Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit
Volume 237, Issue 2, February 2023, Pages 137-145

https://doi.org/10.1177/09544097221093486

Original Article

Use of deep convolutional neural networks and change detection technology for railway track inspections

Ryan M Harrington [1], Arthur de O Lima [1], Richard Fox-Ivey [2], Thanh Nguyen [2], John Laurent [2], Marcus S Dersch [1], and J Riley Edwards [2]

Abstract

Railroad track inspections conducted in accordance with federal regulations and internal railway operating practices result in significant labor costs and occupy valuable network capacity. These factors, combined with advancements in the field of machine vision, have encouraged a transition from human visual inspections to machine-based alternatives. Commercial machine vision technologies for railway inspection currently exist, and automated analysis approaches—which deliver objective results—are available in some systems. However, they are limited to a "pass/fail" approach through the detection of components which fail to meet maintenance or geometry thresholds, as opposed to being able to detect subtle changes in track conditions to identify evolving problems. To overcome these limitations, this paper presents results from the field deployment and validation of a system that pairs three-dimensional (3D) machine vision with automated change detection technology. The change detection approach uses a deep convolution neural network (DCNN) to accurately characterize track conditions between repeat runs. Current automated track inspection technologies were studied, and the applicability of change detection is discussed. The paper presents the process for 3D image capture, DCNN training, and evaluation by comparing DCNN results to an expert human evaluator. Finally, it presents change detection results for fastener presence and spike height. Results indicate that this technology can successfully identify fasteners and spikes with percent accuracies greater than 98% and that it can successfully generate change detection results for comparison of track condition among runs.

Keywords

3D laser triangulation, artificial intelligence, deep convolutional neural networks, algorithms, change detection, safety, operational efficiency, inspection

[1]Rail Transportation and Engineering Center - RailTEC, Department of Civil and Environmental Engineering - University of Illinois at Urbana-Champaign, Urbana, IL, USA
[2]Pavemetrics Systems Inc., Québec, QC, Canada

Corresponding author(s):
Ryan M Harrington, Rail Transportation and Engineering Center - RailTEC, Department of Civil and Environmental Engineering - University of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801, USA. Email: rhrrngt2@illinois.edu

Introduction

Background

In the United States, the Department of Transportation (DOT) Federal Railroad Administration (FRA) regulates track inspection intervals based on track class. As track class increases, so does the inspection frequency requirements—as often as twice per week for FRA Classes 4 and 5.[1] Additionally, railroads often impose more restrictive inspection practices than the FRA's minimum requirements.

To perform these inspections, railroad staff typically perform visual inspections from hi-rail vehicles. While inspectors often possess significant knowledge and experience, their inspections are subjective and challenging given the logistics of performing detailed inspections at hi-rail speeds. These challenges drove the industry to seek the development of machine-based inspection tools which leverage digital imaging, image processing, and artificial intelligence (AI) to augment human inspections.

Several companies have developed machine vision inspection systems that focus on defect identification and the exceedance of maintenance threshold values or FRA-mandated safety threshold. Aurora ***X**iv* inspects and grades timber and concrete crossties, identifies rail base corrosion, and quantifies ballast levels by capturing light reflected off the track.[2,3] High powered xenon lights have been employed in a track imaging system to illuminate rails, where the reflected light is captured by a set of two-dimensional (2D) cameras to identify joint bar cracks.[4] Similarly, Railcheck also captures reflected light from light emitting diodes to produce 2D images which are analyzed to identify rail, concrete crosstie, fastener, and ballast level anomalies.[5] The chord-based Rail Corrugation System employs optics to identify rail corrugations through 2D laser profiling.[6] Lastly, LR*AI*L captures three-dimensional (3D) profiles and 2D images simultaneously (along with inertial data) to measure rail geometry, and inspect rail surfaces, fasteners, spikes, ballast, and crossties.[7]

The AI deployed in this research involves deep convolution neural networks (DCNNs), which are machine learning programs that use deep learning for image recognition. Over the last two decades, these programs have outperformed other pattern recognition methods given they incorporate reinforcement learning to interact with unknown environments.[8] Several studies pair DCNNs with image capture technology to identify specific features of interest along the railway. Chen et al. employed a DCNN to identify catenary system support defects.[9] Similarly, DCNNs have also been used to detect other features along the track from 2D images (e.g. rail surface defects,[10] fasteners,[11] and concrete crosstie cracks).[12,13] Despite the success of DCNN feature identification demonstrations, comparatively little research has investigated how DCNNs can be paired with other algorithms to expand their utility beyond binary "pass/fail" inspections to more sophisticated tasks such as change detection.

Change detection programs automatically identify changes in features over a given period by comparing new and preexisting datasets which were processed through image recognition software. This potential benefit led to the development of two FRA-funded change detection research programs. The first—conducted by ENSCO—comprised of a study collecting and evaluating 2D "before" and "after" images within two time intervals to identify alterations in fastener condition, crosstie condition, and rail surface in 2D track images.[14] Railmetrics conducted the second FRA research program through the execution of two project phases, with the first documenting an initial proof of concept and the second showing a further-developed prototype of a 3D imaging and scanning system.[15,16] In the first project, LR*AI*L technology was used to evaluate datasets captured two months apart on Amtrak's Northeast Corridor (NEC). The prototype system was shown to be capable of identifying changes to crosstie skew angle, fastener presence, joint gap, joint bar bolt count and ballast levels with repeatability in excess of 95%.[16] The success of the LRAIL prototype led the FRA to sponsor further development of this technology.

Objective and scope

Due to the success of the LRAIL prototype as demonstrated in preliminary field tests conducted on Amtrak and the desire for further refinement, this paper documents the use of an updated version of LR*AI*L to collect field data during the fall 2019 cycle of the Facility for Accelerated Service Testing (FAST) at the Transportation Technology Center's (TTC) High Tonnage Loop (HTL). Thus, the focus of this research is the development and field validation of a 3D laser-based approach which uses AI to objectively analyze 3D scans to create an input into change detection algorithms. The scope of this paper includes a description of the measurement system, data processing methods, and detailed results for spikes and fasteners from the collected data.

Methodology

Overall approach

The research team deployed the LR*AI*L system for data collection on the HTL at TTC six times over a seven-week period between September 10, 2019 and October 23, 2019. This period overlapped with the fall 2019 FAST train operating schedule. Each deployment involved two or more scans of the HTL captured in the forward and/or reverse direction for a total of thirty-five 3D scans of the HTL. For certain deployments, a subset of the HTL was unavailable due to maintenance operations, thus comparisons were not feasible on those sections. Among the selected novel datasets used for DCNN evaluation, the amount of overlapping data between them represented approximately 66% of the HTL.

The project team collected five scans of the HTL on the first day of field testing (September 10) which provided sufficient data to permit one inspection run to serve as a change baseline dataset and a separate dataset to serve for training the DCNN. The team also performed a physical walking inspection of the HTL loop on the first day of field testing to provide ground-truth information to assist with subsequent data interpretation.

Once a DCNN was trained for a particular feature, the research team evaluated its performance against an expert human

evaluator (a University of Illinois at Urbana-Champaign employee) with a performance target of 75% agreement. Following the performance evaluation, the trained DCNN processed a track scan from each deployment to create datasets to compare against the September 10 baseline for change detection purposes. The team developed a change detecting sensitivity target which reported present fastener counts per kilometer as a percentage (±1%), and changes in the position of multiple fasteners (≥10 mm for three consecutive fasteners in the same position along the same rail). These thresholds and the performance target were established prior to the start of the project through discussion with the project team and a subset of industry partner railroads.

Field data collection

Data collection at the HTL involved both a walking ground-truth inspection (initial deployment only) as well as 3D laser scanning for DCNN training and change detection analysis.

Three-dimensional laser scan data collection

Three-dimensional (3D) scans were captured via two 3D laser sensor heads mounted on a hi-rail test trailer that was pulled by a hi-rail pick-up truck (Figure 1). The sensors project high speed (up to 28,000 Hertz) laser lines across the track bed while synchronized cameras use custom filters to capture images of each projected line. During capture, software automatically compiled and merged successive lines into a continuous 3.6 m (11.8 ft) wide image of the track.

Figure 1. Hi-rail pick-up truck pulling hi-rail test trailer.

An optical encoder mounted on a rear wheel of the hi-rail trailer measured vehicle speed and triggered image capture on a distance basis (every 2 m, 6.6 ft, of travel along the track). Additionally, a blended inertial navigation system (GPS coupled with an Inertial Measurement Unit) captured and integrated the test trailer's latitude, longitude, and elevation into each 3D profile.

The resulting dataset contained geo-referenced 2D intensity and 3D range data with a longitudinal, transverse, and vertical resolution of one millimeter by one millimeter by 0.1 mm (0.039 in by 0.039 in by 0.0039 in) respectively, which is consistent with previous research.[16] Range data usable in both as 3D point cloud (LAS) and JPEG formats (grayscale image)— wherein the shade of pixels corresponds to the distance between the sensor and the surface—are shown in Figure 2 (light-shaded pixels corresponded to higher elevations).

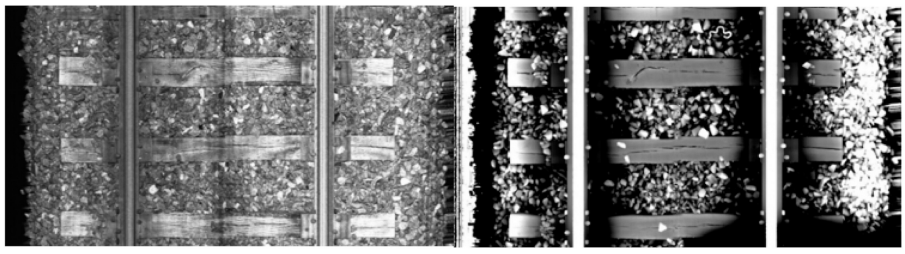

Figure 2. Images from system, 2D intensity (left) and 3D rectified range (right).

Latitude, longitude, and elevation data were subsequently used to preliminarily align repeat runs for change detection with 3D shape matching of components (e.g., crossties and fasteners), which helped refine positions and precisely match runs on a crosstie-by-crosstie basis.

Field test site characteristics

Beyond FAST's expedited accumulation of tonnage, one of the primary benefits of collecting data on the HTL was the high level of variability in track components. This variability permitted inspections of several identified features of interest including spikes and several clip types. The increased wear rate due to FAST operations provided quantifiable changes in feature condition each week of FAST train operations, which shortened this research program's timeline by decreasing the gap between necessary data collection deployments. In total, the HTL accumulated approximately 57 million gross tons (MGT) during the seven-week period.

Deep convolutional neural network (DCNN) training

Configuration

The DCNN consists of a multilayer region proposal network (RPN) and a multilayer classifier network based on fast region based convolutional neural network architecture. The output layers of both the RPN and the classifier network are softmax and regressor classifiers.

Training data preparation

The project team selected the second inspection run from the first deployment as the DCNN training set. This dataset contained 3800 images (1900 2D intensity and 1900 3D range) and formed a continuous run that the ground-truth inspection could validate. These 1900 images comprise a larger portion of the HTL (87%) than the subsequent testing data (66% of the HTL) due to improved track availability on the day that test data were collected.

To prepare a DCNN training set, the inspection data was first automatically inspected using a preliminarily-trained DCNN used by LR*AI*L. Following this initial processing, the team individually reviewed each resulting processed image to visually mark features that the preliminary DCNN missed or incorrectly classified (Figure 3). The research team also used the ground-truth inspection data during this task to add supplemental information when necessary.

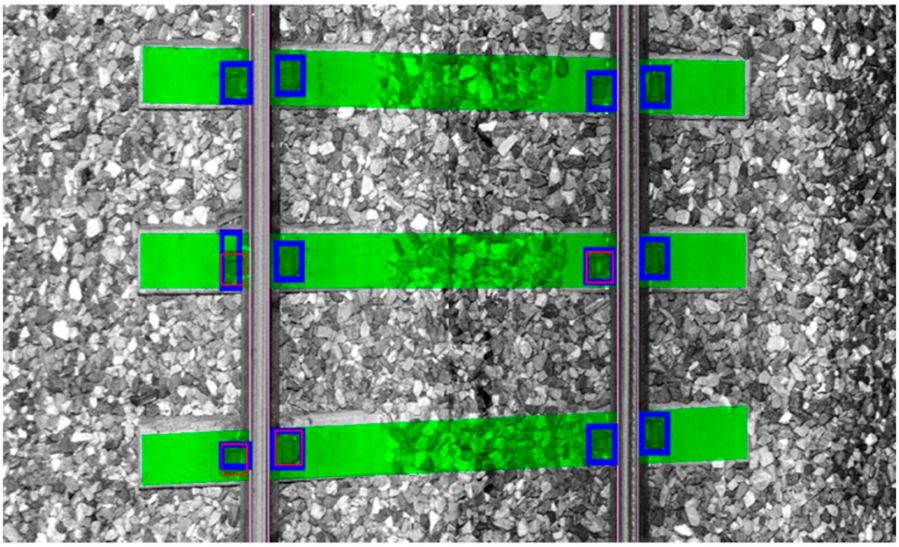

Figure 3. Intensity image showing DCNN-detected bounding boxes (blue) and expert review correction bounding boxes (red).

This process reproduced the expert evaluator's assessment in a digital format with the exact pixel position of each feature demarcated as well as the relevant conditions noted (e.g. broken e-clip, missing spike, etc.) in both 2D intensity as well as 3D range data. The DCNN training process used this enhanced dataset to align AI evaluations with that of the expert evaluator.

DCNN training

TensorFlow Core was used as both a training and validation tool with manually-labeled intensity and range datasets used as an input for the DCNN training process. In total three rounds of training and validation were repeated with more than 1,200,000 epochs (a complete cycle through the training images) completed in each round.

Elastic fastener (clip) training

Training for clip inspection first involved clip identification and then classification of clips into five numeric types which mapped to Pandrol e-clips and similar variants (e.g. PR clips), Safelok clips (including Safelok I and III), Skl tension clamps, and Pandrol Fastclips. While other clip types were present in the HTL, these five types made up most of the data collected and represent most of the systems currently used in North America. Clip "status" categorizations derived by the algorithm included: present, missing, loose, covered, and damaged. The missing classification was straight-forward for standard installations which involved just one clip on each side of each rail. However, unique components such as full and half-frame crossties provided instances where multiple possible clip positions existed for one or both sides of each rail, making identification by the DCNN harder.

Spike training

Training for spikes involved the detection of spike presence for each tie plate and the detection of missing spikes using pattern analysis. The DCNN was trained to compare the spiking pattern in place for the left rail against the right rail, with non-equal patterns (e.g. number of spikes) being flagged as either "missing spike" (when DCNN does not detect a spike) or "broken spike" (when the DCNN does not detect a proper head size). Additionally, a second pattern analysis involved the comparison of the spiking pattern for a given plate to the pattern present on both preceding as well as following crosstie plates (Figure 4). The final analysis simplifies the results to "missing spike" for the results.

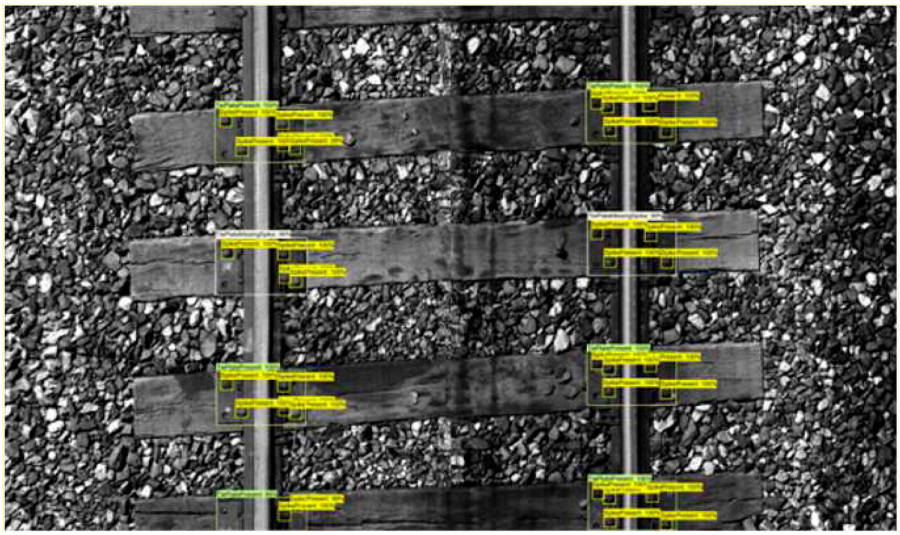

Figure 4. Two missing spikes detected since the surrounding crosstie plates have five spikes.

Retraining and performance tracking

Following each round of training, the DCNN models were implemented into the inspection software and again used to automatically process a dataset that was separate from the training set but captured on the same day. After automated processing, the research team (again) manually reviewed the results to mark features which the DCNN missed or incorrectly classified. Following this manual review, the DCNNs underwent retraining to account for errors. This process was repeated until DCNN detection precision could not be significantly improved through additional training cycles.

DCNN Performance was tracked in terms of the precision and recall. Precision indicates the level of false positives (FP) which the DCNN generates; or the tendency to report a condition as true when it is actually false (TN) (equation (1)). Recall indicates the level of false negatives (FN) which the DCNN generates; or the tendency to report a condition as false when it is actually true (TP) (equation (2)).

$$\textbf{Precision} = \frac{TP}{FP + TP} \tag{1}$$

$$Recall = \frac{TP}{FN + TP} \tag{2}$$

Overall performance of the DCNN was evaluated in terms of the combined precision and recall performance by plotting precision against recall and calculating the area under the curve with the result being expressed as mAP Loss, a general metric for algorithm error. Thus, mAP Loss achieved for each round of training was used as an indicator of the overall effectiveness and progress of training; with lower mAP Loss values indicating improved performance. An example of this practice from the project is presented in Figure 5, which presents the mAP loss (Y-axis) for each complete cycle (X-axis) through the training images.

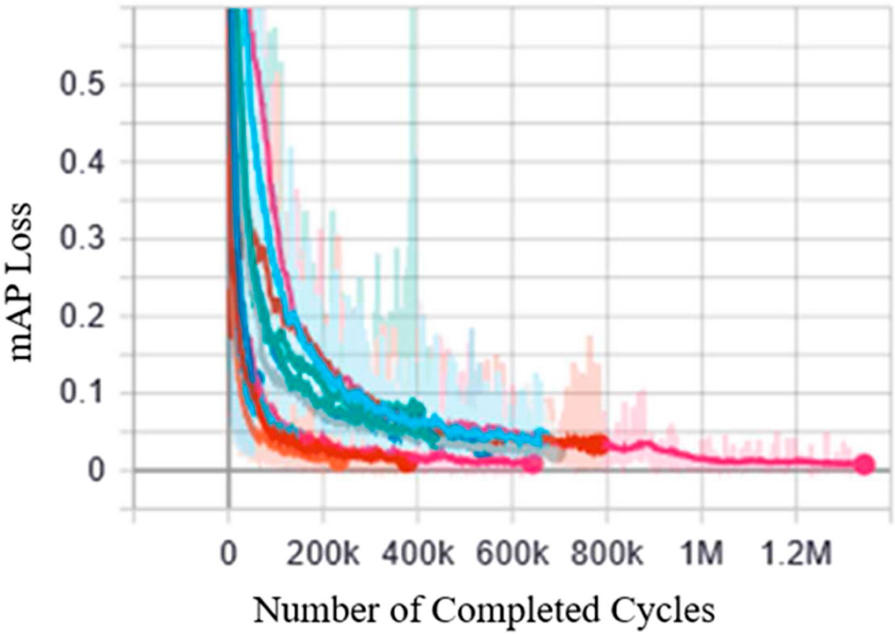

Figure 5. Summary of mAP losses (level of accuracy) for each competed cycle.

As can be expected, at the start of the training cycle (e.g., cycles 0–10,000) the general level of error (mAP Loss) for the DCNN was high across all aspect that were being trained (indicating poor DCNN performance). However, as cycles increase, the DCNN model error (mAP Loss) decreases which indicates a successful training approach (low FP and low FN).

DCNN detection results

The final performance evaluation of the DCNNs used the standard metrics of sensitivity (equation (3)), specificity (equation (4)), and percent agreement (equation (5)). These results are described later in the data analysis section.

$$Sensitivity = \frac{TP}{TP + FN} * 100\% \tag{3}$$

$$Specificity = \frac{TN}{FP + TN} * 100\% \tag{4}$$

$$Percent\ Agreement = \frac{TP + TN}{TP + FP + FN + TN} * 100\% \tag{5}$$

Change detection process

As a first step in the change analysis, the algorithm positionally aligned inspection pairs to ensure that each run compared the same position along the track to report changes between them. The change detection algorithm first used latitude, longitude, and elevation of each dataset to create a preliminary rough match (within ±1 mm on average) in position between runs. A second automated step conclusively matches positions between runs using 3D shape analysis of components (crossties, fasteners, tie plates, etc.) which the algorithm detects in the rough match region. The result was a crosstie-by-crosstie match between runs allowing for the development of change statistics on a crosstie-by-crosstie basis as well as using higher summary values.

Data analysis

DCNN detection results

Following training, the project team generated inspection results for two novel datasets; the first from the initial deployment (September 10) and the second from the final deployment (October 23). The research team selected bounding (initial and final) data given the accumulated tonnage (57 MGT) and maintenance operations between the initial and final deployment would likely produce the greatest change in the overall track condition and feature detection quantities. The first dataset came from a different run on the same day as the training dataset and facilitated evaluation of the DCNN under track conditions that were essentially unchanged from the training dataset but that had not been used for actual training.

The values for sensitivity, specificity, and percent agreement can be easily calculated when summarizing DCNN and expert evaluator results in a confusion matrix (Figure 6). Subsequent confusion matrices within Figure 6 display the results, which compares the results between features that were present or covered to those that were missing, loose, or damaged.

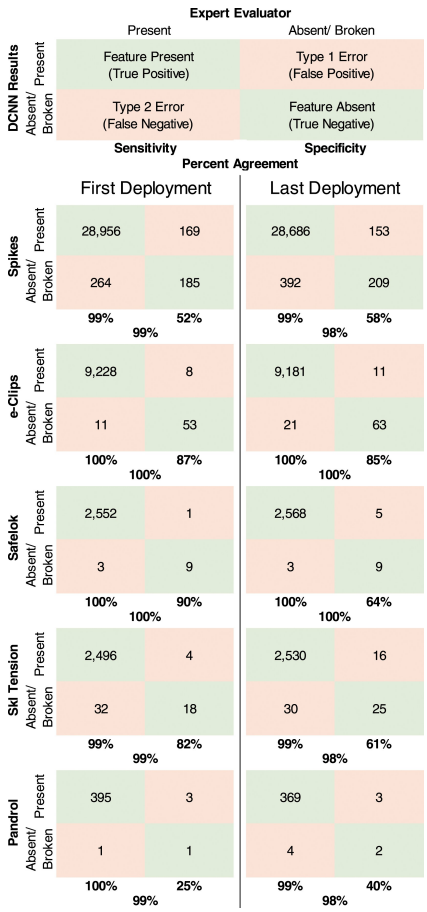

Figure 6. Generalized confusion matrix and performance evaluation.

Percent agreement and sensitivity results greatly exceeded the pre-established project goal of 75% regardless of feature type and time of deployment. However, the discrepancy between absent and present features and the low sample size for these absent

features may have caused specificity values to be under 75%. The overall quality of the DCNNs proved encouraging, and the data were used for subsequent change detection.

Change detection results

To evaluate track changes, DCNNs analyzed data from the first day (September 10) of fall 2019 FAST operations to produce a baseline dataset. Next, the change detection algorithm compared that dataset to a subsequent dataset (October 3) and the final dataset (October 23).

Fastener count

Change detection regarding clips involved run-to-run comparisons related to the number of acceptable (present and properly seated/installed) fasteners versus unacceptable fasteners (missing or loose). Fastener counts from the deployments were compared, and clusters of three or more fastener status changes within 10 meter (33 ft) segments of track are summarized in Table 1 for September 10 versus October 3 and October 23. The evaluator confirmed these results produced by the DCNN and the change detection algorithm. Despite obtaining maintenance data being obtained and reviewed for the HTL, no explanation could be found for the large cluster of fastener changes 2316 meters from the starting location.

Table 1. Threshold fastener changes.

Distance from starting location (meters)	Sep.10 to Oct. 3 changes	Sep.10 to Oct. 3 changes
832	3	<3
848	3	4
1140	3	<3
1156	<3	7
1220	4	4
1724	3	<3
1852	<3	4
1944	<3	3
2316	19	20
2448	<3	7
2588	12	11

Two change detection results from the second cluster are shown in Figure 7. While at first a fastener is incorrectly classified as present on September 10, its condition changes (further deteriorates) leading up to October 23. This change was detected, and the fastener is now correctly evaluated as missing/damaged (Figure 7). Additionally, a change is detected in fastener status as two fasteners which were covered as of September 10 are subsequently uncovered as of October 23.

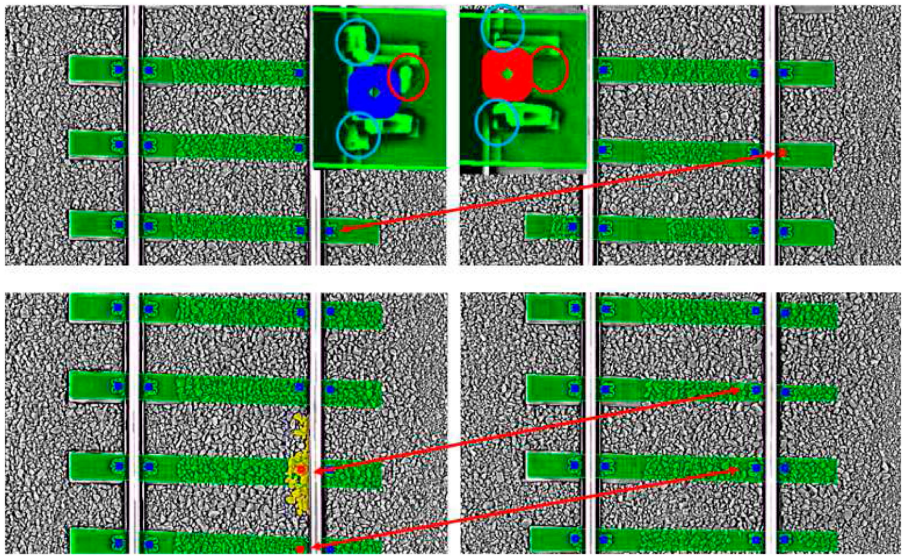

Figure 7. Fastener change identification examples (Sep. 10 vs Oct. 23).

Spike height

Spike height changes (measured directly from detected spikes) greater than seven millimeters between September 10 and October 23 are displayed in Figure 8 with large red triangles representing spike height increases and small green triangles representing spike height decreases. An example of a large spike height change is shown in Figure 9 where a spike saw a 19.01 mm increase between September 10 and October 23.

Figure 8. Areas with significant spike height changes (Sep. 10 vs Oct. 23).

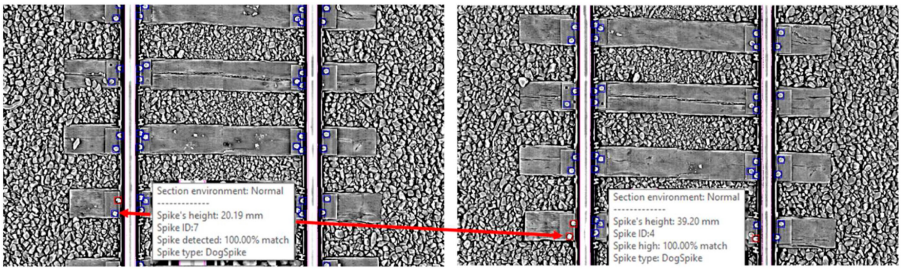

Figure 9. Example of spike height increase (Sep. 10 vs Oct. 23).

Discussion

DCNN agreement for spike and fastener condition between initial and final deployments was 98%; exceeding the 75% target. Additionally, while there were decreases in percent agreement between the initial and final deployments for spikes and fasteners, the difference between these percent agreement values was less than 1%.

These data indicate that the DCNNs were successful in determining when features were present regardless of fastener type given the high sensitivity results (**99% or greater**). It is possible that the DCNNs were biased towards detecting positive results due to the sample size discrepancy between present and absent features (e.g. more spikes present than absent in a tie plate), which reflects in the low values for specificity. This can be attributed to the small sample sizes for missing spikes when compared to the number of present spikes and that the sample size for missing spikes may have proved too small to be properly trained by the DCNN. This imbalance increased difficultly for the DCNN to properly define missing spikes. Considering that present spike sensitivity values for both deployments was **99%**, improvement with missing spike identification through increased sample size is fully expected.

A similar sample size challenge affected the results for Safelok and Pandrol Fastclip detection. Since the number of candidate examples for absent fasteners is very small (just 10 Safelok fasteners out of 2552 and just four Pandrol Fastclip out of 395), type one and type two errors are skewed. Additionally, since the HTL's sample size for present Safelok fasteners (2552 or just 9% of the total) and present Pandrol Fastclips (just 395, or 1%, of the total) remained small, improved assessments for DCNN accuracy may be reliant on further DCNN training for these features.

Despite the potential for training the DCNN to detect spiking patterns, the unique environment of the HTL presented challenges with respect to the training dataset. Many unique track components and configurations are tested on the HTL, and many different spiking patterns are present over a relatively short distance. For instance, while spiking patterns on Class I railroads are mostly consistent over many crossties, the HTL's patterns change within very short distances (e.g. as little as 10 to 20 ft) for which the DCNN sees spiking anomalies instead of the spiking patterns. While these patterns could be theoretically trained-for by using data from revenue service Class I track with longer and more homogenous track, the best alternative involves consultation with the operating railroad for their preferred spike pattern and setting thresholds for spike requirements (e.g. minimum number of spikes on gauge and field side).

Change detection analysis using DCNN results as inputs for both fastener and spike presence and condition proved successful, with the program properly aligning all crossties and correctly comparing almost all conditions. While some inaccuracies exist between spike and fastener changes, these errors are the result of DCNN detection accuracy and not the change detection algorithm. Further detection refinement should correspond with greater change detection accuracy.

Conclusions

The objective of this study was to evaluate the potential for the use of 3D laser triangulation, DCNN, and change detection technology to provide value-added inspection data to existing geometry car inspection systems.

In this study, 3D laser triangulation and DCNN analysis proved to be an accurate and reliable inspection method for fasteners and spikes. The agreement between the DCNNs and the expert evaluator for both spike and fastener presence showed 98% agreement; exceeding the 75% target. Change detection analysis using DCNN results as inputs was also successfully demonstrated for fastener and spike presence and condition.

This research has demonstrated encouraging progress in automating the track inspection process to help meet and exceed current FRA track inspection standards, with the ultimate objective of improving both safety and efficiency of rail operations. The system can detect subtle changes in track condition and goes beyond the traditional "pass/fail" approach to meet maintenance and safety thresholds. With sample size increases stemming from revenue service deployments on Class I track, DCNNs will be further refined to build on this prototype concept.

Based on this project's results, this technology may be used to develop a condition change index to facilitate track condition assessment and to provide continuous change-detection capabilities across an entire division or network. Additionally, the team plans to deploy this technology on a revenue service Class I railroad to collect data and evaluate DCNN performance under differing environmental conditions.

Authors' note

The material in this paper represents the position of the authors and not necessarily that of its sponsors. Finally, the authors acknowledge the involvement and support from, Amtrak, BNSF, and CN in their role as Industry Partners. J. Riley Edwards has been funded in part through grants to the Rail Transportation and Engineering Center (RailTEC) from CN and Hanson Professional Services, Inc.

Author contributions

The authors confirm contributions to the paper as follows: study conception and design: Ryan Harrington, Richard Fox-Ivey, Arthur De Oliveira Lima, Marcus S. Dersch, Thanh Nguyen, Van Nguyen, Marcus Dersch, John Laurent, and J. Riley Edwards; data collection: Ryan Harrington, Richard Fox-Ivey, and Arthur De Olivera Lima; analysis, and interpretation of the results: Ryan Harrington, Marcus S. Dersch, Richard Fox-Ivey, Thanh Nguyen, and Van Nguyen; draft manuscript preparation: Ryan Harrington, Richard Fox-Ivey, Arthur De Oliveira Lima, and J. Riley Edwards. All authors reviewed the results and approved the final version of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Federal Railroad Administration (FRA), part of the United States Department of Transportation (US DOT). This work was also supported by the National University Rail Center, a U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology Tier 1 University Transportation Center.

ORCID iDs

Ryan M Harrington https://orcid.org/0000-0002-4691-0400

Arthur de O Lima https://orcid.org/0000-0002-9642-2931

J Riley Edwards https://orcid.org/0000-0001-7112-0956

Marcus S Dersch https://orcid.org/0000-0001-9262-3480

References

1. Federal Railroad Administration (FRA). Chapter 1 - track safety standards, classes 1 through 5. In: *Track and Rail Infrastructure Integrity Compliance Manual*. Washington, DC: U.S. Department of Transportation, Federal Railroad Administration; 2014.
2. XivTM Aurora. Aurora Xiv, https://georgetownrail.com/Track-Inspection/Aurora-Xiv-1 (2020, accessed 18 April 2020. Crossref.
3. Orrell SC, Nagle JA, Villar C. *System and Method for Inspecting Railroad Track*. Austin, TX: US 7.616, 329 B2, 2009.
4. Berry A, Gibert-Serra X, McNew D, et al. High speed video inspection of joint bars using advanced image collection and processing techniques. In: Proceedings of the world congress on railway research (WCRR), Seoul, South Korea, Apr 18 2008. Available from: https://www.researchgate.net/publication/261309156_High_Speed_Video_Inspection_of_Joint_Bars_Using_Advanced_Image_Collection_and_Processing_Techniques.
5. RailCheck. RailCheck - bvSys Bildverabeitungssysteme Gmbh, https://www.bvsys.de/58.html (2020, accessed 18 April 2020)
6. Rail Corrugation. MERMEC Group. http://www.mermecgroup.com/inspect/track-measurement/61/rail-corrugation.php.AccessedApr18,2020 (2020, accessed 18 April 2020)
7. Laser Rail Inspection System. Pavemetrics, http://www.pavemetrics.com/applications/rail-inspection/laser-rail-inspection-system/ (2020, accessed 27 June 2020).
8. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks* 2015; 61: 85–117. Crossref. PubMed. ISI.
9. Chen J, Liu Z, Wang H, et al. Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Trans Instrum Meas* 2018; 67: 257–269. Crossref.
10. Faghih-Roohi S, Hajizadeh S, Nunez A, et al. Deep convolutional neural networks for detection of rail surface defects. In: 2016 international joint conference on neural networks (IJCNN). Vancouver, Canada, 24–29 July 2016: IEEE, pp. 2584–2589. Crossref.
11. Wei X, Yang Z, Liu Y, et al. Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Eng Appl Artif Intelligence* 2019; 80: 66–81. Crossref.
12. Giben X, Patel VM, Chellappa R. Material classification and semantic segmentation of railway track images with deep convolutional neural networks. In: 2015 IEEE international conference on image processing (ICIP). Quebec, QC, Canada, 27–30 Sept 2015: IEEE, pp. 621–625. Crossref.
13. Chellappa R, Xavier G, Vishal P. *Robust Anomaly Detection for Vision-Based Inspection of Railway Components*. Technical Report DOT/FRA/ORD-15/23, Washington, DC: Federal Railroad Administration. 2015, https://railroads.dot.gov/sites/fra.dot.gov/files/fra_net/14789/TR_UMD_AnomalyDetection_final.pdf
14. Henderson H, Borsholm A. *Demonstration of Commercial-Off-The-Shelf Change Detection on Railway Images*. Washington DC, USA: Technical Report SERV-REPT-0001686: Department of Transportation, Federal Railroad Administration, 2019
15. Fox-Ivey R, Nguyen T, Laurent J. *Laser Triangulation for Track Change and Defect Detection*. Technical Report DOT/FRA/ORD-20/08, Washington, DC: Department of Transportation, Federal Railroad Administration, 2020
16. Fox-Ivey R, Nguyen T, Laurent J. *Extended Field Trials of LRAIL for Automated Track Change Detection*. Technical Report DOT/FRA/ORD-20/14, Washington, DC: Department of Transportation, Federal Railroad Administration, 2020