TrackNet - A Deep Learning Based Fault Detection for Railway Track Inspection

Ashish James*†, Wang Jie*†, Yang Xulei†, Ye Chenghao‡, Nguyen Bao Ngan‡, Lou Yuxin⊥, Su Yi⊥, Vijay Chandrasekhar†, and Zeng Zeng§†
†Institute for Infocomm Research, A*STAR, Singapore
‡SMRT Trains Ltd, Singapore
⊥SMRT Corporation Ltd, Singapore
Email: {ashish_james, wang_jie, yang_xulei, vijay, zengz}@i2r.a-star.edu.sg
{YeChenghao, NguyenBaoNgan, LouYuxin, SuYi}@smrt.com.sg

***Abstract*—Reliable and economical inspection of rail tracks is paramount to ensure the safe and timely operation of the railway network. Automated vision based track inspection utilizing computer vision and pattern recognition techniques have been regarded recently as the most attractive technique for track surface defect detection due to its low-cost, high-speed, and appealing performance. However, the different modes of failures along with the immense range of image variations that can potentially trigger false alarms makes the vision based track inspection a very challenging task. In this paper, a multiphase deep learning based technique which initially performs segmentation, followed by cropping of the segmented image on the region of interest which is then fed to a binary image classifier to identify the true and false alarms is proposed. It is shown that the proposed approach results in improved detection performance by mitigating the false alarm rate.**

***Index Terms*—Railway track inspection; track fault detection; deep learning; deep convolution neural networks**

I. Introduction

Recently, feature learning using deep neural networks has proved to be successful when applied to a variety of computer vision and classification problems in diverse application domains. The accuracy of such systems for several benchmark datasets have improved over classical hand-crafted feature learning approaches and have achieved state-of-the-art performance on many use cases [1], [2]. Some of the advances made in these domains can be applied for detection and identification of faults in railway tracks which is crucial for the safety and availability of railway networks.

Around the world rail systems are among the most preferred public transportation methods and are becoming busier requiring them to operate with increasing levels of availability and reliability [3]. In addition, the speed and loads of trains have also been increasing greatly in recent years, and all these factors inevitably raise the risk of producing rail defects. There are different reasons for the occurrence of rail surface defects, for example as a result of fatigue, due to the repetitive passing of rolling stock over rail components such as welds, joints, and switches; or because of the impacts from damaged wheels [4]. Railway operators worldwide have been highly concerned about such defects as if treated late, they may lead to major revenue loss and have bigger implications like loss of life due to accidents.

Traditionally, a trained person visually inspects the rail for defects which makes the whole process slow, subjective and dangerous. This led to many advanced non-destructive testing (NDT) techniques, which acquire the condition of a rail from sensors (such as visual and ultrasonic) with the information fed to some sophisticated software to detect defects. Currently, the available NDT techniques for rail inspection utilizes visual cameras, ultrasonics, eddy current, etc. [5], [6]. One of the best performance for detecting internal rail cracks has been inspection utilizing ultrasonics [7], [8]. However, its inspection speed is slow (no more than 75 km/h) [5] and it cannot detect surface defects. In order to improve the inspection speed, several improved ultrasonic techniques such as electromagnetic acoustic transducers, lasers, and air-coupled ultrasonics were proposed, but they did not achieve enough progress to detect surface defects [5].

The NDT technique using eddy current uses magnetic field generated by eddy currents to identify defects [9]. This technique has relatively high inspection speed and is able to detect surface defects, so it is widely combined with ultrasonics for rail inspection. However, the sensor of eddy current is very sensitive to the lift-off variation with the probe positioned at a constant distance (no more than 2 mm) from the surface of the rail head [10]. As a result, the operation of eddy current testing is complex and sensitive; furthermore, the reported highest speed of this testing is also no more than 100 km/h [5].

With the recent advances in computer vision techniques, visual based track inspection system (VTIS) for rail surface detection have been developed. In VTIS, a high speed camera, installed under a test train captures the images of the track as it moves over them; with further analysis of the images being performed by an image processing software for custom applications such as bolt detection [11], corrugation inspection [12], and crack detection [13]. Visual based track inspection systems have the advantages of high speed, low cost, and

*Joint first author
§Corresponding author

978-1-5386-7528-1/18/$31.00 ©2018 IEEE

appealing performance and is regarded as the most attractive technique for track surface defect detection [5]. However, many of the commercial off-the-shelf (COTS) VTIS systems have high false alarm rate resulting in a considerable amount of maintenance man-hours spent to screen such images for eliminating the false alarms. This paper focuses on the VTIS system and proposes a multiphase deep learning based rail surface anomaly detection and classification technique.

The proposed technique, named TrackNet performs image segmentation in the first phase to extract the rail tracks and locate the Regions of Interest (ROI) from the raw image. This step benefits the classification being performed in the final phase by discarding the noisy background and other non-relevant information. The cropped images focusing on the ROI are then fed to the final classification phase where the rail defects are identified and classified.

The rest of the paper is organized as follows. In Section II, a review of the related works on rail defect detection is presented. Section III describes the problem addressed in this work followed by the proposed approach in Section IV. In Section V, the experimental results are presented with a comparison of different deep learning based VTIS techniques. Finally, Section VI provides the conclusions and suggests the possible future investigations.

II. Related Works

Vision based track inspection technology has been gradually adopted by the railway industry since the pioneering work by Torsino *et. al.* [14], [15]. Classically, the common choices of features for detection from visual data has been gradient-based features such as the histogram of oriented gradients (HoG), scale-invariant feature transforms (SIFT), spacial pyramids, and basis function representations such as Gabor filters. In [11], [16], a two 3-layer neural network running in parallel is used to detect hexagonal headed bolts. In [17], a VisiRail system which collects images on each rail side, and find cracks on joint bars using edge detection and a support vector machine (SVM) classifier that analyzes the extracted features from the edges is proposed. A system for detecting tie plates and missing spikes using an AdaBoost-based object detector is proposed in [18].

In recent years, the expansion of feature learning using neural networks has provided a better tool for extracting features that are specifically tailored for each domain. In [19], a convolutional neural network trained on a database of photometric stereo images for detecting steel defects on rail surfaces is proposed. By means of differently colored light-sources illuminating the rail surfaces, the defects are made visible in a photometric dark-field setup. A max-pooling convolutional neural network is used for steel defect classification in [20]. In [21], [22], deep convolutional neural networks have been used for rail fastening condition monitoring with the focus on identification of track components such as ballast, concrete, wood, and fastener. In [4], deep convolutional neural networks with different network architectures characterized by different depth and activation functions are studied for the detection and classification of rail surface defects.

III. Problem Description

In VTIS, the images are initially captured by the image acquisition subsystem which are then fed to the image analysis subsystem for rail surface anomaly detection and classification. The defects flagged by the image analysis subsystem are then inspected by a human reviewer and it has been observed that there is a high false alarm rate. This is caused due to varying reasons such as animal droppings, writings etc. on the rail track which are then subsequently classified as a rail surface defect by the VTIS. A case of true and false alarm (caused by writing on the track) flagged by a COTS VTIS is shown in Fig. 1.

(a) True alarm (b) False alarm

Fig. 1: Images of true and false alarm from a COTS VTIS

The high false alarm rate results in considerable amount of maintenance man-hours expended in just screening through thousands of images to identify the correct defects. However, such a task is quite cumbersome and the multiphase deep learning technique proposed in this paper, named TrackNet will enhance the VTIS performance by mitigating the false alarm rate and is explained in the next Section.

IV. Approach

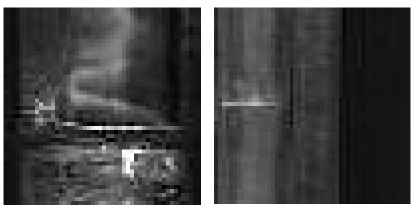
Conventionally, the images from the VTIS with additional data augmentation are used for fault classification. Deep convolutional neural networks are popular models when dealing with image classification and fault detection problems. In this work, state-of-the-art convolutional neural networks such as ResNet and DenseNet are adopted as the baseline techniques for performance comparison with the proposed TrackNet.

TrackNet

The proposed technique named TrackNet is a multiphase deep learning approach which integrates track segmentation and true/false alarm classification tasks. This is achieved through two neural networks with one dealing with semantic track segmentation and another on classifying the segmented images as either true or false alarms.

For semantic segmentation, a U-Net is used to extract rail tracks and locate ROI. The U-Net is a convolutional neural network architecture for fast and precise image segmentation [23] and have up to now outperformed the prior best method (a sliding-window convolutional network). The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. It introduces the shortcut/skip connections to preserve the pixel-level information for different image resolutions. In the present context, this is crucial to minimize the ROI as the rail tracks appear in various forms, i.e. tracks split and merge for different routes. In TrackNet, the initial segmentation is done by the U-Net with the architecture specified in [23] and trained using Adam algorithm with an initial learning rate of $1e-4$ and binary cross entropy as the loss function. The model is trained with images in mini-batches of 4 and the best model is selected from 10 epochs.

After the initial segmentation by the U-Net and extraction of ROI, image processing tools are used to crop the portion of the potential faulty region in the images indicated by the bounding boxes. The size of the cropping window is approximately 64x64 and a case of true and false alarm image after the cropping is shown in Fig. 2. The cropped images are saved and then fed to the next phase for classification.



(a) True alarm (b) False alarm

Fig. 2: Cropped images of true and false alarm case after segmentation by U-Net

For the last phase in TrackNet, the cropped images with ROI are fed to a neural network that classifies the images into either True or False alarms. The performance of the proposed technique is compared for different classifier architectures such as ResNet [24] and DenseNet [25], which are currently the pinnacles of neural network architectures for image classification. Among this DenseNet have several compelling advantages such as alleviating the vanishing-gradient problem, strengthening feature propagation, encouraging feature reuse, and substantially reducing the number of parameters. This makes the optimization of very deep neural networks trackable and robust. The architecture of the proposed technique, TrackNet is illustrated in Fig. 3.

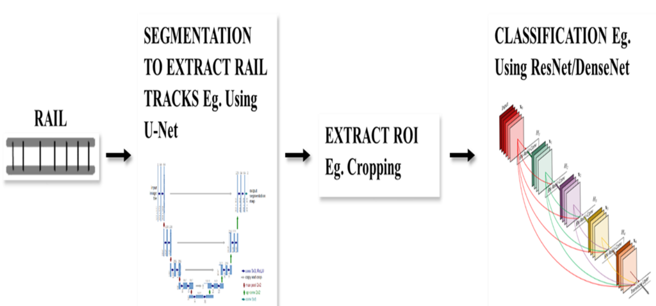


Fig. 3: TrackNet architecture

For the final phase of the TrackNet, the weights of the ResNet/DenseNet model are initialized from a model trained on ImageNet [26]. Typical weights for pre-trained model is based on ImageNet as the training is based on classification of real-life objects for 1000 classes. The use of ImageNet weight substantially reduces the cost for fine-tuning. The network is trained end-to-end using stochastic gradient descent with standard parameters and using images in mini-batches of size 16. In this work, the final fully connected layer is replaced with a small customized convolutional neural network model that consists of two layers. The number of units in the two layers are 256 and 2, respectively. The final fully connected layer corresponds to the two-classes representing the true/false alarm cases.

The fine-tuning process for training the TrackNet consists of two phases. In the first phase, all layers except the customized convolutional neural network layers are frozen and trained so as to customize the added layers according to the present dataset. By making use of the features from the previous frozen blocks, the final customized layers can be well tuned to the dataset. In the second phase, all layers are unfrozen and a typical classification training is executed. This two phase fine-tuning can improve the average accuracy of classification as well as speed up the fine-tuning process. For each of the cases, the training is performed for 50 epochs and one with the lowest validation loss is selected as the best model.

V. Experimental Results

In this section, initially the experimental setup is described followed by the performance results of the proposed technique.

A. Experiment Setup & Data Description

All images used in this experiment are actual rail tracks that are collected by a COTS VTIS. The top view of the rail track is captured by the VTIS as shown in Fig. 1. The dataset consists of 138 images with each image having potentially at least one faulty area. Of the 138 images only 14 are true alarms and rest of the images are false alarms flagged by the COTS VTIS.

The raw images from the image acquisition subsystem contains many information such as machine generated text-based label and comment, other parts of the rail track, rocks and ties, etc as shown in Fig. 1, which are irrelevant for the type of track defect classification namely rail discontinuity considered in this paper. This motivates to use segmentation for extracting the rail tracks in TrackNet. The images are initially resized to 512×512 before being fed to the U-Net for segmentation. The segmented images are then cropped to a size 64×64 around the ROI. The images are then upscaled to 224×224 before being fed to the ResNet/DenseNet for the final classification task. This is performed according to the mean and standard deviation of images in the ImageNet training dataset. Further, data augmentation is performed on the images being fed to the U-Net and DenseNet/ResNet architectures. The images in the training dataset are augmented using standard parameters such as rotation = 0.2, shift = 0.05, shear = 0.05, zoom = 0.05, and by vertical and horizontal mirroring of the images. This basically expands the dataset by a factor of around 30 resulting in nearly 4000 images for training.

An additional trick is used in the final phase of TrackNet when training it for classification. In the first phase of fine-tuning the ResNet/DenseNet, instead of training the entire model at runtime, a typical ResNet/DenseNet model trained on the ImageNet dataset is used to make predictions. The bottleneck features at the last layer before the final fully connected layer are extracted and saved in static data files. The customized layers of the TrackNet will use these saved features as input during training. This saves considerable amount of runtime computational resources.

The experiments are run on an Ubuntu 16.04 machine with four Nvidia Titan X GPUs. For the classification task, the dataset is divided into training and testing set with 75% allocated for training and 25% for testing. Further, K-fold cross validation is performed on the model with K being 4.

B. Performance Results

In the initial phase of TrackNet, segmentation is performed to extract the rail track from the random camera shot image as shown in Fig. 4.

(a) Original image

(b) Segmented image

Fig. 4: Extracted rail track from the original image

The performance of this phase is validated in terms of the dice coefficient, which is defined by

$$DICE = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{1}$$

where X and Y corresponds to the prediction and the target, respectively. For the initial segmentation task, a dice coefficient of 0.99 is obtained for the configuration explained in Section V. The segmented images are then sharpened by setting a threshold of 0.6. The output after the sharpening is shown in Fig. 5.

The cropped images after the first phase of TrackNet are then fed to the final classification phase which differentiates the true and false alarms. In this paper, two neural network architectures namely DenseNet and ResNet are employed and performance is compared for the classification phase. The performance metric used for comparison is the accuracy of classification and the results are illustrated in Table I. It can be observed that the performance of the TrackNet with ResNet and DenseNet as the classifier are quite close in terms of accuracy with the DenseNet appearing to be a slightly better at distinguishing true/false alarms.

(a) Segmented image

(b) Sharpened image

Fig. 5: Segmented image sharpened through thresholding operation

TABLE I: Performance Comparison

Fold	ResNet	DenseNet	TrackNet using ResNet Classifier	TrackNet using DenseNet Classifier
1	0.6332	0.6746	0.8644	0.9036
2	0.6050	0.6988	0.9211	0.8795
3	0.7095	0.6482	0.8892	0.8915
4	0.5914	0.5549	0.8774	0.9390
Average	0.6347	0.6441	0.8880	0.9034

C. Baseline Comparison

As mentioned in Section IV, state-of-the-art deep learning models for classification such as ResNet and DenseNet trained on the raw images from the image acquisition subsystem are used as the baseline system for comparison with the proposed TrackNet. The performance comparison results are shown in Table I. Recall that the difference in TrackNet is that instead of using the raw images for classification as in the baseline systems, it uses the track only content extracted from the raw images for classification.

Given that the faulty region in majority of our dataset only occupies a few thousandth of the area of the whole image as shown in Fig. 1, hence a large margin of image content does not contribute directly to track fault namely rail discontinuity classification. By extracting the ROI, it can be clearly seen that TrackNet satisfactorily distinguishes the true and false alarms. When compared with the baseline systems which has the best accuracy of 71%, the TrackNet brings in a huge improvement by having an average accuracy of 90%. Intuitively, by focusing on the ROI, TrackNet minimizes the input noise from external and environmental noise. This makes it a suitable model for detecting track faults on large scale industry level environments for railway track inspection.

D. Limitations

The limitations of the proposed TrackNet has been identified into three. Firstly, this work focuses only on one type of track fault namely rail discontinuity for classification which enhances the need for extracting the ROI. However, in many other types of fault detection and classification scenarios rely on the status of surrounding objects and environmental condition. Secondly, the machine generated comment region on raw images from the image acquisition subsystem introduces unnatural noise, which have been shown to degrade the classification accuracy. Finally, majority of images in the dataset show a clear pattern where the tracks are perfectly arranged vertically, which might not be the case when track crossings are involved. However, this can be addressed by training the TrackNet with images of track at different alignments.

VI. Conclusion & Future Works

In this paper, a multiphase deep learning technique is introduced for detecting rail surface defects in vision based railway track inspection system. The first phase in this technique extracts the track through segmentation and the extracted tracks are then used for classification. Such an approach enables the classifier to focus on the ROI and results in better performance. In the current context, binary classification is performed with emphasis on mitigating the false alarms in VTIS. However, there are several types of rail defects and exploring a generalized deep learning approach that can automatically detect other types of defects will be the future direction of our work.

Acknowledgment

This work was performed as one of the initiatives under the A*STAR-SMRT Urban Mobility Innovation Centre and the labelled images used were provided by SMRT Trains Ltd.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proc. Intl. Conf. Neural Information Processing Systems*, 2012, pp. 1097–1105.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Intl. Conf. Learning Representations (ICLR)*, vol. abs/1409.1556, 2014.

[3] J. Chen, C. Roberts, and P. Weston, “Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems,” *Control Engineering Practice*, vol. 16, no. 5, pp. 585–596, 2008.

[4] S. Faghih-Roohi, S. Hajizadeh, A. Nez, R. Babuska, and B. D. Schutter, “Deep convolutional neural networks for detection of rail surface defects,” in *IEEE Intl. Joint Conf. Neural Networks (IJCNN)*, Jul. 2016, pp. 2584–2589.

[5] M. P. Papaelias, C. Roberts, and C. L. Davis, “A review on non-destructive evaluation of rails: State-of-the-art and future development,” *Proc. Institution Mech. Eng., Part F: J. Rail Rapid Transit*, vol. 222, no. 4, pp. 367–384, 2008.

[6] R. Clark, “Rail flaw detection: Overview and needs for future developments,” *NDT & E Intl.*, vol. 37, no. 2, pp. 111 – 118, 2004.

[10] H.-M. Thomas, T. Heckel, and G. Hanspach, “Advantage of a combined ultrasonic and eddy current examination for railway inspection trains,” *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 49, no. 6, pp. 341–344, 2007.

[7] R. Clark, S. Singh, and C. Haist, “Ultrasonic characterisation of defects in rails,” *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 44, no. 6, pp. 341–347, 2002.

[8] R. S. Edwards, S. Dixon, and X. Jian, “Characterisation of defects in the railhead using ultrasonic surface waves,” *NDT & E Intl.*, vol. 39, no. 6, pp. 468–475, 2006.

[9] M. Bentoumi, P. Aknin, and G. Bloch, “On-line rail defect diagnosis with differential eddy current probes and specific detection processing,” *Eur. Phys. J. Appl. Phys.*, vol. 23, no. 3, pp. 227–233, 2003.

[11] F. Marino, A. Distante, P. L. Mazzeo, and E. Stella, “A real-time visual inspection system for railway maintenance: Automatic hexagonal-headed bolts detection,” *IEEE Trans. Syst., Man, Cybern. C*, vol. 37, no. 3, pp. 418–428, May 2007.

[12] C. Mandriota, M. Nitti, N. Ancona, E. Stella, and A. Distante, “Filter-based feature selection for rail defect detection,” *Machine Vision and Applications*, vol. 15, no. 4, pp. 179–185, 2004.

[13] L. Jie, L. Siwei, L. Qingyong, Z. Hanqing, and R. Shengwei, “Real-time rail head surface defect detection: A geometrical approach,” in *IEEE Intl. Symp. Indust. Electron.*, Jul. 2009, pp. 769–774.

[14] J. J. Cunningham, A. E. Shaw, and M. Trosino, “Automated track inspection vehicle and method,” U.S. Patent 6 064 428, May, 2000.

[15] M. Trosino, J. J. Cunningham, and A. E. Shaw, “Automated track inspection vehicle and method,” U.S. Patent 6 356 299, Mar., 2002.

[16] P. D. Ruvo, A. Distante, E. Stella, and F. Marino, “A gpu-based vision system for real time detection of fastening elements in railway inspection,” in *IEEE Intl. Conf. Image Processing (ICIP)*, Nov. 2009, pp. 2333–2336.

[17] X. Gibert, A. Berry, C. Diaz, W. Jordan, B. Nejikovsky, and A. Tajaddini, “A machine vision system for automated joint bar inspection from a moving rail vehicle,” in *ASME/IEEE Joint Rail Conf. & Internal Combustion Engine Spring Technical Conf.*, Mar. 2007, pp. 289–296.

[18] Y. Li, H. Trinh, N. Haas, C. Otto, and S. Pankanti, “Rail component detection, optimization, and assessment for automatic rail track inspection,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 760–770, Apr. 2014.

[19] D. Soukup and R. Huber-Mörk, “Convolutional neural networks for steel surface defect detection from photometric stereo images,” in *Advances in Visual Computing (ISVC)*, 2014, pp. 668–677.

[20] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout, “Steel defect classification with max-pooling convolutional neural networks,” in *Intl. J. Conf. Neural Networks (IJCNN)*, Jun. 2012, pp. 1–6.

[21] X. Gibert, V. M. Patel, and R. Chellappa, “Deep multitask learning for railway track inspection,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.

[22] X. Giben, V. M. Patel, and R. Chellappa, “Material classification and semantic segmentation of railway track images with deep convolutional neural networks,” in *IEEE Intl. Conf. Image Processing (ICIP)*, Sep. 2015, pp. 621–625.

[23] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, pp. 234–241, 2015.

[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conf. on Computer Vision Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778.

[25] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *IEEE Conf. Computer Vision Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269.

[26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 115–211, 2015.