduction

The process of visually examining railway infrastructure is labour-intensive and frequently not feasible due to limitations on security measures and a lack of adequately trained employees [1]. Hence, an increasing demand for automated solutions is evident. In the realm of transportation engineering, machine-learning techniques have been utilized in the past as effective tools for data analysis, decision-making, and performance prediction [2]. However, the recent emergence of deep learning (DL) has significantly propelled the progress of this discipline. Deep learning neural networks, such as Convolutional Neural Networks (CNNs), possess the capability to effectively handle unprocessed data and autonomously acquire the essential representations required for the tasks of recognition and classification [3].

The exploration of Convolutional Neural Networks (CNNs) in the realm of railway infrastructure maintenance has sparked significant interest among researchers aiming to enhance defect detection and classification processes. Tulbure et al. [4] and Yang et al. [5] have contributed to this field, highlighting the potential of CNNs in advancing the accuracy of multi-class recognition tasks within railway maintenance. By leveraging extensive datasets, these researchers have demonstrated substantial improvements in identifying various classes of defects, thus emphasizing the efficacy of CNNs as powerful tools in this domain.

DOI: https://doi.org/10.5592/CO/CETRA.2024.1684

Moreover, Faghih-Roohi et al. [6] have showcased remarkable achievements in classification accuracy, surpassing the 92% threshold in categorizing railway photographs without relying on traditional bounding box methodologies. This remarkable accomplishment underscores a significant departure from conventional approaches, underscoring the robustness and adaptability of CNNs in handling diverse data modalities inherent in railway infrastructure analysis.

The integration of CNNs has indeed revolutionized defect detection methodologies, especially in processing image data. By autonomously extracting pertinent features and patterns, CNNs have firmly established themselves as indispensable instruments in contemporary railway maintenance practices. Their ability to discern subtle anomalies within vast datasets not only enhances the accuracy of defect identification but also streamlines maintenance operations, thereby contributing to improved safety standards, regulatory compliance, and cost efficiency within railway systems [7-9]. This paper investigates the efficacy of Convolutional Neural Networks (CNNs) in detecting railway rail defects, both with and without data pre-processing. Through experimentation with two CNN configurations and various data augmentation techniques, we analyse their performance using a dataset comprising 556 images of defective and non-defective rails and offering insights into optimizing railway maintenance defect detection systems.

2 Data and methodology

The present study involved the design and implementation of two convolutional neural network (CNN) models for classifying railway defects. The first model consisted of two layers, while the second model consisted of five layers. In addition, two resampling approaches have been explored to handle the imbalanced issue, and augmentation technics were also implemented aiming to improve models robustness and generalization capacity. In the next subsections, more details are provided about the different components that integrates the adopted methodology.

2.1 Dataset

The dataset used in this study was derived from the railway track defect detection dataset available on Kaggle [10]. The dataset has a total of 556 images depicting instances of rail surface failures as well as images of healthy rails. These images were analysed in detail and classified into two distinct categories: defect and non-defect. To evaluate the model's performance, the K-fold cross-validation technique was utilized. This method includes repeatedly partitioning the dataset into K subsets or folds, utilizing K-1 folds for training and one fold for testing, ensuring a robust evaluation of the model's performance [10]. A sample of images from the defect and non-defect database is shown in Figure 1.

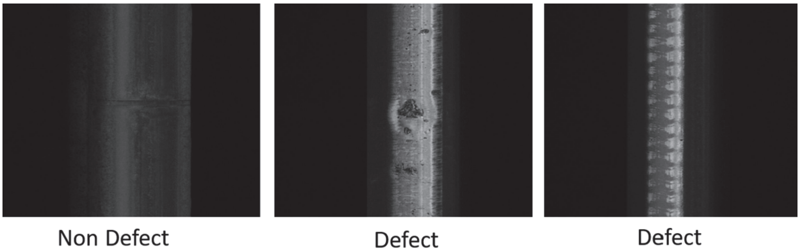

Figure 1 A sample of images of defects and non-defects

2.2 Data preprocessing

Data preprocessing includes the mitigation of dataset imbalances and the utilization of data augmentation strategies. Two methodologies were utilized to address the issue of class imbalance. The initial approach included a weighting technique [11] in which data samples were assigned weights according to the significance of their respective classes. This prioritized the training of underrepresented classes. The second option employed the Synthetic Minority Over-Sampling Technique (SMOTE) [12] to generate synthetic samples, thereby achieving a balanced distribution of classes. Furthermore, a range of data augmentation techniques were employed [12]. The employed procedures encompassed augmenting images to a predetermined size, implementing random alterations to color characteristics, randomly flipping and rotating images, applying composite modifications, converting images to grayscale, and converting enhanced images to PyTorch tensors. The study implemented six strategies to evalu-ate the influence of data augmentation and balancing approaches on classification performance. The six stratagies explored are described as following:

- Strategy 1: Training and testing on the initial database without preprocessing.
- Strategy 2: Training on the initial database with weight balance.
- Strategy 3: Training on the initial database with SMOTE balance.
- Strategy 4: Training on an unbalanced augmented database.
- Strategy 5: Training on a weighted balance augmented database.
- Strategy 6: Training on the SMOTE-enhanced database balance.

Ultimately, the outcomes of the two CNN models were thoroughly examined and assessed for each strategy.

2.3 Model architecture

The outline of the proposed framework is shown in Figures 2 and 3.

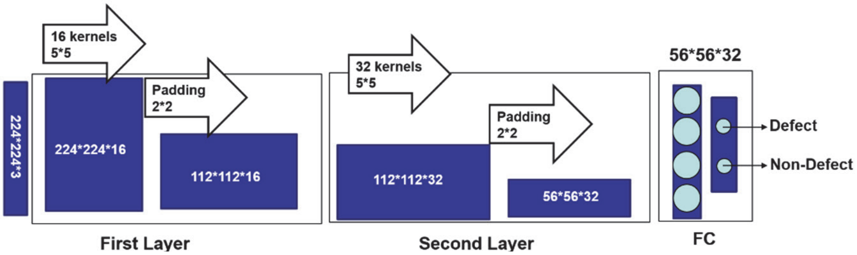

Figure 2 Two-layer CNN architecture.

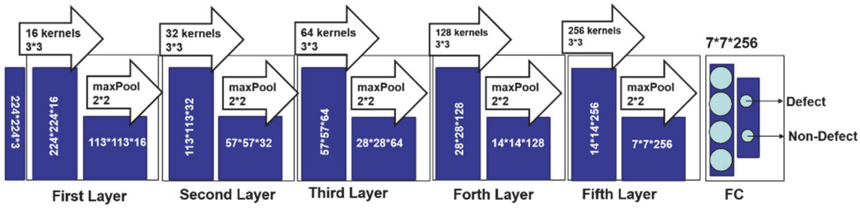

Figure 3 Five-layer CNN architecture.

2.3.1 Two-layer CNN network architecture

The first layer applies a (5, 5) kernel with padding to extract 16 feature maps, capturing low-level and mid-level features. Batch normalization, max pooling, and ReLU activation are applied after the first layer. Batch normalization stabilizes train-ing, improves convergence speed and gradient flow, while max pooling reduces spatial dimensions and preserves important information. ReLU introduces nonlinearity, enabling the model to learn complex patterns. The second layer takes the 16 feature maps from max pooling and generates 32 feature maps. An output layer is added for final classification.

2.3.2 Five-layer CNN network architecture

Similar to the two-layer CNN, the first and second layers produce 16 and 32 fea-ture maps, respectively. Additional sets of convolution, ReLU activation, batch nor-malization, and max pooling layers gradually increase the number of feature maps to 64, 128, and 256. These additional layers capture increasingly abstract and high-level features. A fully connected final layer is defined for classifying the input into predefined classes, with the number of neurons equal to the number of problem classes.

2.4 Models performance assessment

Forming the confusion matrix is one of the best methods of evaluating the models performance developed for classification tasks. By comparing the classifications from the model with the observed classes, this matrix shows the answers obtained from the model in each of the categories with four possible states: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). After determining the components of the clutter matrix, the efficiency and performance of the models in the classification and correct diagnosis of rail failures are evaluated in each strategy using prediction accuracy, precision and F1 score [11].

3 Results and discussion

The experimental results obtained from the CNN models used in the classification of rail defect images are presented. The test results, 2-layer and 5-layer CNN models are shown as confusion matrices in Table 1. According to the obtained results, which can be seen in Figure 4, the performance of two-layer and five-layer CNN models is as follows.
Strategy 1 (no augmentation and no balance): The two-layer CNN model outperformed the five-layer CNN model in terms of prediction accuracy, precision, and F1 score. This suggests that increasing the depth of the CNN architecture does not necessarily improve performance in this particular strategy. The execution time of the 5-layer model in this strategy was 314 seconds, which is 30 seconds less than the execution time of the two-layer model.
Strategy 4 (with augmentation and no balance): Introducing data augmentation resulted in different performance. The two-layer CNN generally performs better than the five-layer CNN, indicating that the simpler model is more effective in using augmented data. In this strategy, the execution time of both models was the same and equal to 1630 seconds.
Strategies 2 and 5 (with weight balancing): weight balancing improved the performance of both two-layer and five-layer CNN models. The two-layer CNN consistently achieved higher precision, accuracy, and F1 score, indicating that it is more robust when applying weight balance than unbalanced datasets. In the second strategy, the 5-layer CNN was executed about 50 seconds earlier (327 seconds) than the two-layer CNN, while in the fifth strategy, the execution time of both models is the same.

Strategies 3 and 6 (with SMOTE balancing): SMOTE balancing did not significantly improve performance for two-layer or five-layer CNN models. Both models achieved similar precision, accuracy and F1 score, indicating that SMOTE balance has a limited effect in this regard. In the third strategy, both models had almost the same execution time (500 seconds), while in the sixth strategy, the two-layer model was executed about 300 seconds earlier (1900 seconds) than the 5-layer CNN model.

Table 1 Average Confusion matrix of the models on strategies

Model 1: CNN 2 layers						
Strategies	**Strategy 1**		**Strategy 2**		**Strategy 3**	
Confusion matrix	Defect	Non-Defect	Defect	Non-Defect	Defect	Non-Defect
Defect	76.62 (68.90%)	2.98 (2.68%)	72.88 (65.54%)	6.72 (6.04%)	43.48 (27.31%)	36.12 (22.69%)
Non-defect	6.4 (5.76%)	25.2 (22.66%)	5 (4.50%)	26.6 (23.92%)	31.12 (19.55%)	48.48 (30.45%)
Model 2: CNN 5 layers						
Defect	76.28 (68.60%)	3.32 (2.99%)	76.36 (68%)	3.24 (2%)	38.24 (24.02%)	41.36 (25.98%)
Non-defect	21.36 (19.21%)	10.24 (9.21%)	15.96 (14%)	15.64 (14%)	26.28 (16.51%)	53.32 (33.49%)
Model 1: CNN 2 layers						
Strategies	**Strategy 4**		**Strategy 5**		**Strategy 6**	
Confusion matrix	Defect	Non-Defect	Defect	Non-Defect	Defect	Non-Defect
Defect	263.88 (59.33%)	54.52 (12.26%)	208.76 (46%)	109.64 (24%)	145.88 (22.91%)	172.52 (27.09%)
Non-defect	58 (13.04%)	68.4 (15.38%)	37.36 (8%)	89.04 (20%)	109.52 (17.2%)	208. 88 (32.8%)
Model 2: CNN 5 layers						
Defect	213.64 (48%)	104.76 (23%)	193.24 (43.44%)	125.16 (28.14%)	129.84 (20.36%)	188.56 (29.61%)
Non-defect	57.36 (12%)	69.04 (15%)	32.96 (7.41%)	93.44 (21.01%)	106.92 (16.79%)	211.48 (33.21%)

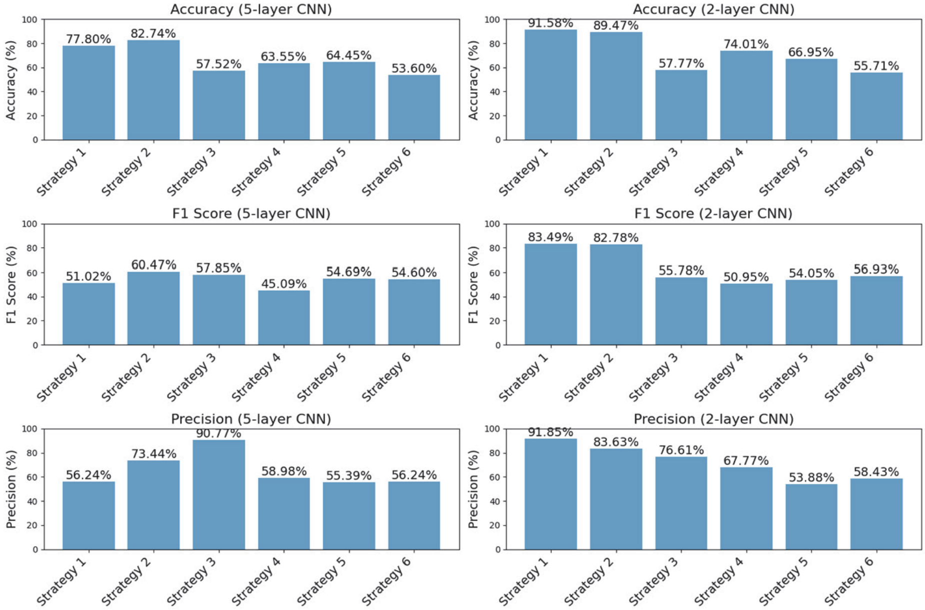

Figure 4 Models performance comparison based on precision, F1 score and accuracy metrics

4 Conclusion

The results indicate that the two-layer convolutional neural network (CNN) consistently outperforms the five-layer CNN across various methodologies. These findings offer valuable insights for selecting appropriate CNN architectures and methodologies in future phases of the research project. Specifically, they emphasize the importance of considering factors such as model complexity, data augmentation, and class balance techniques when designing CNN-based approaches for rail defect detection. By showcasing the superior efficacy and efficiency of the simpler two-layer CNN in detecting railway rail defects, this study underscores the significance of network architecture in optimizing defect detection systems for railway maintenance applications. Future work could focus on further refining and enhancing CNN-based methodologies, potentially incorporating advanced techniques or exploring alternative architectures, to continue advancing the safety and reliability of transportation systems.

Acknowledgements

This work is financed by national funds through FCT – Foundation for Science and Technology, under grant agreement 2022.13248.BD attributed to the Saeed Mohammad Khah.

References

[1] Mohamadkhah, S., Tinoco, J., Matos, J.C.: Effect of climate change on railway maintenance: a systematic review, 2021.

[2] Akhtar, M., Moridpour, S.: A Review of Traffic Congestion Prediction Using Artificial Intelligence, Journal of Advanced Transportation, 2021, 8878011

[3] SARITAS, M.M., et al.: Railway Track Fault Detection with ResNet Deep Learning Models

[4] Tulbure, A.A., Tulbure, A.A., Dulf, E.H.: A review on modern defect detection models using DCNNs – Deep convolutional neural networks, Journal of Advanced Research, 35 (2022), pp. 33-48

[5] Yang, C., et al.: Developing machine learning-based models for railway inspection, Applied Sciences, 11 (2021) 1

[6] Faghih-Roohi, S., et al.: Deep convolutional neural networks for detection of rail surface defects, International Joint Conference on Neural Networks (IJCNN), 2016.

[7] Dong, S., Wang, P., Abbas, K.: A survey on deep learning and its applications, Computer Science Review, 40 (2021), 100379

[8] Liu, H., Zhang, Y.: Deep learning based crack damage detection technique for thin plate structures using guided lamb wave signals, Smart Materials and Structures, 29 (2019) 1, 015032

[9] Wen, T., et al.: Automated pavement distress segmentation on asphalt surfaces using a deep learning network, International Journal of Pavement Engineering, pp. 1-14, 2022.

[10] Oscar van hees: Insulation Joint Training Set - ProRail, www.kaggle.com/datasets/oscarvanhees/insulation-joint-training-set-prorail

[11] Brownlee, J.: Master Machine Learning Algorithms: discover how they work and implement them from scratch, Machine Learning Mastery, 2016.

[12] Steininger, M., et al.: Density-based weighting for imbalanced regression, Machine Learning, 110 (2021), pp. 2187-2211