Sensors and Materials, Vol. 32, No. 2 (2020) 549–561
MYU Tokyo

S & M 2121

Development of Artificial-intelligence Vision System for Measurement On-service Train for Train-track Inspection

Anakkapon Saenthon,[1] Somyot Kiatwanidvilai,[2] and Suparat Pumyoy[2*]

[1]Department of Manufacturing System Engineering, King Mongkut's Institute of Technology Ladkrabang,
1, Soi Chalongkrung 1, Chalongkrung Street, Ladkabang, Bangkok, 10520
[2]Department of Electrical Engineering, King Mongkut's Institute of Technology Ladkrabang,
1, Soi Chalongkrung 1, Chalongkrung Street, Ladkabang, Bangkok, 10520

(Received June 29, 2019; accepted January 24, 2020)

Keywords: rail, track, artificial intelligence, vision system inspection

We studied and inspected the evaluation of safety during the operation of a rail system. In addition, in order to find patterns in the interaction between wheels and rails, the interaction is investigated using an automatic vision system to analyze the patterns that occur. The model of a faster region with a convolutional neural network (R-CNN) can be used to design artificial intelligence (AI) models with vision systems for detecting abnormal operations that occur, causing rail accidents, and to monitor the measurement of tracks and interactions between wheels and tracks in real time.

1. Introduction

Thailand is developing a basic rail infrastructure with a length of more than 4500 km, including dual-rail trains and city buses, and mass transit trains. The length of the railway (Railway Track) is up to a distance of more than 7000 km. However, the system is required to be ready to provide service under the safety standards so the rail infrastructure must be reliable. According to the data of rail accidents collected by the Federal Railway Association (FRA), 45 % of rail accidents were caused by irregularities of railways and various devices on the trains. Risk prevention can be achieved by effective measurement and maintenance planning. The rails currently in use can be analyzed using images. Oh *et al*. presented a vision-based tracking system with a train signal system to stop trains when there is an emergency.(1) Rikhotso *et al*. presented 3D rail modelling and measurement for rail profile condition assessment.(2) Yaman *et al*. presented a method for the detection of abnormalities of railway tracks from photos taken by cameras.(3) Liu *et al*. used image processing to detect the damage of the rail surface.(4) In addition, Girshick *et al*. presented the region with convolutional neural network (R-CNN) method together with photographs from which objects of interest are detected.(5) Later, Girshick presented fast R-CNN, which takes less time than R-CNN.(6) Therefore, we are interested in applying interdisciplinary knowledge, such as knowledge of electrical and electronic engineering, computer engineering, mechanical engineering, rail transport engineering, and

*Corresponding author: e-mail: suparat.pu@kmitl.ac.th
https://doi.org/10.18494/SAM.2020.2529

ISSN 0914-4935

other related fields such as vision systems, artificial intelligence (AI), and Internet of Things (IoT), to monitor interactions between wheels and tracks in real time. Nassu and Ukai applied a basic vision system to evaluate the changes in train tracks.(7) The main purpose of this research is to develop a tool for measuring the abnormalities of a rail system. Such a tool will be beneficial for the safe use of a rail system by both rail system providers and rail system users. It will also help to reduce the possibility of accidents involving rail systems, which would affect the credibility of the country as well.

2. Theory

In this paper, we present a vision system and AI based on image processing.

2.1 Vision system

The operation of the vision system involves image processing or computer calculation to obtain the required quality and quantity of information. There are various important steps to derating sharp images to eliminate interference in images. Objects in photographs are isolated to obtain an image of the object for analysis to acquire quantitative data such as size, shape, and the direction of movement of objects in the photograph. We can subsequently use these quantitative data for analysis and to create systems for use in various applications, such as the fingerprint recognition system, the quality inspection system of products in the manufacturing process of industrial plants, the grade separation or inspection system of the quality of agricultural crops, the automatic postal code reading system that sorts letters from images of the postal code on the envelope, the data collection system of vehicles that enter and exit a building using photographs of the vehicle license plate, the road traffic monitoring system in which the number of cars on the road is counted in a photograph taken by closed-circuit television (CCTV), the face recognition system for surveillance of terrorists in buildings, at landmarks, or immigration areas, the advanced license plate recognition system for car parking,(8) the image processing technique in a real-time vision system for automatic weeding strategy,(9) the vision system in the weaving loom industry,(10) and the vision system for monitoring intermodal freight trains.(11) These systems require much image processing that must be repeated. It is impossible for human beings to analyze this data because it would be excessively time-consuming. Therefore, computers play an important role in these functions in place of humans.

2.2 AI

AI is a field of computer science concerning how to make computers have capabilities like those of humans or imitate human behavior, especially the ability to think by themselves or have intelligence. Such humanlike intelligence is created by the computer and therefore called AI. People differ depending on how they use intelligence considering behavior in the environment or the results of AI (the ability of AI to think). Hence, the definition of AI is based on the abilities of humans divided into the following four groups:

• acting humanly,
• thinking humanly,
• thinking rationally, and
• acting rationally.

The neuron model structure used in this research is a faster R-CNN,[6] which has the ability to detect the position of an object in an image and recognize the object type. For the Faster R-CNN, the main idea is to bring the selective search, which is the area of interest containing the target object, included in the neuron model in the same structure. However, Liu improved the Faster R-CNN for object detection.[12] In addition, Cai *et al*. presented street object detection based on the Faster R-CNN.[13]

As shown in Fig. 1, this model has the following three main components.

1) the base that performs the extraction feature or considers the distinctive features of the image,
2) the region proposal network (RPN) area that separates "expecting or maybe" of an object from the feature map, and
3) the classification section that processes the feature map and region obtained from the RPN by region of interest (RoI) pooling to determine which areas of the image contain objects.

VGG-16 is a CNN model proposed by Simonyan and Zisserman.[16] The VGG-16 structure shown in Fig. 2 can support colorized black and white images. The first layer can support pictures with a size of 224 × 224 pixels and RGB images. The work structure is divided into five convolution steps. Then, the final convolution is the fully connected classifier. Details are shown in Fig. 3.

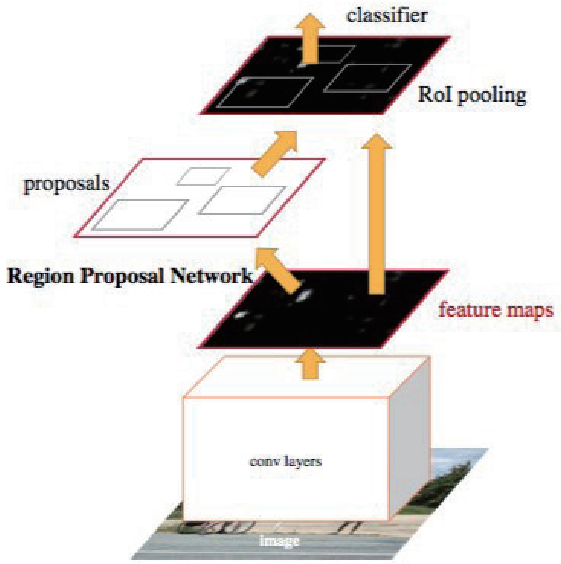


Fig. 1. (Color online) The faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.[14]

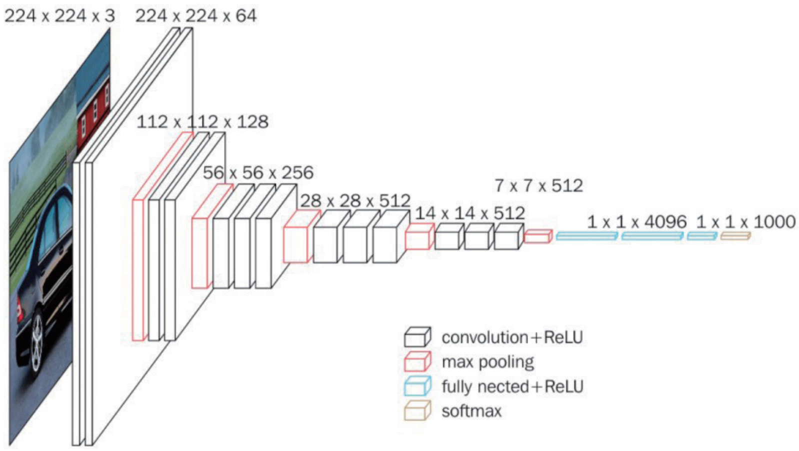


Fig. 2. (Color online) Structure of VGG-16 used as the base of the model.(15)

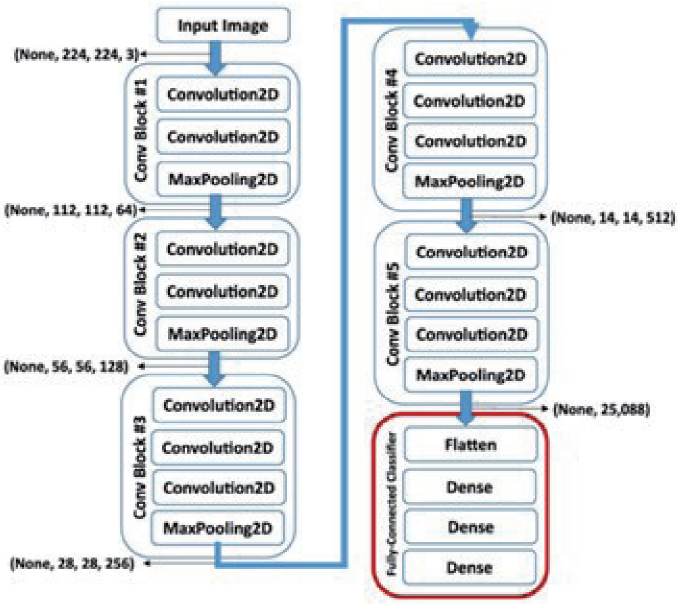


Fig. 3. (Color online) Details of each VGG-16 layer.

3. Experimental Results

3.1 Determining the scope of consideration for preparing images

Figure 4 shows an example of images from the video camera installed under the carriage. In data collection, a camera is used to record data as a video of the normal state of movement for the purpose of quality assessment analysis of safety and railway conditions. When using the image signal to consider the nature of the risk, the characteristics of wheels and rails are as shown in Fig. 5.

From Fig. 5, we found that the wheel rim was very close to the edge of the rail, causing the wheel rim to hit the track. This collision may result in damage to both the rail and the wheel. To determine the area of interest, the process converted the video to 32400 images suitable for studying and memorizing by AI (Fig. 6).

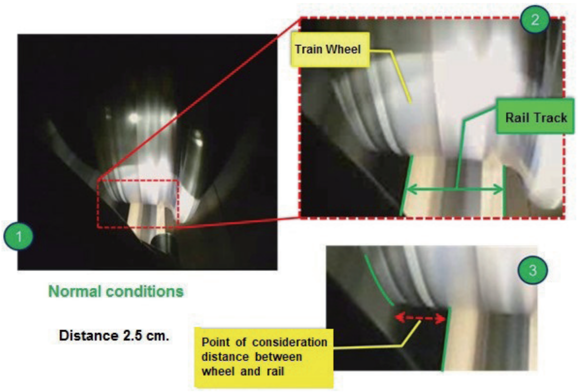


Fig. 4. (Color online) Images of wheel and rail area to determine the point of consideration.

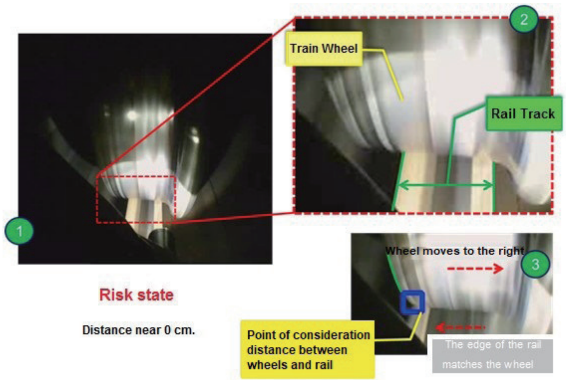
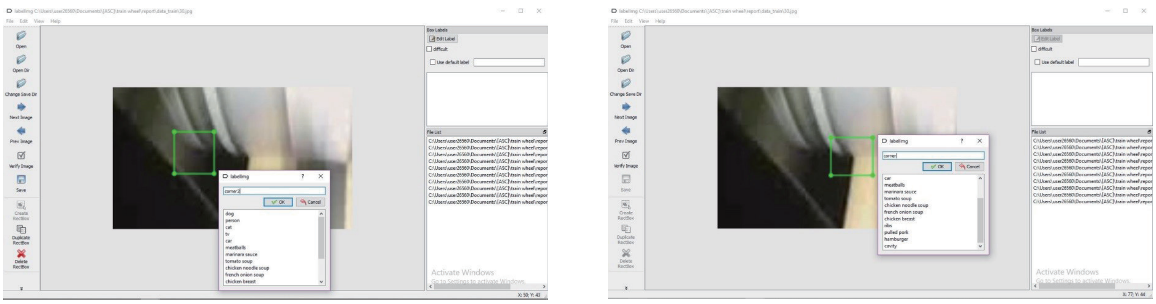


Fig. 5. (Color online) Images showing wheels and the risk of damage to wheels.

Fig. 6. (Color online) Transforming video to images.



(a) (b)

Fig. 7. (Color online) Selection of the wheel rim and the rail rim area. (a) Rim of the wheel. (b) Rim of the rail.

On the basis of the total amount of data, the data were separated into the following two main parts: information used for learning by the AI and information used to test AI capabilities. The procedure specifies two areas, namely the wheel rim and the edge of the track, as shown in Figs. 7(a) and 7(b), respectively.

3.2 Detailed structure of AI model

Figure 8 shows the detailed structure of the Faster R-CNN. First, import the image into VGG-16 to find the hidden layer in a total of five repetitions. Then, perform convolution to find the strength of the object in the image and apply the rectified linear unit (ReLU) to eliminate

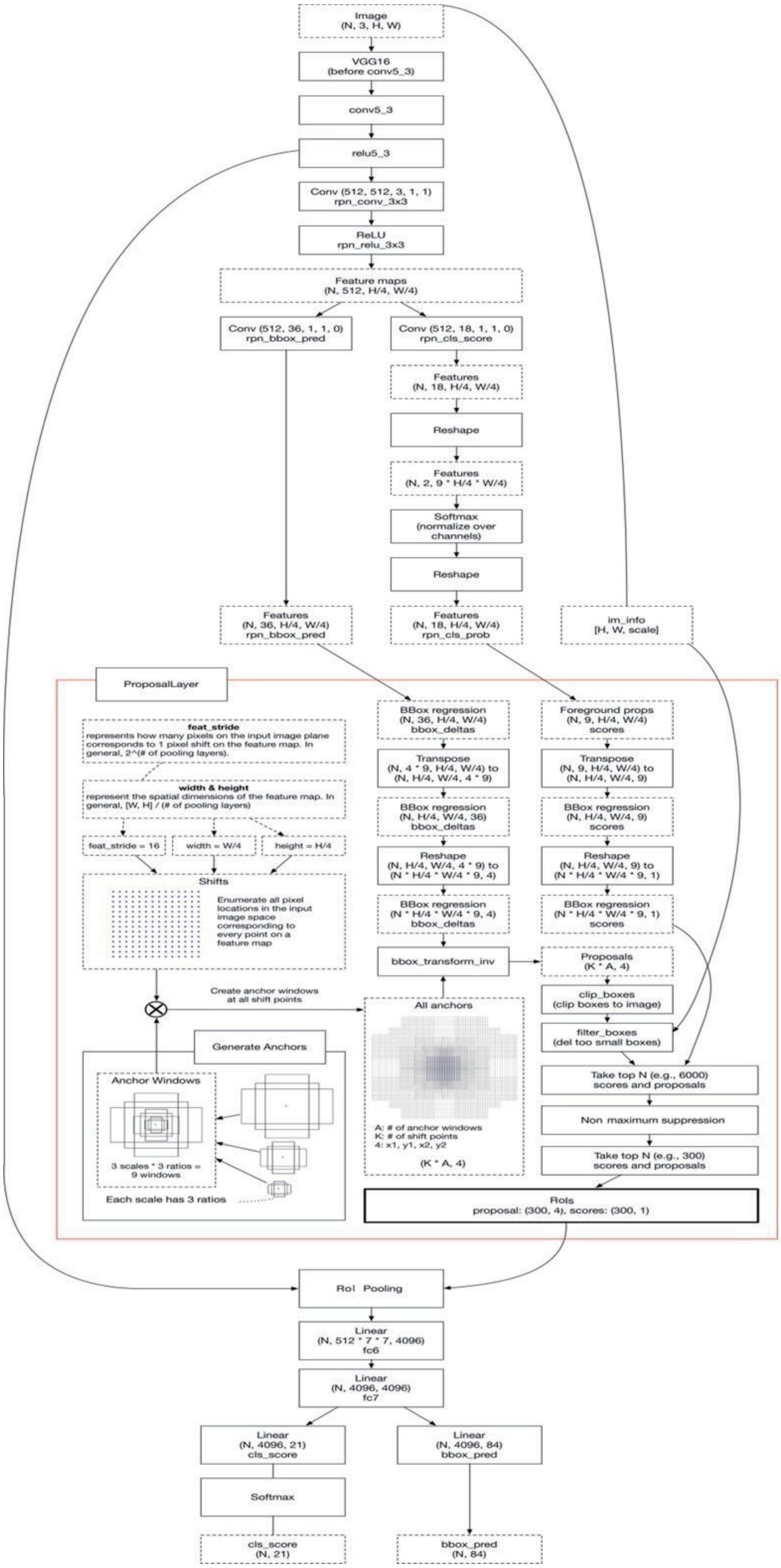


Fig. 8. (Color online) Structure of the Faster R-CNN.[5]

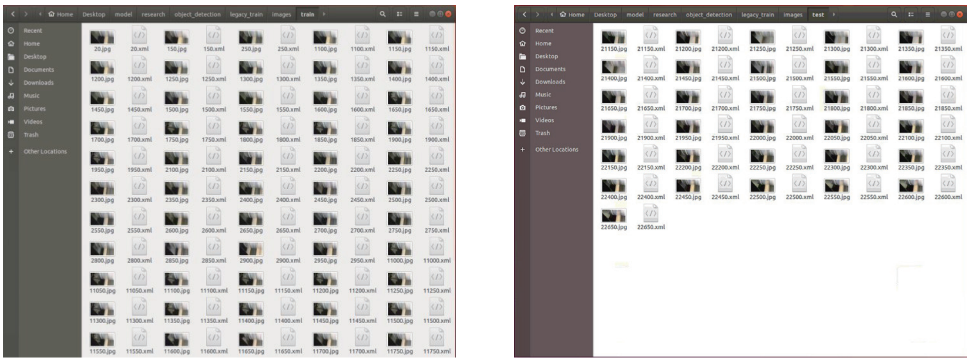
linearity. Next, import data into the RPN to separate the parts that are expected to be objects. Then, obtain images of many objects using only the feature map from the final convolution layer and determine the coordinates of the objects on the image. Use the object image to find the proposal layer to obtain the RoI. Then, carry out RoI pooling to obtain the fixed size vector. After that, send the information to the classification section.

3.3 Learning to remember to analyze the positions of wheels and rails

Figure 9 shows the determination of the area on the rim and the edge of the rail using the LabelImg program. A total of 81 images were used, 50 images for learning and 31 images to test the AI results.

The prepared image data for locating the object of interest is in the form of a Label file. The file information consists of the path, the location of the file, *X* position, *Y* position, width, and height. This information is used to verify the accuracy of the AI results.

Figure 10 shows the TensorBaord graph. The TensorBoard graph is a visualization tool that appropriates with the AI training process. The TensorBoard graph in Fig. 10 represents the



(a) (b)

Fig. 9. (Color online) Images and labels for (a) training and (b) testing.

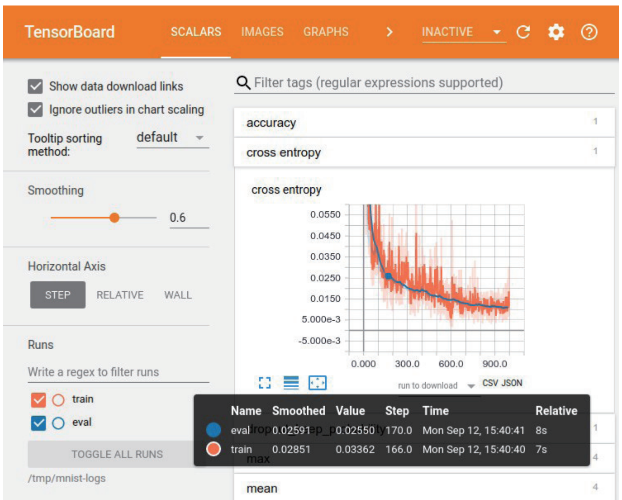


Fig. 10. (Color online) TensorBoard graph.

Cross Entropy graph, which is the result of training the model. The results are obtained from the differential value between the distributed probability regarding the actual data and the distributed probability regarding the utilized model as

$$H(p,q) = -\sum p \log q \,. \tag{1}$$

Then, when the differential value between the distributed probability regarding the actual data and the distributed probability regarding the utilized mode is improved to be as low as possible by utilizing the loss function, this particular loss function becomes the differential value between the output and the target, which is utilized to estimate and update the weight of the model.

Typically, in the training process, the learning data is fed to AI, then AI analyzes the results compared with the labels. If the results are largely different from the labels, the weight will be constantly updated until the results become close to the labels. The differential values of AI recognition results are called loss values. The loss values can be divided into two sections, which are the loss from the locating process and the loss from the classification process. After running the test for 10000 repetitions, the results show that the loss values are 0.02 and 0.04. Figure 11 delineates the graph of the results from AI, clearly showing that the graph is gradually converged into the horizontal axis.

3.4 Learning results for position by AI

The AI model is composed of the learning process and the adjustment of the appropriate weight. The test results for the analysis of the positions of wheels and rails are shown in Fig. 12.

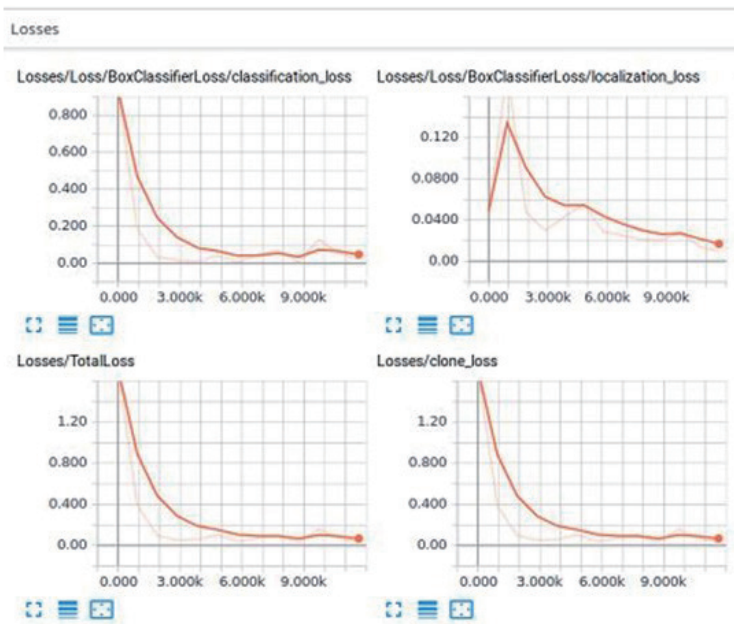


Fig. 11. (Color online) Resulting loss determined by TensorBoard running for 10000 repetitions.

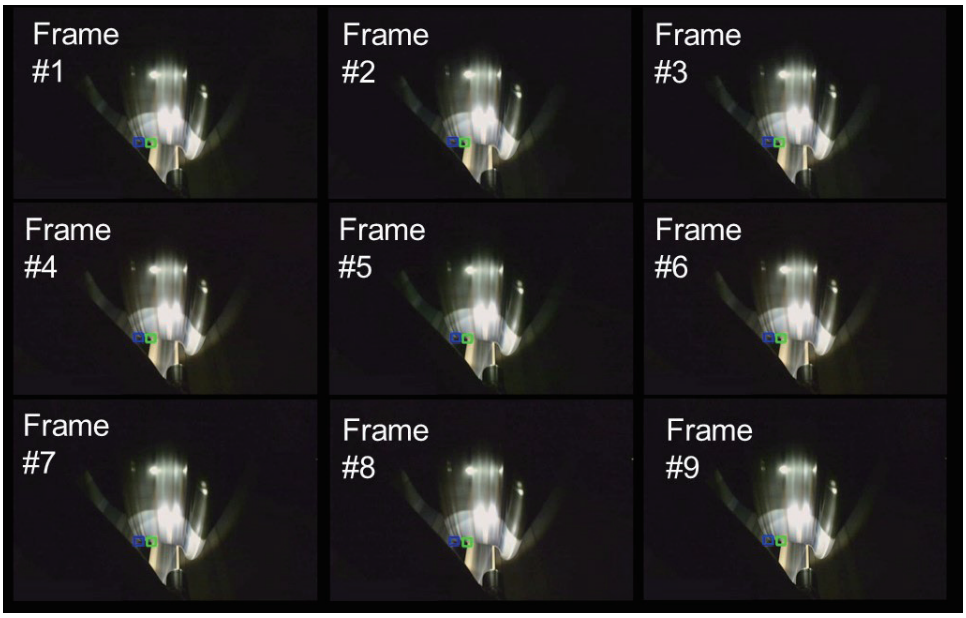


Fig. 12. (Color online) Results of wheel and rail edge detection.

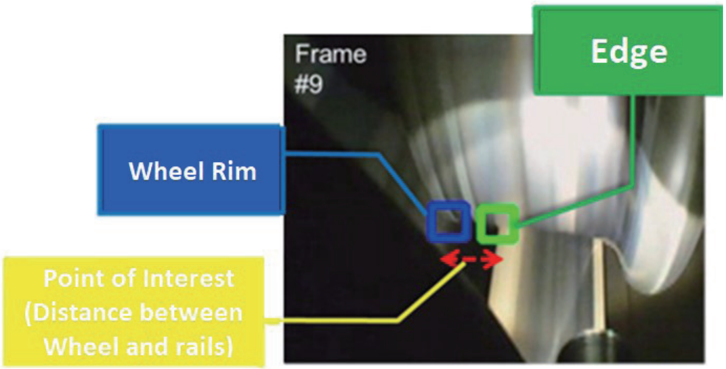


Fig. 13. (Color online) Results of analysis of wheel and rail positions by AI.

In Fig. 13, the blue square shows the wheel rim and the green square shows the trough. The distance between the two areas was analyzed to assess safety and railway conditions by testing using video signals of about 90 s length from the train tracks.

The data of 2699 image signals for rail position are divided into three ranges, as shown in Figs. 14–16.

The analysis results of the first range of signals from the 1st to the 1000th image are shown in Fig. 14. It was found that the average distance between wheels and rails was 2.40 cm. The standard deviation was 0.2854. Two wheels moving near the edge of the rail were detected two times.

The analysis results of the second range of signals from the 1001st to the 2000th image are shown in Fig. 15. It was found that the average distance between wheels and rails was 2.54 cm. The standard deviation was 0.2408. No wheel moving near the edge of the rail was detected.

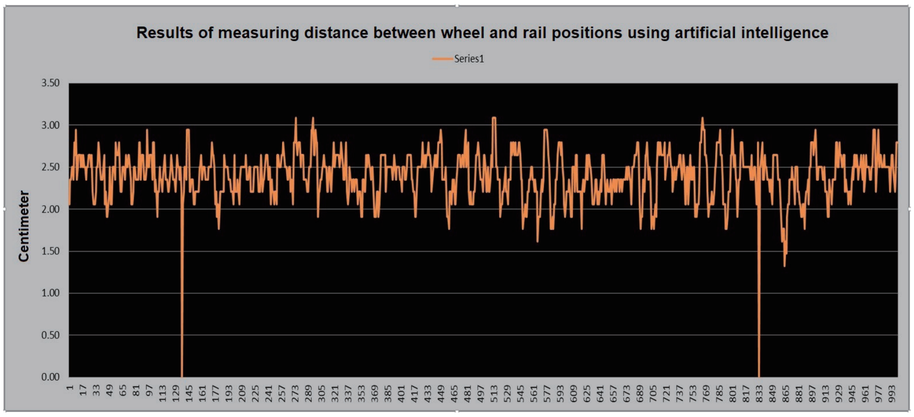


Fig. 14. (Color online) Results of measuring the distance between the wheel and rail positions for image signals from 1 to 1000.

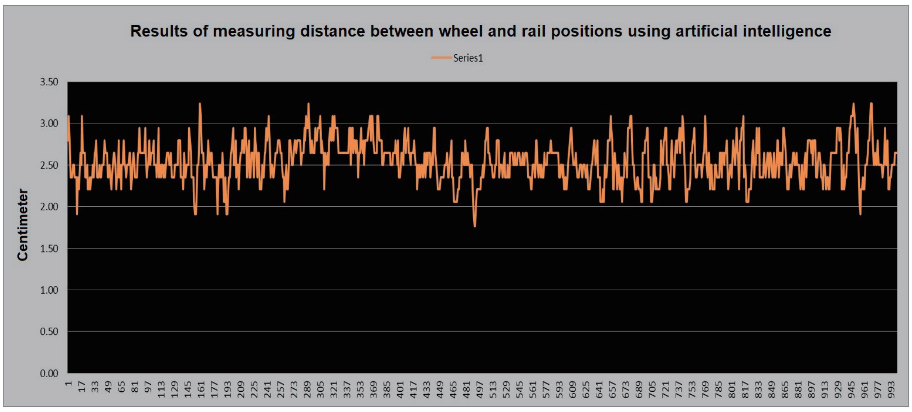


Fig. 15. (Color online) Results of measuring the distance between the wheel and rail positions for image signals from 1001 to 2000.

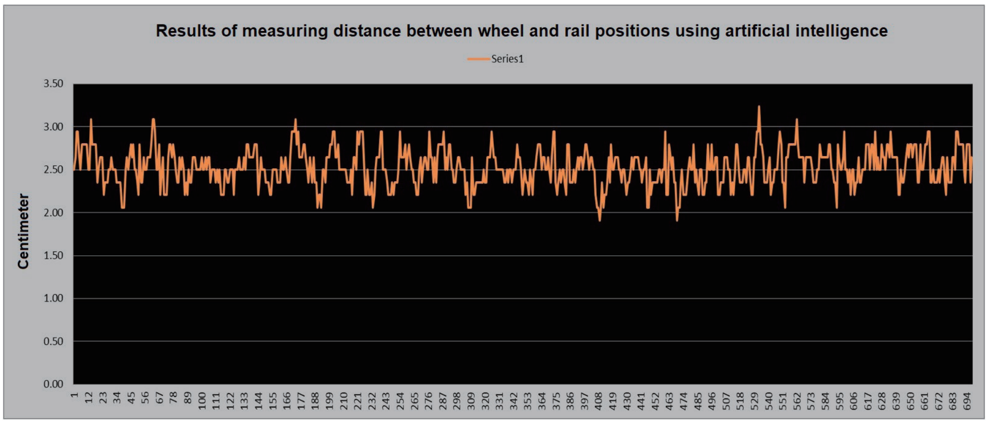


Fig. 16. (Color online) Results of measuring the distance between the wheel and rail for image signals from 2001 to 2699.

The analysis results of the last range of signals from the 2001st to the 2699th image are shown in Fig. 16. It was found that the average distance between wheels and rails was 2.53 cm. The standard deviation was 0.2145. Again, no wheel moving near the edge of the rail was detected.

Table 1
Mean and standard deviation of analyses results.

Range	Image signals	Mean distance (cm)	Standard deviation
1	1–1000	2.40	0.2854
2	1001–2000	2.54	0.2408
3	2001–2699	2.53	0.2145

From the analysis of these three ranges of data, it was found that the average distance in the first trough was the lowest with an average of 2.40 cm and a standard deviation of 0.2854, and the movement of the wheel near the edge of the draft was detected two times. These results are shown in Table 1.

Table 1 shows that the distances in the second and third ranges have similar averages of 2.54 and 2.53 cm, with average standard deviations of 0.2408 and 0.2145, respectively.

4. Conclusions

The developed AI systems for the visual measurement of the train can be used to examine the safety and railway conditions used in the route information of the train. The AI model for measuring the images of the rail condition and the interaction between wheels and rails shown here is the Faster R-CNN (VGG-16). The results show the effectiveness of the model in determining the rail condition and the efficiency of the assessment of the interaction between wheels and rails by distance measurement. The result of measuring the distance between the wheel and rail positions from three consecutive sets of images shows the efficiency of the evaluation of the interaction between wheels and rails in the three periods. The average distance in the first range was the lowest at 2.40 cm with the standard deviation of 0.2854. The movement of the wheel near the edge of the rail was detected 2 times. In the second and third ranges of image analysis, the average values were similar, being 2.54 and 2.53 cm, with the average standard deviations of 0.2408 and 0.2145 cm, respectively. The results of this experiment can be further analyzed and used to evaluate the safety and conditions of rails in the future.

Acknowledgments

We express special gratitude to the Development Research Center, Test and Transfer Technology of High Speed Train of Thailand for providing research funding.

References

1 S. Oh, S. Park, and E. Joung: Proc. 2009 Int. Conf. Information and Multimedia Technology (IEEE, Jeju Island, 2019) 59–63. https://doi.org/10.1109/ICIMT.2009.34
2 V. Rikhotso, N. Steyn, and Y. Hamam: Proc. 2017 IEEE AFRICON (IEEE, Cape Town, 2017) 1522–1527. https://doi.org/10.1109/ AFRICON.2017.8095708
3 O. Yaman, M. Karakose, and E. Akin: Proc. 2017 Int. Conf. Computer Science and Engineering (UBMK, Antalya, 2017) 713–718. https://doi.org/10.1109/UBMK.2017.8093511

4 Z. Liu, W. Wang, X. Zhang, and W. Jia: Proc. 2010 2nd Int. Asia Conf. Informatics in Control, Automation and Robotics (CAR, Wuhan, 2010) 472–475. https://doi.org/10.1109/CAR.2010.5456793
5 R. Girshick, J. Donahue, T. Darrell, and J. Malik: Proc. 2014 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2014) 580–587. https://doi.org/10.1109/CVPR.2014.81
6 R. Girshick: Proc. 2015 IEEE Int. Conf. Computer Vision (ICCV, Santiago, 2015) 1440–1448. https://doi.org/10.1109/ICCV.2015.169
7 B. T. Nassu and M. Ukai: Proc. IEEE Trans. Intelligent Transportation Systems (IEEE, 2012) 1763–1771. https://doi.org/10.1109/TITS.2012.2204052
8 E. J. Sen, K. D. M. Dixon, A. Anto, M. V. Anumary, D. Mieheal, F. Jose, and K. J. Jinesh: Proc. 2014 Int. Conf. Embedded Systems (ICES, Coimbatore, 2014) 162–165. https://doi.org/10.1109/EmbeddedSys.2014.6953109
9 M. M. Mustafa, A. Hussain, K. H. Ghazali, and S. Riyadi: Proc. 2007 IEEE Int. Symp. Signal Processing and Information Technology (IEEE, Giza, 2007) 632–635. https://doi.org/10.1109/ISSPIT.2007.4458197
10 G. Vladimir, B. Dmitriy, T. N. Win, and I. Evgen: 2017 IEEE Conf. Russian Young Researchers in Electrical and Electronic Engineering (EIConRus, St. Petersburg, 2017) 669–671. https://doi.org/10.1109/EIConRus.2017.7910645
11 A. Kumar, N. Ahuja, J. M. Hart, U. K. Visesh, P. J. Narayanan, and C. V. Jawahar: Proc. 2007 IEEE Workshop Applications of Computer Vision (WACV '07, Austin, TX, 2007). https://doi.org/10.1109/WACV.2007.9
12 Y. Liu: Proc. 2018 11th Int. Symp. Computational Intelligence and Design (ISCID, Hangzhou, 2018) 119–123. https://doi.org/10.1109/ISCID.2018.10128
13 W. Cai., J. Li, Z. Xie, T. Zhao, and K. Lu: Proc. 2018 37th Chinese Control Conference (CCC) 9500–9503. https://doi.org/10.23919/ChiCC.2018.8482613
14 S. Ren, K. He, R. Girshick, and J. Sun: Proc. IEEE Trans. Pattern Analysis and Machine Intelligence (IEEE, 2017) 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
15 Y. T. Li and J. I. Guo: Proc. 2018 IEEE Int. Con. Consumer Electronics-Taiwan (ICCE-TW). https://doi.org/10.1109/ICCE-China.2018.8448674
16 K. Simonyan and A. Zisserman: Proc. 2015 6th Int. Conf. Learning Representations (ICLR 2018). https://arxiv.org/pdf/1409.1556.pdf

About the Authors

Anakkapon Saenthon received his B.S. and M.S. degrees from Narasuan University, Thailand, in 1996 and 2008, respectively, and his Ph.D. degree from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2011. Since 2015, he has been an assistant professor at King Mongkut's Institute of Technology Ladkrabang, Thailand. His research interests are in vision systems, image processing, and manufacturing systems.
(anakkapon.sa@kmitl.ac.th)

Somyot Kiatwanidvilai received his B.S. and M.S. degrees from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 1996 and 2000, respectively, and his Ph.D. degree from Asian Institute of Technology (AIT) in 2005. From 2006 to 2012, he was an assistant professor at King Mongkut's Institute of Technology Ladkrabang, Thailand. Since 2013, he has been an associate professor at King Mongkut's Institute of Technology Ladkrabang, Thailand. His research interests are in power electronics, machine drives and systems, electrical power and energy, control systems, and artificial intelligence. (somyot.ki@kmitl.ac.th)

Suparat Pumyoy received her B.S. and M.S. degrees from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2009 and 2014, respectively. Her research interests are in power electronics, genetic algorithms, and vision systems. (suparat.pu@kmitl.ac.th)