

Article

An Investigation of Railway Fastener Detection Using Image Processing and Augmented Deep Learning

Praneeth Chandran *, Johnny Asber, Florian Thiery, Johan Odelius and Matti Rantatalo

Division of Operation and Maintenance Engineering, Lulea University of Technology, 97187 Lulea, Sweden; john.asber@hotmail.com (J.A.); florian.thiery@ltu.se (F.T.); johanodelius@ltu.se (J.O.); Matti.Rantatalo@ltu.se (M.R.)

* Correspondence: praneeth.chandran@ltu.se; Tel.: +46-761-342-418

Abstract: The rail fastening system forms an indispensable part of the rail tracks and needs to be periodically inspected to ensure safe, reliable and sustainable rail operations. Automated visual inspection has gained significant importance for fastener inspection in recent years. Position accuracy, robustness, and practical limitations due to the complex environment are some of the major concerns associated with this method. This study investigates the combined use of image processing and deep learning algorithms for detecting missing clamps within a rail fastening system. The images used for this study was acquired during field inspections carried out along the Borlänge-Avesta line in Sweden. The image processing techniques proposed in this study enabled the improvement of the fastener position and removal of redundant information from the fastener images. In addition, image augmentation was carried out to enhance the data set, ensure experimental reliability and replicate practical challenges associated with such visual inspection. Convolutional neural network and ResNet-50 algorithms are used for classification purposes, and both the algorithms achieved over 98% accuracy during training and validation and over 94% accuracy during the test stage. Both the algorithms also maintained a good balance between the precision and recall scores during the test stage. CNN and ResNet-50 algorithms were also tested to analyse their performances when the clamp areas were covered. CNN was able to accurately predict the fastener state up to 70% of clamp area occlusion, and ResNet-50 was able to achieve accurate predictions up to 75% of clamp area occlusion.



Citation: Chandran, P.; Asber, J.; Thiery, F.; Odelius, J.; Rantatalo, M. An Investigation of Railway Fastener Detection Using Image Processing and Augmented Deep Learning. *Sustainability* **2021**, *13*, 12051.

<https://doi.org/10.3390/su132112051>

Academic Editor: Giovanni Leonardi

Received: 4 October 2021

Accepted: 28 October 2021

Published: 31 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rail transport has emerged as a significant mode of transportation as it forms a major contributing factor in the economic and industrial development of the society, through mobilization and transportation of people and commodities. Rail freight transport and passenger traffic has increased rapidly in Europe to overcome heavy congestions of road and sky, increasing energy costs, and carbon emissions. In EU15 countries, there has been an increase of 28% in passenger-kilometres and an increase of 15% in rail freight ton-kilometres, between 1990 and 2007 [1]. In Sweden, between 1960 and 2010, there has been an average annual growth of 1.1% traffic on the railway network and a further annual increase of 1% in traffic tonnage up to 2050 is anticipated [2]. The state of the existing infrastructure and the increase in volumes of freight and passenger traffic are the issues that require significant attention in the field of rail transportation [3]. Capital expansion of the infrastructure could be a possible solution to improve the rail performance, however this is a time consuming and cost-intensive approach. An ideal solution to improve the availability, capacity and service quality of the existing infrastructure would be to improve the maintenance and renewal (M&R) process. An efficient M&R operation would ensure optimization in resources, leading to smarter and more sustainable infrastructure [4].

The quality of a given railway infrastructure and its utilisation methods plays a significant role in determining its operational capacity [5]. The condition or the given state of the infrastructure and its operational capacity are highly inter-dependent and forms a crucial aspect in railway infrastructure maintenance. When the quality and condition of the railway infrastructure is high, higher operational capacity is achieved with higher service quality. With the increase in the operational capacity, the infrastructure is exposed to higher traffic and load. This increase in traffic and load leads to the deterioration of the infrastructure and deformation to its components, which results in a higher number of M&R interventions. These M&R intervention processes demand track possession, which in turn reduces operational capacity. The down time arising from such maintenance and renewal of networks is responsible for nearly half of all the delays to passengers, reducing the service quality. In Sweden, an average of 572.7 h and 670.3 h of delays are incurred due to failure of components in track and switches and crossings (S&C), respectively [6]. The track and its components need to be inspected periodically to avoid such delays due to failure, and to ensure safe and reliable operation.

To control the state of the given railway infrastructure and to prevent catastrophic accidents, track inspection has to be carried out periodically [7]. Traditionally, trained inspectors carry out the task of rail inspection by walking along the track length to look for visible defects and technical deviations. This mode of manual inspection poses safety issues for maintenance staff, and are slow, labour intensive and prone to human errors, especially in tough winter conditions. Further, such manual inspections are time consuming and expensive for railroad companies, especially for long-term and large-scale development projects. Recent technological enhancements have seen automated rail inspection systems based on machine vision being widely used for inspection of the track and its components. Moving towards autonomous visual inspection will facilitate a reduction in resource consumption arising due to manual labour, thus making the railway sector more sustainable. Such automated rail inspection systems consists of various functions such as rail profile measurement, rail surface defect detection, gauge measurements and rail fastener detection [8]. Rail fastening systems are pivotal components in the rail infrastructure as they clamp the rail to the sleepers, preventing the transverse and longitudinal deviations of rails from the sleeper. They also aid in maintaining the gauge and preserving the designed track geometry. Failures of fasteners can cause an increase in wheel flange wear, reduce the safety of train operations, and may lead to catastrophic accidents [9]. In the last two decades, application of automated machine vision systems for fastener inspection has gained significant importance; however, the detection methods from these rail images have varied over time.

Image processing and deep learning-based methods are the two widely employed detection methods for fastener defect detection [10]. The image processing-based method has the three following aspects: (1) locating and segmenting the fastener region; (2) extracting fastener features; (3) using classification algorithm for fastener defect recognition. In 2007, Marino et al. [11] made use of a multilayer perceptron neuron classifier for detecting missing hexagonal-headed bolts. Stella et al. [12] used wavelet transform and principal component analysis for fastener image pre-processing and employed a neural classifier to detect missing hook-shaped fasteners. Yang et al. [13] used the direction field as a template to match the fastener images and obtained the weight coefficient matrix by employing linear discriminant analysis (LDA) for matching. Ruvo et al. [14] used an error back propagation algorithm on rail images to model two types of fasteners and implemented the same on a graphical processing unit to achieve real time performance. The AdaBoost algorithm was used by Xia et al. [15] for detecting fasteners from rail images. Li et al. [16] used image processing techniques to detect fasteners and their various components from images acquired during visual inspections. To model fasteners and learn from the probabilistic representation of different components in rail images, H.Feng et al. [8] used the structure topic model (STM) on the acquired rail images. To differentiate between normal fasteners and broken fasteners, H.Fan et al. [17] used line local binary pattern (LLBP) on the rail

images. Edge detection methods [18], support vector machines (SVM) [19] and Gabor filters [20] are other commonly used techniques to detect fasteners from rail images. These traditional methods facilitate the inspection of fasteners with reduced equipment resources and manpower; however, the detection accuracy could easily stagnate as it becomes difficult to manually design accurate and robust features for such railway components due to the diversity of shapes and backgrounds [21]. In recent years, the application of deep learning methods [22–25] for fastener detection has gained significant importance due to the increase in computing power and development of the graphical processing units (GPU).

Significant progress has been made in detecting fasteners and identifying the defects from railway track images, however there are some underlying concerns associated with this method. The position accuracy and robustness are the two major concerns associated with this detection method [23]. The practical implementation of this technique is relatively expensive to carry out, and they are difficult to mount and maintain on an in-service train as they are integrated in the operation and are subjected to vibrations, brightness fluctuations and motion blurring during high-speed travel, which can reduce the accuracy of detection. The detection task also becomes complicated when the rail and fasteners are obscured due to the presence of dust and rust. Visual inspection can also be hindered by the presence of snow, stones (ballast occlusion) and other debris, or during heavy rain, minimising the efficiency in detecting the rail and its components. Considering these problems, a new method combining the image processing technique and deep learning technique is investigated in this article for missing clamp detection. The image processing steps aid in improving positioning of the fastener area and removing extra content from the raw images. Deep learning algorithms such as Convolutional neural network (CNN) and deep residual network (ResNet-50) are investigated for classification purposes. In addition, image augmentation techniques are implemented to investigate the performance of the detection algorithm, reciprocating the various practical limitations mentioned above. The remainder of the paper is structured as follows. Section 2 elaborates the research methodology used for this study. The results and analysis are explained in Section 3, and the conclusions are discussed in Section 4.

2. Research Methodology

The most common observed fault within a rail fastening system is missing clamps. The clamping force holding the rail on the sleeper is reduced when a clamp is missing from the fastening system. When clamps are missing from fastening system in consecutive sleepers, the track integrity is affected, as it may lead to slipping, excessive gage widening and low lateral resistance, which can further lead to derailment. This study makes use of Image processing techniques to pre-process the rail image captured during track inspection and feed them as an input to deep learning algorithms for detecting missing clamps within a rail fastening system. This study makes use of a standard laptop (Dell Ultrabook) with Matlab (R2019b), Python 3.6 (with necessary packages such as Numpy, Pandas, and Keras) and Jupyter Notebook.

2.1. Raw Data

The rail images were collected along the Borlänge-Avesta line in Sweden using a greyscale CMOS line camera (see Figure 1). Each line is triggered by a wheel encoder in 0.4 mm intervals at 20 km/h. Two thousand such lines are combined into one image and compressed with JPEG. The raw images obtained are RGB images with a resolution of 2000×2048 pixels. Due to the complex environment of the railway network and the vibration of the measurement vehicle, the rail images collected from the field are prone to noise and asymmetrical illumination. Even though sleepers should be mounted at an equal distance, in practice the distance can vary between different lines or slightly within the same line. Thus, the positioning of the sleeper and fastening system within the collected images can vary during the image segmentation procedure. Hence, it becomes necessary to incorporate image processing techniques to tackle additional noises and error

in positioning, which can deteriorate the performance of the detection algorithm. Figure 2 depicts few examples of the typical problems associated with the raw images collected from the field. The presence of such half sleepers and more than one sleeper in segmented images can deteriorate the detection accuracy.



Figure 1. Image acquisition system mounted on an inspection car.

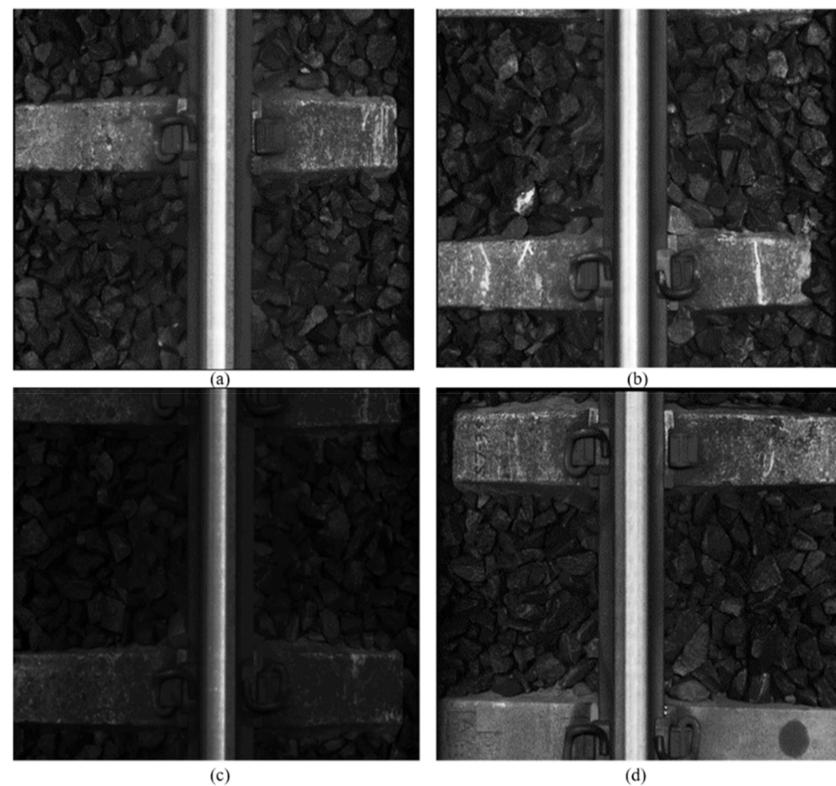


Figure 2. Raw images from the measurement car; (a) position of rail and fasteners not centred, (b) position of sleeper with fasteners towards the bottom with presence of additional noise, (c) two sleepers with four intact fasteners with low illumination, and (d) two sleepers with one clamp missing in the top sleeper.

2.2. Image Processing

To improve the detection accuracy and reduce computational cost, the fasteners need to be positioned precisely within the images. Taking the characteristics of the raw railway track line images and the positional relationship between the track and fasteners, the following image processing steps were adopted to reduce the fastener positioning error:

- The raw images were merged to form a concatenated long image of the railway track line as depicted in Figure 3a. The image is converted to gray-scale image.
- The concatenated gray-scale image was converted to a binary image (binary matrix) by using the adaptive threshold algorithm (Otsu's method) [26], to simplify the positioning process of the targeted area (refer to Figure 3b).
- The binary image was filtered using adaptive noise removal filtering to de-noise the image and improve the accuracy of the positioning result.
- The binary matrix was summed both horizontally (along the sleeper direction) and vertically (along the rail direction) to create a column vector and a row vector. The column vector was used to position the fasteners and the row vector was used to position the rail. Moving average filters was then used on these vectors to smooth them in the same way (refer to Figure 3c).
- The filtered vectors were converted to binary vectors by thresholding it to 75% of its maximum value. The centre position of the sleepers was extracted by finding the peaks and the width of the same within the binary column vector as shown in Figure 3d. Similarly, the centre of the rail was identified from the binary row vector.
- The centre position of both sleeper and rail was used to cut the concatenated images, such that the sleeper and rail was centred within a single frame and each frame contained one sleeper with two fasteners (refer to Figure 4).

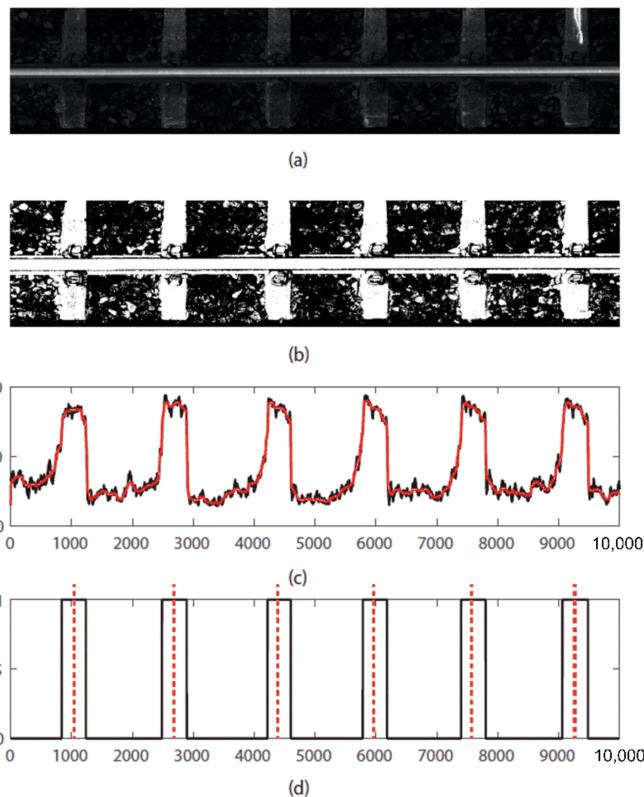


Figure 3. Image processing steps. (a) Concatenated rail images, (b) filtered binary image of the rail, (c) the result of horizontal summation of the binary matrix to a column vector marked in black colour and the smoothened curve using the moving average filter (marked in red), and (d) the binary matrix after thresholding indicating the sleeper position with the centre marked in red dotted lines.

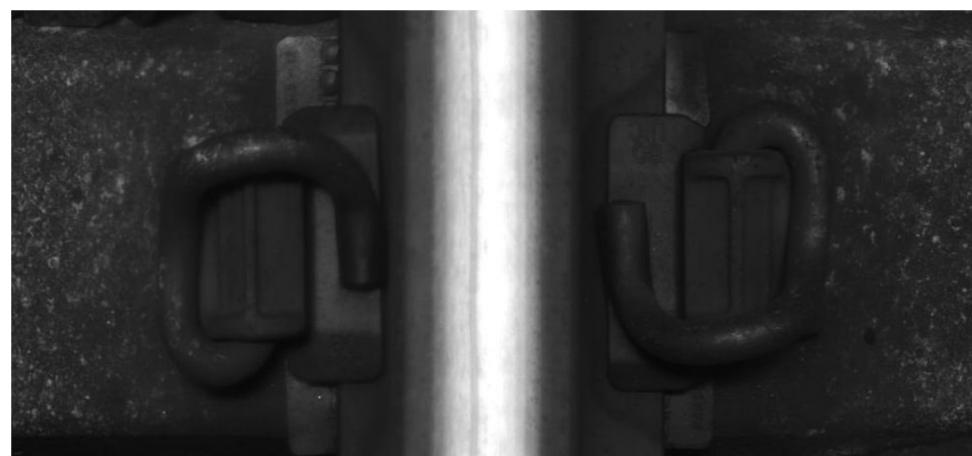


Figure 4. Fastener image after image processing.

Figure 4 depicts the fastener image acquired after the image processing techniques applied on the raw image. The image processing steps aid the pre-processing of the raw images and helps to remove redundant information from the raw images and improve the fastener positions. The image processing method presented in this study can also be used to extract individual fasteners from images captured during high-speed travel.

2.3. Deep Learning

The fastener detection task investigated in this study is a multi-class classification problem with three classes, i.e., healthy fastening system with both clamps intact, fastening system with one clamp missing, and fastening system with both clamps missing. Deep learning (DL) algorithms for image recognition and detection have gained significant importance in the last decade as these are designed in such a way that they try to replicate the function of the human cerebral cortex [27]. For this study, a convolutional neural network (CNN) and residual network (ResNet-50) are used for the classification task.

2.3.1. CNN

Convolutional neural networks are deep learning algorithms that convolve the input images with filters or kernels to extract features [27]. Subsampling, weight sharing and local field are three main traits of a CNN that allows it to minimise the trainable parameters as compared to a traditional artificial neural network. These traits also aid them to decrease overfitting and achieve shift invariance property, thus increasing the model robustness [28]. A comprehensive description regarding CNN can be found in [29,30].

The CNN architecture for the classification task used in this study is depicted in Figure 5. Three convolutional layers are used for the fastener classification purpose. The first layer consisted of 32— 3×3 filters, the second layer was composed of 64— 3×3 filters and the third convolutional layer had 128— 3×3 filters. The pooling and fully connected layers follow the convolutional layer, and a dropout (value of 0.25) were added after each convolutional layer. Pooling was used to simplify the output after convolution. Max pooling layer of size 2×2 was used for this study. The strides for the convolutional layer was 1 and for max-pooling was 2. Strides defines the number of blocks to move forward after each calculation.

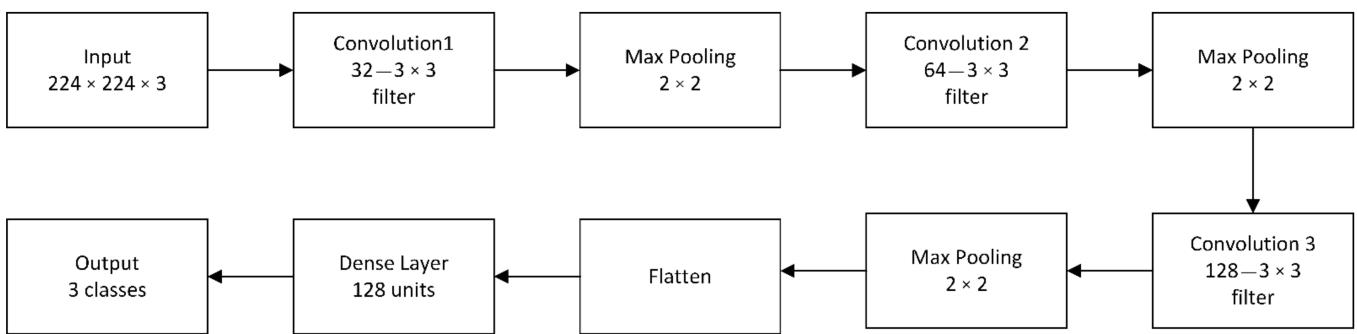


Figure 5. CNN architecture for fastener detection.

2.3.2. ResNet-50

ResNet-50 is a very deep network that stacks building blocks of the same connecting shape called the residual units [31]. Compared to other deep networks, ResNet employs a shortcut or skip connection that allows the gradients to be back propagated directly to previous layers, thus protecting the network from the vanishing gradient problem. ResNet mitigates the problem of covariate shifts by using batch normalization at its core, thus adjusting the input layer to increase the performance of the network. A comprehensive description about ResNet-50 can be found in [31,32].

Depending on the input/output dimensions, there are two main types of blocks used in ResNet, the convolution block and the identity block (ID). An identity block is used where the input activation has the same dimension as the output activation, whereas the convolutional block is used to resize the input to another dimension, such that the dimensions are equal in the final addition. Both the convolutional and identity block are used in this ResNet-50 model, and both these blocks employ skipping over three hidden layers (of kernel sizes 1×1 , 3×3 and 1×1 respectively) rather than the traditional two layers. Within each of these blocks, the shortcut and the input are added together and a ReLU activation function was added afterwards. The ResNet-50 model employed for this study had five stages, as depicted in Figure 6.

The input was zero padded (size $(3, 3)$) before moving on to the first stage. In stage 1, a convolution layer with 64 filters (shape 7×7) and with a stride of 2 was used. Further, batch normalisation and a max-pooling layer (size $(3, 3)$) with a stride of 2 was applied. In stage two, the convolution block used three sets of filters of sizes 64, 64 and 256 respectively for the three layers and a stride of 1. Stage 2 employed two identity blocks with three sets of filters of size 64, 64, and 256 for the three layers within each block. Stage 2 thus has nine layers associated with it. Similarly, stage 3 used three sets of filters of size 128, 128, and 512 respectively for the three layers within both the convolutional block (stride value of 2) and the identity block. Stage 3 had one convolutional block and three identity blocks, thus having 12 layers altogether. Stage 4 adopted three sets of filters of size 256, 256, and 1024 respectively for the three layers within both the convolutional block (stride value of 2) and the identity block. Five such identity blocks were used in stage 4, along with one convolutional block, thus having 18 layers associated with it. Stage 5 employed three sets of filters of size 512, 512, and 2048 respectively for the three layers within both the convolutional block (stride value of 2) and the identity block. The last stage had one convolution block and two identity blocks, thus making it nine layers within stage 5. Average pooling and flattening were used at the end before passing on to a dense layer (fully connected). The dense layer reduces its input to the number of classes using the softmax activation function.

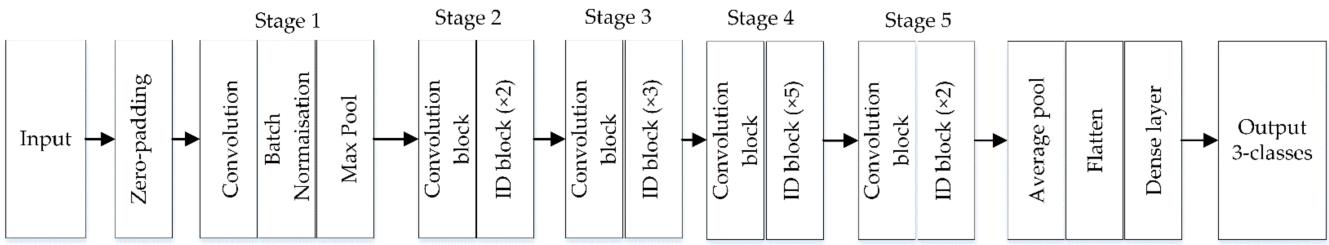


Figure 6. ResNet-50 architecture for fastener classification.

2.3.3. Training, Validation and Testing

After image processing and positioning of the track and fastener within the image, three types of fastener images are retrieved from the images captured during the track inspection, i.e., healthy fastening system with both clamps intact, fastening system with one clamp missing, and fastening system with both clamps missing. The data set for this study was collected during a real time track inspection carried out along the southern part of Sweden. The data set contained over 6000 instances of healthy clamps, 116 instances of fastening systems with one missing clamp, and 47 instances of fastening system with both clamps missing. The data set was imbalanced as the number of healthy fasteners were much higher than those with one and both clamps missing, as this is the expected behaviour of an operational (in-traffic) track section. It is more challenging to detect railway components with a limited dataset under diverse conditions, than to detect components from a large dataset under identical conditions [22]. Image augmentation was implemented to expand the data set and to ensure experimental reliability. The augmentation was carried out on all three classes. The augmentation techniques employed for this study made use of only those that are practically possible during real time measurements. Brightness, contrast, saturation, blur, noise, and rotation of the images are the parameters that were used for augmentation (refer Figure 7), and these parameters were selected based on expert opinions from the field. Presence of snow on the fastener and ballast occlusion (stones covering the fasteners) are frequently encountered problems in railway fastener inspection. Hence, the data set used for this study also made use of instances where the fasteners were partially or fully covered by ballast and snow. The augmentation techniques used for this study aims to incorporate realistic practical variations that can occur during high-speed visual inspection and not just enhance the data set with non-practical parameters.

The final data set contains 3000 images (1000 images for each class). Each class in the final data set contains instances of both the actual, as well as the augmented, images. The input images for both the models were resized to 224×224 in RGB form. The data set made use of 2550 images for training and validation (2040 samples for training and 510 samples for validation) and 450 images for testing. The parameters of the designed CNN and ResNet-50 algorithms were updated through the Adam stochastic optimization algorithm (with a learning rate of 0.01) to minimize the loss function. Cross-entropy (sparse categorical cross entropy), which estimates the divergence between the distribution of the network output and the ground truth, was considered as the loss function for this study.

Performance indicators are used to understand and evaluate how effective the model is. Different evaluation metrics underline different aspects of the performance of the classification algorithm. The classification approach used in this study is a multiclass classification model. The models were evaluated based on the performance indicators such as accuracy and cross-entropy (loss) during the training and validation stages. Indicators such as precision and recall were investigated during the test stage, along with accuracy and loss.

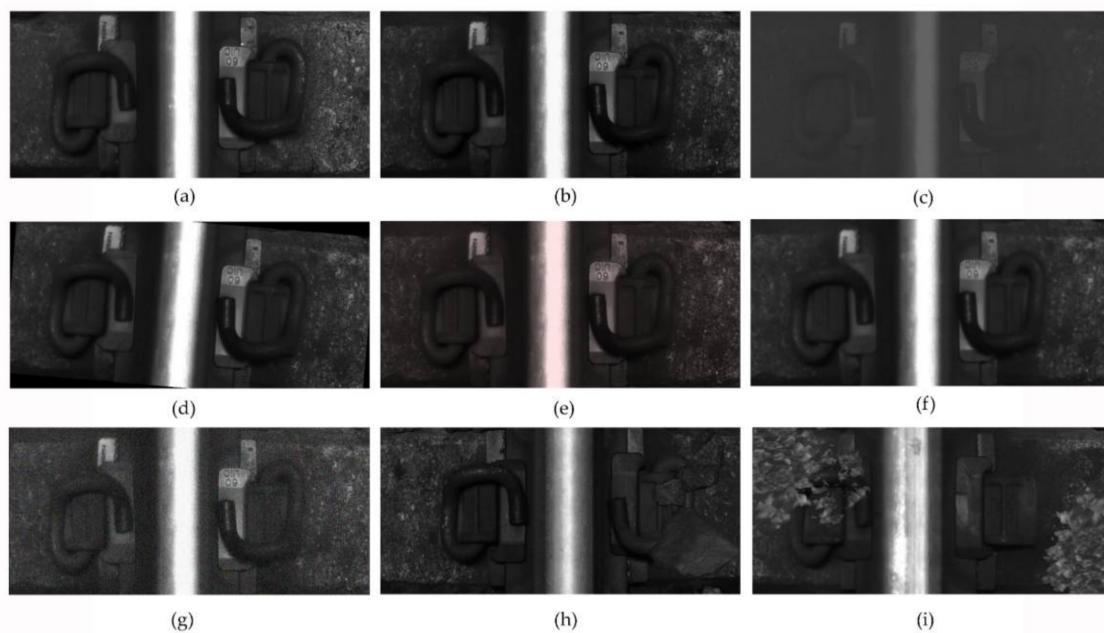


Figure 7. Augmentation of the fastener images; (a) original, (b) brightness, (c) contrast, (d) rotation, (e) saturation, (f) blur, (g) noise, (h) fasteners partially covered with stones, and (i) fasteners partially covered under ice/snow.

3. Results

The training and validation performance of the two deep learning algorithms are presented in Table 1. The CNN employed 20 epochs and ResNet had 50 epochs during the training phase. Epochs during training represents the number of passes of the entire training data set the algorithm has completed. The number of epochs was determined based on the lowest validation loss achieved and when the difference between the training and validation loss was found to be the least. The batch size was 50 for both the algorithms. Both CNN and ResNet-50 exhibited a high accuracy during training and validation, of over 98%. ResNet-50 exhibited the highest accuracy among the two algorithms considered, during both training (99.02%) and validation (98.24). The loss for both the algorithms was well below 0.05. The loss was found to be the least in ResNet-50 during both the training and validation phase, with values of 0.0086 and 0.0205, respectively. The average training time per epoch was lower for CNN (127 s) compared to ResNet-50 (1029 s). This is due to the larger network structure and higher number of trainable parameters observed in the ResNet-50 algorithm. The average training time per sample for the CNN model was 63 milliseconds, and for the ResNet-50 model was 509 milliseconds. The average training time per sample was low for CNN (63 S) compared to the ResNet-50 (509 s). There were no huge variations in accuracy and loss during both training and validation for both algorithms, indicating that the models did not over-fit or under-fit the data. Figure 8 depicts the learning curves for both the algorithms with respect to the number of epochs. The learning curves for both accuracy and loss are depicted in the same way for the two algorithms. For both the deep learning algorithms, the validation scores (both accuracy and loss) tend to converge to a value close to the training score, indicating low bias and low variance. Since the training accuracy was high with low loss, the training data was well fitted by both the models, indicating a low bias. Furthermore, the gap between the training and validation curves for both the algorithms was nominal, indicating a low variance.

Table 1. Performance of the algorithms during training and validation.

Parameters	CNN	ResNet-50
Training Accuracy (%)	98.73	99.02
Training Loss	0.015	0.0086
Validation Accuracy (%)	98.18	98.24
Validation Loss	0.024	0.0205
Total Parameters	11,169,347	23,643,011
Trainable Parameters	11,169,347	23,589,891
Number of Epochs	20	50
Average Training Time per Epoch (seconds)	127	1039
Average Training Time per Sample (milliseconds)	63	509

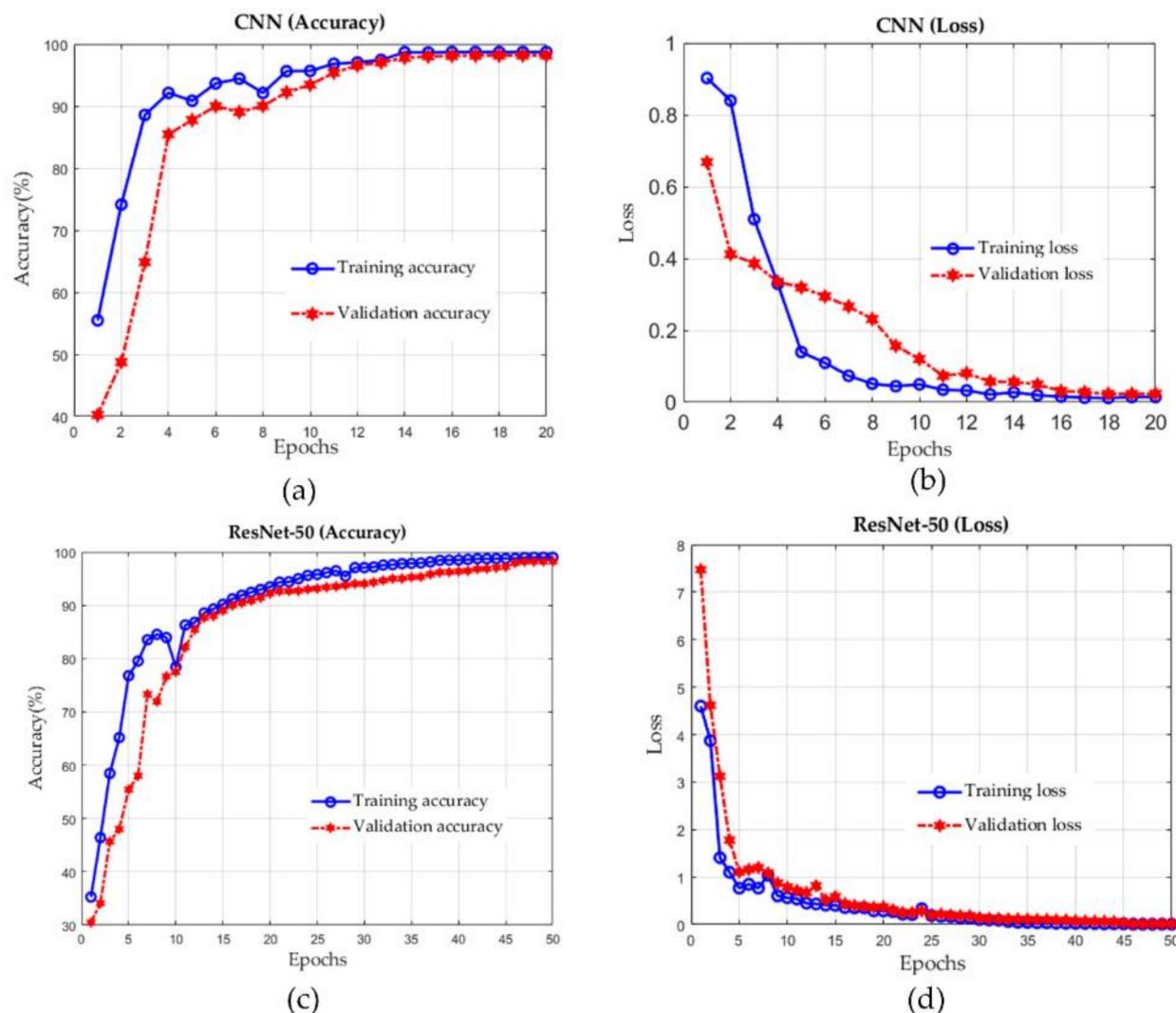
**Figure 8.** Learning curves with respect to the number of epochs and performance indicator; (a) CNN epochs vs. accuracy, (b) CNN epochs vs. loss, (c) ResNet-50 epochs vs. accuracy, and (d) ResNet-50 epochs vs. loss.

Table 2 depicts the performance of CNN and ResNet-50 on the testing set. In the test set, 450 samples were used, which contained instances that correlated with the complex situations occurring during vision-based inspections. CNN and ResNet-50 both exhibited relatively high accuracy of 94% and 94.4% respectively, even under such circumstances. The loss for CNN during testing was slightly higher than ResNet-50 during the testing phase. CNN had a loss of 0.56, whereas ResNet-50 had a loss score of 0.47. The time taken for the prediction of a sample is lower for both the CNN and ResNet-50 models. CNN

was able to make predictions on the test set in 8 s and ResNet-50 model took 87 s. Both the models exhibited a good balance between the precision and recall score. The precision score on the test set was high for both the ResNet-50 model (95%) and the CNN model (94%). The recall score on the test set was similar for both the algorithms (94%). In railway application, it is essential to have both high precision and recall in order to balance the risk of failure and the cost of inspection. A higher precision minimizes the false positive rates, thus contributing to better detection of the fastener state, ensuring safe and reliable operation of the railway. A higher recall ensures minimal false negatives, thus minimizing cost due to unwanted inspection.

Table 2. Performance of the algorithms during testing.

Parameters	CNN	ResNet-50
Accuracy (%)	94	94.4
Loss	0.56	0.47
Precision	94	95
Recall	94	94
Time for testing (seconds)	8	87
Time per sample (milliseconds)	16	190

Out of 450 samples used for testing, CNN misclassified 27 samples and ResNet-50 misclassified 25 samples. Some of the misclassified fastener images are depicted in Figure 9. Among the six instances depicted in Figure 9, five instances (refer Figure 9a–e) were misclassified by both the algorithms in similar manner. The instance depicted in Figure 9f was correctly predicted by the ResNet-50, but was wrongly predicted by the CNN algorithm. Both the algorithms performed significantly well when the fasteners in the test images were rotated, had synthetic noise added to them, when the saturation level of the images varied, and when the fastener images were blurred. Both CNN and ResNet-50 models performed well in detecting the fastener state when the fasteners were partially covered with stones and snow. However, for both the deep learning algorithms, the majority of the false predictions occurred when the fasteners in the images were obscured heavily under snow or stones. The algorithms also had difficulties in predicting the right class when the illumination level was poor (low level of brightness).

To further understand the performance of the deep learning algorithms on predicting the fastener state where the clamps were covered, an additional test set was created as depicted in Figure 10. A black box was used to cover the clamp area on one side of a healthy fastening system and was tested for both the algorithms. The clamp area was covered incrementally in steps of 5% of the total clamp area. A total of 30 images were created in the new test set, such that 10 images had no clamp area covered and the remaining 20 images had clamp areas covered (5% to 100% of the clamp areas). The CNN algorithm was able to predict the detect fastener state precisely up to 70% of clamp area occlusion. The algorithm misclassified the fastener state to one missing for those fasteners where the clamp areas were covered by 75% of the total fastener areas. The ResNet-50 algorithm was slightly better when compared to the CNN in this regard, as it was able to detect accurately up to 75% of clamp area occlusion. The prediction, however, was not accurate when the clamp area was occluded from 80% and above.

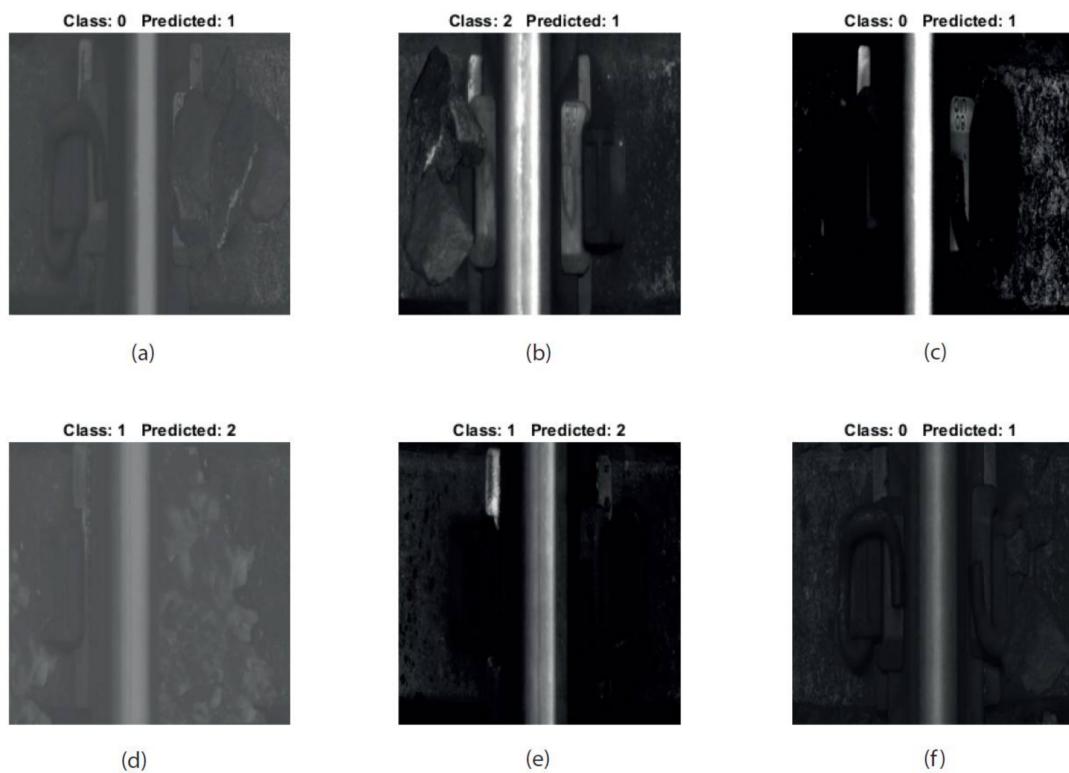


Figure 9. Misclassified fastener images. Labels 0, 1, and 2 represents healthy state, one missing clamp within a fastening system, and two missing clamps within a fastening system, respectively. (a) True class ‘healthy’ predicted ‘one missing’, (b) true class ‘two missing’ predicted ‘one missing’, (c) true class ‘healthy’ predicted ‘one missing’, (d) true class ‘one missing’ predicted ‘two missing’, (e) true class ‘one missing’ predicted ‘two missing’, and (f) true class ‘healthy’ predicted ‘one missing’.

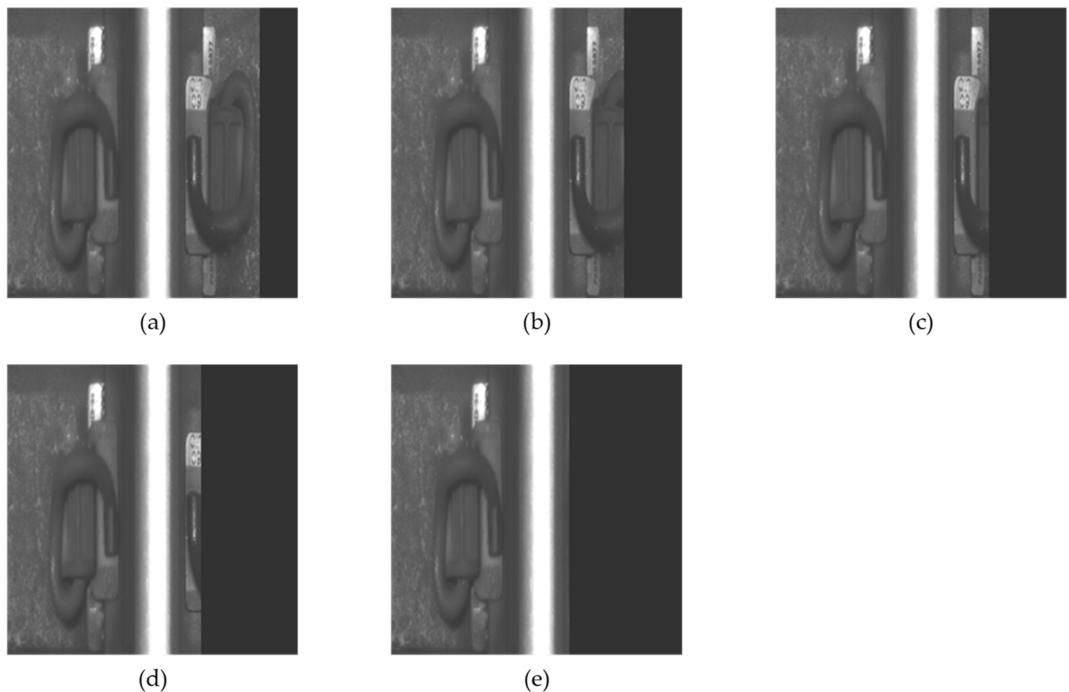


Figure 10. Additional test set to analyse the performance of the deep learning algorithms when the clamp areas were covered; (a) 0% clamp area covered, (b) 25% clamp area covered, (c) 50% clamp area covered, (d) 75% clamp area covered, and (e) 100% clamp area covered.

4. Conclusions and Future Work

In recent years, with the development of high-speed railways, automated fastener detection technologies based on machine vision have gained significant attention. Automated visual inspection makes use of rail images for detecting fasteners. The positional accuracy, complex railway environment, practical implementation and robustness are the major concerns associated with this method for fastener detection. This article aims to investigate a method combining image processing and deep learning algorithms for detecting missing clamps within fastening systems. The images used for this study were obtained during the field inspection along the Borlänge-Avesta line in Sweden. The image processing technique was successfully able to improve the positional accuracy of the fastener and rail, while removing the redundant information from the rail images. Data augmentation was then carried out to replicate the complex scenarios associated with the visual inspections. Two deep learning algorithms, namely CNN and ResNet-50 models, are investigated for detecting missing clamps from the rail images.

The results of the study show that combining image processing with the deep learning algorithms was effective in achieving high accuracy for fastener detection. The training and validation accuracies for both CNN and ResNet-50 models were above 98% with minimal loss. The training and testing time per epoch was found to be lesser for CNN than the ResNet-50 algorithm, due to the larger network structure and a higher amount of trainable parameters in ResNet-50. The training and testing time per sample was found to be lower for CNN among the two algorithms. Both the algorithms were able to achieve over 94% accuracy in detecting fasteners from different complex environments during the testing phase. The models were reliable when the fasteners were rotated, had additional noise, when the images were blurred, when the saturation level varied, and when the fasteners were covered partially under snow or ballast. The two models, however, had difficulties in predicting fastener state when the brightness was affected and when the fasteners were heavily occluded by the presence of ballast and snow. An additional test set was created by covering the clamp areas to further analyse the detection capabilities of both the algorithms, when fasteners were covered in the images. The CNN failed to predict the fastener state for all scenarios when the occlusion covered over 70% of the clamp area. The ResNet-50 algorithm failed to predict the fastener state when the occlusion was above 75% of the clamp area. Further studies need to be carried out to analyse the prediction capabilities of the algorithms where different scenarios of occlusion along the clamp area are considered to estimate the triggering mechanism of such an algorithm. This study will be carried out in the future research. On comparing both the algorithms, the complexity and time required for training and testing was lesser for the CNN algorithm, which can add value for real time application. The performance in terms of accuracy and precision was marginally better for the ResNet-50 when compared to the CNN algorithm. The ResNet-50 algorithm was able to detect fastener state slightly better than the CNN, even when the fasteners were occluded by about 75% of the total fastener area. A better detection of fastener state will ensure less disruptions or downtime arising from M&R, leading to safe, reliable and sustainable rail transportation.

In Sweden, the tracks are covered with snow for the majority of the year and would thus require additional rail surface treatment or a removal process that adds to the expenses of the railroad companies. One possible solution to overcome this difficulty and ensure safe, sustainable and reliable rail operation, is by combining automated visual inspection with non-destructive testing such as eddy current sensors [6,33] for fastener inspection. The presence of non-conductive materials (such as ballast and snow) in the sensor-to-target gap do not affect the eddy current sensors. This allows their use in complex environments, such as those involving stones, water, oil, machine fluids and snow. The differential eddy current inspection was able to detect the fastener state with a precision and recall of 96.64% and 95.52%, respectively [6]. The results presented in the previous studies were based on controlled measurements carried out along the heavy haul line in the northern part of Sweden, where the likelihood of disturbances were minimal. The measurements were

carried out by mounting the sensor system 65 mm above the railhead on a trolley system. A detailed comparative study between the eddy current inspections (measurements from an actual train) and machine vision-based inspection for detecting missing clamps from a fastening system for the same track section will be carried out in the future studies.

The future of this study will also focus on detecting different types of rail fastening systems and other track components simultaneously from rail images. Further studies will be carried out to investigate the performance of the deep learning algorithms for various scenarios of occlusion. The possibility of using pre-trained weights and transfer learning for fastener state detection will be investigated in the future study. Future research will aim to combine automated visual inspection with eddy current inspection for improving the fastener detection to ensure safe, reliable and sustainable rail transportation.

Author Contributions: Conceptualization, P.C., F.T., J.O. and M.R.; methodology, P.C., J.A., F.T. and M.R.; software, P.C., F.T., J.A. and J.O.; validation, P.C. and F.T.; formal analysis, P.C., F.T., J.A. and M.R.; investigation, P.C., J.A. and F.T.; resources, P.C., J.O. and M.R.; data curation, P.C., J.A. and F.T.; writing—original draft preparation, P.C.; writing—review and editing, P.C., F.T., J.A., J.O. and M.R.; visualization, P.C. and F.T.; supervision, M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This research was supported and funded by the Luleå Railway Research Centre (JVTc), and Trafikverket (Swedish Transport Administration). The work has been carried out within the strategic innovation program InfraSweden2030, supported by Vinnova, Formas and Energimyndigheten, and the Shift2Rail project IN2SMART. The authors would like to express their gratitude to Raildoc AB for the measurement support. The authors would also like to thank Jesper Engberg of Raildoc AB, Ulf Ranggard of ElOptic i Norden AB, Håkan Lind, Anders Thornemo and David Lindow of Bombardier Transportation Sweden, and Olavi Kumpulainen of Consisthentic AB, for their guidance in this study. The authors would also like to thank Bony Thomas from the department of Engineering, Science and Mathematics at Luleå University of Technology for his support.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Whiteing, T.; Menaz, B. Thematic Research Summary: Rail Transport. Transport Research Knowledge Centre—European Commission. 2009. Available online: http://www.eurosfaire.prd.fr/7pc/doc/1257761907_trs_rail_transport_2009.pdf (accessed on 16 February 2021).
- Trafikverket. *Prognosis of Swedish Goods Flow for the Year 2050 (Prognos Över Svenska Godsströmmar år 2050: Underlagsrapport. Publikationsnummer)*; Technical Report; Trafikverket: Borlänge, Sweden, 2012; Volume 112.
- European Commission. Roadmap to a Single European Transport Area—Towards a Competitive and Resource Efficient Transport System. 2011. Available online: https://ec.europa.eu/transport/themes/european-strategies/white-paper-2011_en (accessed on 17 February 2021).
- Castillo-Mingorance, J.M.; Sol-Sánchez, M.; Moreno-Navarro, F.; Rubio-Gámez, M.C. A critical review of sensors for the continuous monitoring of smart and sustainable railway infrastructures. *Sustainability* **2020**, *12*, 9428. [CrossRef]
- Patra, A.P.; Kumar, U.; Kraik, P.-O.L. Availability target of the railway infrastructure: An analysis. In Proceedings of the 2010 Annual Reliability and Maintainability Symposium (RAMS), Wellington, New Zealand, 2–4 December 2010; pp. 1–6. [CrossRef]
- Chandran, P.; Thierry, F.; Odelius, J.; Famurewa, S.; Lind, H.; Rantatalo, M. Supervised Machine Learning Approach for Detecting Missing Clamps in Rail Fastening System from Differential Eddy Current Measurements. *Appl. Sci.* **2021**, *11*, 4018. [CrossRef]
- Aytekin, C.; Rezaeitabar, Y.; Dogru, S.; Ulusoy, I. Railway Fastener Inspection by Real-Time Machine Vision. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *45*, 1101–1107. [CrossRef]
- Feng, H.; Jiang, Z.; Xie, F.; Yang, P.; Shi, J.; Chen, L. Automatic Fastener Classification and Defect Detection in Vision-Based Railway Inspection Systems. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 877–888. [CrossRef]

9. Dai, P.; Du, X.; Wang, S.; Gu, Z.; Ma, Y. Rail fastener automatic recognition method in complex background. In Proceedings of the Tenth International Conference on Digital Image Processing (ICDIP 2018), Shanghai, China, 11–14 May 2018; Volume 10806, p. 1080625. [[CrossRef](#)]
10. Wei, X.; Wei, D.; Suo, D.; Jia, L.; Li, Y. Multi-Target Defect Identification for Railway Track Line Based on Image Processing and Improved YOLOv3 Model. *IEEE Access* **2020**, *8*, 61973–61988. [[CrossRef](#)]
11. Marino, F.; Distante, A.; Mazzeo, P.L.; Stella, E. A Real-Time Visual Inspection System for Railway Maintenance: Automatic Hexagonal-Headed Bolts Detection. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2007**, *37*, 418–428. [[CrossRef](#)]
12. Stella, E.; Mazzeo, P.; Nitti, M.; Cicirelli, G.; Distante, A.; D’Orazio, T. Visual recognition of missing fastening elements for railroad maintenance. In Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems, Singapore, 6 September 2003; pp. 94–99. [[CrossRef](#)]
13. Yang, J.; Tao, W.; Liu, M.; Zhang, Y.; Zhang, H.; Zhao, H. An Efficient Direction Field-Based Method for the Detection of Fasteners on High-Speed Railways. *Sensors* **2011**, *11*, 7364–7381. [[CrossRef](#)] [[PubMed](#)]
14. De Ruvo, P.; Distante, A.; Stella, E.; Marino, F. A GPU-based vision system for real time detection of fastening elements in railway inspection. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 2333–2336. [[CrossRef](#)]
15. Xia, Y.; Xie, F.; Jiang, Z. Broken Railway Fastener Detection Based on Adaboost Algorithm. In Proceedings of the 2010 International Conference on Optoelectronics and Image Processing, Haikou, China, 13 January 2011; Volume 1, pp. 313–316. [[CrossRef](#)]
16. Li, Y.; Otto, C.; Haas, N.; Fujiki, Y.; Pankanti, S. Component-based track inspection using machine-vision technology. In Proceedings of the 1st ACM International Conference on Multimedia Retrieval, Trento, Italy, 18–20 April 2011; pp. 1–8. [[CrossRef](#)]
17. Fan, H.; Cosman, P.C.; Hou, Y.; Li, B. High-Speed Railway Fastener Detection Based on a Line Local Binary Pattern. *IEEE Signal Process. Lett.* **2018**, *25*, 788–792. [[CrossRef](#)]
18. Singh, M.; Singh, S.; Jaiswal, J.; Hempshall, J. Autonomous Rail Track Inspection using Vision Based System. In Proceedings of the 2006 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, Alexandria, VA, USA, 16–17 October 2006; pp. 56–59. [[CrossRef](#)]
19. Mazzeo, P.; Ancona, N.; Stella, E.; Distante, A. Visual recognition of hexagonal headed bolts by comparing ICA to wavelets. In Proceedings of the 2003 IEEE International Symposium on Intelligent Control ISIC-03, Houston, TX, USA, 8 October 2003; pp. 636–641. [[CrossRef](#)]
20. Mandriota, C.; Nitti, M.; Ancona, N.; Stella, E.; Distante, A. Filter-based feature selection for rail defect detection. *Mach. Vis. Appl.* **2004**, *15*, 179–185. [[CrossRef](#)]
21. Li, Y.; Trinh, H.; Haas, N.; Otto, C.; Pankanti, S. Rail Component Detection, Optimization, and Assessment for Automatic Rail Track Inspection. *IEEE Trans. Intell. Transp. Syst.* **2013**, *15*, 760–770. [[CrossRef](#)]
22. Wang, T.; Yang, F.; Tsui, K.-L. Real-Time Detection of Railway Track Component via One-Stage Deep Learning Networks. *Sensors* **2020**, *20*, 4325. [[CrossRef](#)] [[PubMed](#)]
23. Wei, X.; Yang, Z.; Liu, Y.; Wei, D.; Jia, L.; Li, Y. Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Eng. Appl. Artif. Intell.* **2019**, *80*, 66–81. [[CrossRef](#)]
24. Liu, J.; Huang, Y.; Zou, Q.; Tian, M.; Wang, S.; Zhao, X.; Dai, P.; Ren, S. Learning Visual Similarity for Inspecting Defective Railway Fasteners. *IEEE Sens. J.* **2019**, *19*, 6844–6857. [[CrossRef](#)]
25. Song, Q.; Guo, Y.; Yang, L.; Jiang, J.; Liu, C.; Hu, M. High-speed railway fastener detection and localization system. *arXiv* **2019**, arXiv:1907.01141.
26. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
27. Chauhan, R.; Ghanshala, K.K.; Joshi, R. Convolutional Neural Network (CNN) for Image Detection and Recognition. In Proceedings of the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 15–17 December 2018; pp. 278–282. [[CrossRef](#)]
28. Yuan, Z.; Zhu, S.; Zhai, W. Damage Detection of Rail Fastening System Through Deep Learning and Vehicle-Track Coupled Dynamics. In *Advances in Dynamics of Vehicles on Roads and Tracks*; Springer: Singapore, 2020; pp. 148–153. [[CrossRef](#)]
29. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6. [[CrossRef](#)]
30. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833. [[CrossRef](#)]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 630–645. [[CrossRef](#)]
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. Available online: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf (accessed on 2 September 2021).
33. Chandran, P.; Rantatalo, M.; Odelius, J.; Lind, H.; Famurewa, S.M. Train-based differential eddy current sensor system for rail fastener detection. *Meas. Sci. Technol.* **2019**, *30*, 125105. [[CrossRef](#)]