

Offline Deep-learning-based Defective Track Fastener Detection and Inspection System

Chen-Chiung Hsieh,^{1*} Ya-Wen Lin,² Li-Hung Tsai,²
Wei-Hsin Huang,³ Shang-Lin Hsieh,¹ and Wei-Hung Hung¹

¹Department of Computer Science and Engineering, Tatung University,
No. 40, Sec. 3, Zhongshan N. Rd., Taipei City 104, Taiwan

²Harbor and Marine Technology, Institute of Transportation,
No. 2, Zhongheng 10th Rd. Wuzi Dist., Taichung City 435, Taiwan

³Department of Media Design, Tatung University, No. 40, Sec. 3, Zhongshan N. Rd., Taipei City 104, Taiwan

(Received April 9, 2020; accepted August 11, 2020)

Keywords: track fastener, defect inspection, deep learning, Yolo model, fastener positioning

Railway track fasteners are used to fasten the railway track onto the crosstie and improve the train's stability and safety. Automatic detection systems have been developed for track safety. Most of these systems deployed line scan sensors to capture high-quality track images. These sensors can capture high-resolution images, but they are also expensive. In addition, the recognition kernels range from traditional computer vision to deep learning methods. In this study, we set up a track fastener sensing device on a flat track car by using general sport cameras and LED lamps to capture images of track fasteners. You Only Look Once (Yolo) v3 is also used instead of earlier convolution neural networks (CNNs) for defect inspection. A cloud server is built for users to queue their captured fastener videos to the first buffer for upload, and uploaded videos can be queued to a second buffer for defective track fastener detection. The trained Yolo v3 neural network classification module is encapsulated as a web application interface (API) for performing the task. In experiments, track fastener videos along a total of 70 km of track were captured with a resolution of 1920×1080 at a speed of up to 35 km/h. Six normal and four defective fastener types were defined for inspection. We split the dataset into 80% for training and 20% for testing. The average precision rates for normal and defective fasteners were 83 and 89%, respectively. Finally, the coordinates of defective fasteners were interpolated from GPS positions recorded by a sport camera. The nearest hectometer stake and the offset of each defective fastener were calculated to assist track workers to find the defective fasteners and fix them.

1. Introduction

A general rail track shown in Fig. 1(a) is a transportation structure composed of ballast, crossties, rails, and fasteners for a train to run on.⁽¹⁾ Fasteners fasten the rails onto the crossties, fixing them when they are subjected to vertical, longitudinal, and lateral forces. The rail fastening system has the functions of buffering the lower structure, distributing the weight,

*Corresponding author: e-mail: cchsieh@gm.ttu.edu.tw
<https://doi.org/10.18494/SAM.2020.2921>

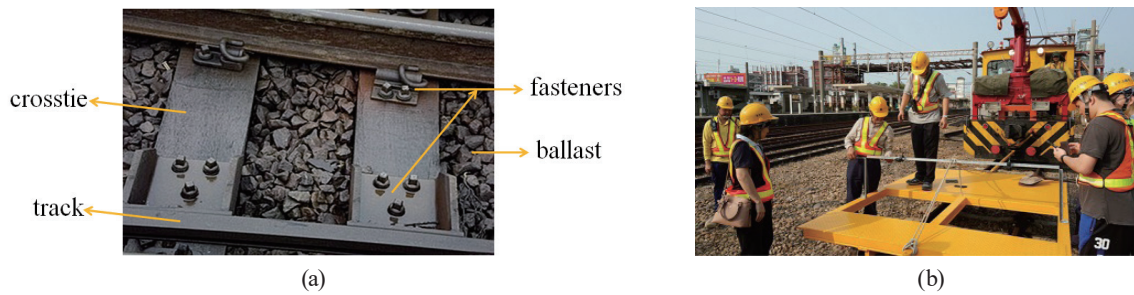


Fig. 1. (Color online) (a) Traditional railway showing track, ballast, crossties, and fixing mechanisms.⁽¹⁾ (b) Our flat car for fastener image sensing.

resisting the disturbance of the rail, and absorbing the sound, so that the rails can be fixed on the track stably. The incorrect functioning of fasteners may cause train derailment.⁽²⁾

To ensure the normal operation of its railroad, Taiwan Railway Company relies heavily on the human visual inspection of the fasteners by workers on a track maintenance vehicle. The inspection and maintenance operations are scheduled at night when there is no train access. However, the results of visual inspection may be limited by the speed of the maintenance vehicle and the view angle. In addition, long-term visual work may cause fatigue and the failure to detect problems. To reduce the workload of staff and ensure the accuracy of inspection, it is necessary to leverage automatic inspection technology to provide more effective solutions. Breakthroughs in deep learning technology have made it possible for computers to learn using neural networks to simulate human thinking, which resulted in rapid advances in computer vision. Although AI may be able to solve many real-world problems, it depends heavily on training data. The more good data available for AI training, the more likely AI will provide good judgment. Therefore, this study is conducted to develop a new method for collecting track images and to choose an appropriate deep learning model for track inspection.

The first step is to set up an image-sensing device for rail fasteners (including two sport cameras and their supporting frames) on a flat car as shown in Fig. 1(b). A GoPro sport camera is selected instead of an expensive line scan image sensor for most track inspection systems owing to its many advantages, such as its compactness, durability, robustness to vibration, and responsiveness. To capture images at night, lamps are mounted on the flat car as low as possible to ensure sufficient brightness for imaging.

As to the convolution neural network (CNN) model, You Only Look Once (Yolo) v3 is chosen owing to its superior performance to Single Shot MultiBox Detector (SSD),⁽³⁾ Faster-RCNN,⁽⁴⁾ and RetinaNet.⁽⁵⁾ Yolo v3 is trained and tested using the collected images of rail fasteners that have been labeled to obtain the detection rates for both normal and defective fasteners. The most time-consuming part was labeling the images. We labelled images of 70 km of rails captured by GoPro with a resolution of 1920×1080 at 60 frames per second (fps). To prove the effectiveness of the system, we found 45950 fasteners that could be searched for by using GPS positions extracted from GoPro images. The nearest hectometer stake and the offset of the detected fastener were calculated to assist track maintenance workers.

This paper is an improved and extended version of our preliminary work cited in Ref. 6. Here, we give a complete and detailed description of our proposed solution. Moreover,

extensive analyses are given and several additional experiments are reported. In summary, the main contributions reported are as follows: (1) an innovative sensing method using general sport cameras installed on a flat track car that can capture images at night is proposed, (2) Yolo v3 is used as the track fastener recognition kernel, which is encapsulated as a thread and works synchronously with the upload thread to achieve high throughput in a cloud service system, and (3) a virtual detection circle (VDC) is also proposed for track workers to quickly locate defective fasteners and their offsets via a hectometer stack. The rest of this paper is organized as follows. In Sect. 2, we introduce related studies on automatic railway inspection. In Sect. 3, we discuss the proposed categories of fasteners and the Yolo v3 model adopted for fastener classification. In Sect. 4, we give an overview of the system and the details of the implementation, and present the experimental results and analysis of the recognition system. In Sect. 5, we conclude this study.

2. Automatic Track Inspection Technologies

Liu⁽⁷⁾ designed a high-speed image capture system that automatically changes the sampling rate of a line-scan camera so that the image resolution of the detected object can be fixed; the wood crosstie positioning system can transmit the detected position of crossties to a computer terminal for engineers to check. Chen *et al.*⁽⁸⁾ proposed a railway monitoring technology for a mass rapid transit system and discussed the real-time identification of rail fasteners from images. The relative positions of the rail and crosstie to a fastener are used as the basis to locate the area of the fastener. Because of the need to detect the rail from the whole image, the execution time of a field-programmable gate array (FPGA) for rail positioning is long owing to the complicated algorithm, which is mainly composed of a gray pixels processing unit, a rail positioning unit, and a synchronous dynamic random access memory.

The Pavemetrics' laser rail inspection system (LRAIL)⁽⁹⁾ of Canada is a recently introduced full-scale system that can be mounted on a vehicle or locomotive. The vehicle/detection speed can reach 180 km/h, and the 3D geometry can be measured at this speed while capturing high-resolution images both during the day and at night. A GPS with an odometer and inertia correction is used for automatic positioning. The detection targets of LRAIL are wood crossties, concrete crossties, fastener inspection, and damage to the rail surface.

Molina Camargo *et al.*⁽¹⁰⁾ collated the most common causes of derailment between 1998 and 2009. After their analysis, the turnout parts near track forks were selected for inspection by using machine vision with a camera. The precision rate of the anchors was only 80%, but the recall rate was 100%. However, their approach was based on traditional pattern recognition methods.

Feng *et al.*⁽¹¹⁾ proposed a probabilistic structure topic model (STM) for modeling fasteners. They trained fastener models using a collection of intact fastener samples. The likelihood was used to measure the similarity between a test fastener and a model. They found that worn fasteners had a lower likelihood than intact ones. The fasteners were classified into three levels on the basis of their likelihood in descending order. The intact fasteners were classified into the high level, partly worn or damaged fasteners were classified into the middle level, and severely worn or missing fasteners were ranked into the low level.

Ladola *et al.*⁽¹²⁾ designed an automatic railway track fault detection system using infrared sensors to detect cracks. Fault locations were recorded by GPS and transmitted using the global system for mobile communication (GSM) short message service (SMS). Ritika *et al.*⁽¹³⁾ proposed a prototype system that combines cameras and GPS to capture rail images and record positions. The camera was carefully designed to withstand the effects of train movement and provide stable images at a speed of about 30 fps. By advanced image analysis and deep learning techniques, the track signals in these camera images were detected and their locations were stored in a database. The railway signal detection system was tested with 150 km of track and 247 signal routes, and the overall accuracy was 94.7%.

Karakose *et al.*⁽¹⁴⁾ proposed a computer-vision-based monitoring method for the detection of faulty tracks. Such a method is being increasingly used in railway systems. The railway condition monitoring process obtained image data, and analysis was carried out with the aid of a computer. In their study, a camera was placed on top of the train to take images of the tracks in front of the train. Edge detection and feature extraction were applied to the images to determine the tracks. The distance between the tracks was used to determine whether there was a fault. The experimental results show that the computer-vision-based method was effective and reliable.

Gibert *et al.*⁽¹⁵⁾ proposed a multidetector to locate the track and fasteners simultaneously. Their design was a full CNN that was trained with 10 classes of materials and produced feature maps with 10 different channels. Their goal was to simultaneously detect the most likely fastener location within each predefined region of interest (ROI), then classify such detections into one of the three basic conditions: background, broken fastener, and undamaged fastener. Then, class labels were assigned for each fastener type (PR clip, e-clip, fast-clip, c-clip, and j-clip).

Wei *et al.*⁽¹⁶⁾ proposed a fastener defect detection and identification method using Dense-SIFT features. They also trained VGG16, a very deep convolution network, for fastener defect detection and recognition. Their results demonstrated that it is possible to detect defective fasteners with a CNN. Finally, Faster R-CNN was used for fastener defect detection to improve the detection rate and efficiency.

Chen *et al.*⁽¹⁷⁾ applied deep convolutional neural networks (DCNNs) in the defect detection of fasteners. Their system cascaded three DCNN-based detection stages in a coarse-to-fine manner, including two detectors to sequentially localize the cantilever joints and their fasteners and a classifier to diagnose the fasteners' defects. They concluded that SSD and Faster R-CNN perform better than Yolo⁽¹⁸⁾ and DPM in terms of accuracy. However, the Yolo network had a much higher detection speed and a shorter training time.

In this paper, we proposed the use of a DCNN for rail track fastener detection. We tested R-CNN, Faster R-CNN, SSD, and Yolo v3. Yolo v3⁽¹⁹⁾ was finally selected owing to its superior performance. Furthermore, both normal and defective fasteners are counted as one type of fastener. One advantage of such classification is that the classification task could be finished within only one stage, thus greatly reducing the training and prediction times. Six normal and four defective fastener types were defined for inspection.

3. Proposed System Architecture

Because most inspection activities are conducted at night, we use a high-speed video camcorder with night vision (or adequate supplementary lighting) to achieve the purpose of this study. The captured images along with their GPS positions can be transmitted offline to the back-end deep-learning-based AI server for defective fastener identification. The system architecture shown in Fig. 2 is divided into a front-end control system running on a flat track car and a back-end cloud server for storage and classification. To ease the operation of the former system for track workers, we designed it as an offline subsystem owing to the unreliability of wireless network communication in rural areas.

The front-end control system mainly sets up the recording and storage function of the videos, and at the same time provides a manual operation interface for video recording and viewing. The image-sensing device is set to operate with a resolution of 1920×1080 at 60 fps during recording. It is expected to work continuously for at least 5 h. In addition to recording fastener images, some necessary data such as time, speed, and GPS position are also recorded. The recorded video is then uploaded offline to the back-end server for storage and deep-learning-based defective fastener identification.

There are four request modes in the back-end server: video upload, fastener classification, fastener enquiry, and fastener positioning. When connecting to the server, it is necessary to have enough input/output throughputs to complete fastener classification within a reasonable time. Two buffers are used to guarantee a fluent workflow despite the inherent large size of video data and long classification time. The fastener enquiry mode provides users with the results of previous analyses to assist judgment. After the uploaded video is classified, the results can be queried by the captured date/time and fastener type. The identified defective fasteners can also be displayed on Google Maps by utilizing their GPS coordinates.

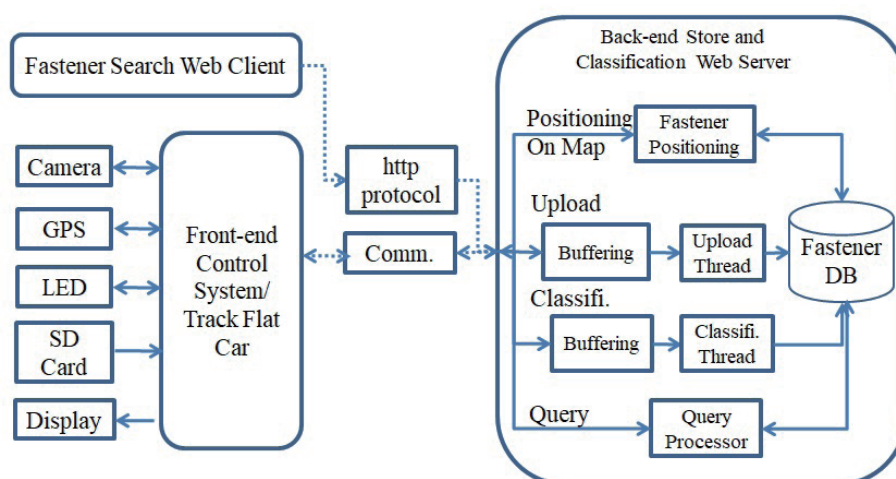


Fig. 2. (Color online) The proposed system architecture includes a front-end client and a storage/classification server.

3.1 Sensing method

We adopt a flat car as the main vehicle, which is cheap and easy to produce compared with a train coach. The designed flat car shown in Fig. 1(a) follows the current specifications of Taiwan rails. It is powered by a train locomotive, and its robustness has been proved in many tests on the side railway near Dajia Station. The flat car can run at nearly 50 km/h and still capture high-quality images during the daytime. If the expected operating environment for fastener inspection is on a train with a speed of 120 km/h and a resolution of 0.5 m/image is required, then a frame sensing rate of 67 images per second is the lower bound. Therefore, if the speed of the maintenance vehicle is only about 30 km/h at night, the high-speed video camera can clearly satisfy this requirement. A high-speed video camera of more than 60 fps is enough to obtain the railway track images. In addition, taking into account the external environment, the high-speed video camera should be waterproof and shockproof, and should ideally be operated by remote control.

We selected GoPro Hero7 Black as the video camera, which is a popular sport video camera, especially for its 4K high quality, light weight, and ease of use with water and dust resistance. Therefore, its suitability for railway fastener inspection was assessed. In addition to its excellent antivibration ability, GoPro has a wide field of vision and GPS positioning. The camera also provides support for real-time streaming, Python application interface (API), and other functions, but more importantly, a variety of optional accessories allow it to be clipped onto a train maintenance vehicle or flat car according to our needs.

GoPro is fixed on an aluminum frame and is adjustable as shown in Fig. 3(a). It remained firmly attached to the frame after several field tests. However, thermal shutdown occurred during long 4K recordings. We solved this problem by unplugging the battery and using an external power supply, which allowed GoPro to operate for more than 4 h.

3.2 Lighting

When choosing the lighting equipment, we considered the brightness of the light. To obtain similar clear images in daylight, we first selected 10 W engineering-edition white LED searchlights. It was found that the brightness of the 10 W LED searchlights was insufficient

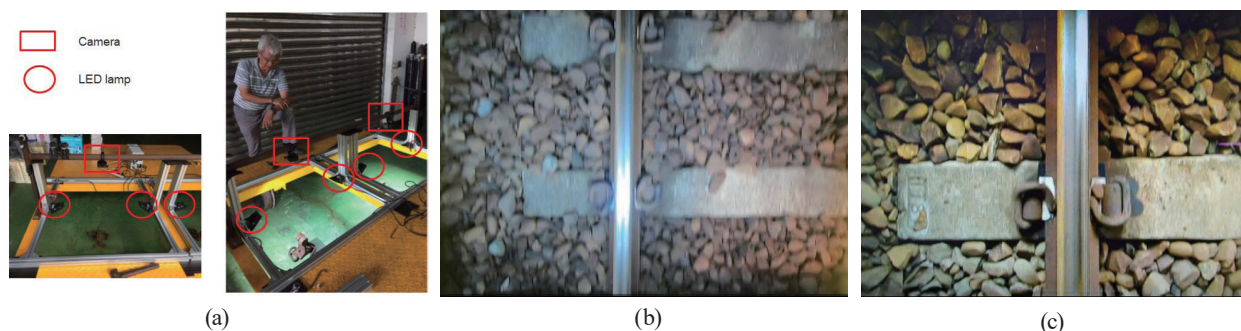


Fig. 3. (Color online) (a) Cameras and lamps fixed on aluminum frames. (b) Lighting by 2×10 W LED lamps. (c) Lighting by 2×200 W LED lamps at lower positions.

[Fig. 3(b)], so 200 W LEDs were used instead. Figure 3(c) shows the improved clarity of images when using two 200 W LEDs and lowering the lamps close to the track for brighter imaging.

The lamps can be adjusted to verify the effects at different lighting positions. Because there are trains running on the main track every day, the surface of the track is smooth with a mirrorlike finish. If the light is directly illuminated over the track, it may be difficult to obtain a clear picture of the rail fastener owing to the reflection. In addition, lighting from the left or right side will cause a shadow, meaning that the position of the lamps should be adjusted repeatedly. To facilitate the adjustment of lamps according to the actual situation, the bracket adopted was an aluminum extrusion frame, which allowed the optimal positioning so that the left and right brightness values were consistent, resulting in minimum shadow.

Regarding the power supply, the original intention was to use the power supply from the train. However, in view of the power load and the safety of the wire connection, we decided to use the power supply of a generator with an uninterruptible power supply system. The voltage-stabilizing function of the uninterruptible power supply system can ensure high-quality power for lighting, GoPro, and other equipment.

3.3 Classification kernel

From the development of R-CNN, Fast R-CNN, and Faster R-CNN to Yolo, Yolo's attractive feature is direct end-to-end rather than multiple stages for object detection. It predicts all information about the target object, including the bounding box coordinate location of the object, the confidence value of the contained object, and the category to which the object belongs, using the whole picture as the input. Yolo v1⁽¹⁸⁾ was fast enough to achieve real-time identification but not accurate enough to predict the position and precision of small objects. Some of the problems of Yolo v1 were solved in Yolo v2 by introducing an anchor box in Faster R-CNN. Yolo v3⁽¹⁹⁾ shown in Fig. 4 does not have any major improvements, but includes refinements based on ideas from other studies, greatly improving its performance.

The advantages of Yolo v3 are its light weight and high efficiency of identification. However, training the neural network model requires a large number of samples. The number of samples determines the generalization ability and accuracy of neural network models. As much training data as possible of both normal and faulty samples are needed. We found that Yolo outperforms other object recognition neural network models in terms of its relative efficiency and accuracy.

3.4 Back-end classification server

The front end receives/displays images from GoPro and packages them to the back end for classification. Regarding the back-end server, if there are not enough GPUs, there will be a wait of at least half a day whenever a large amount of data arrives. For real-world applications, it is recommended to use two or more RTX2080Ti cards installed at the back end as an appropriate balance between efficiency and cost.

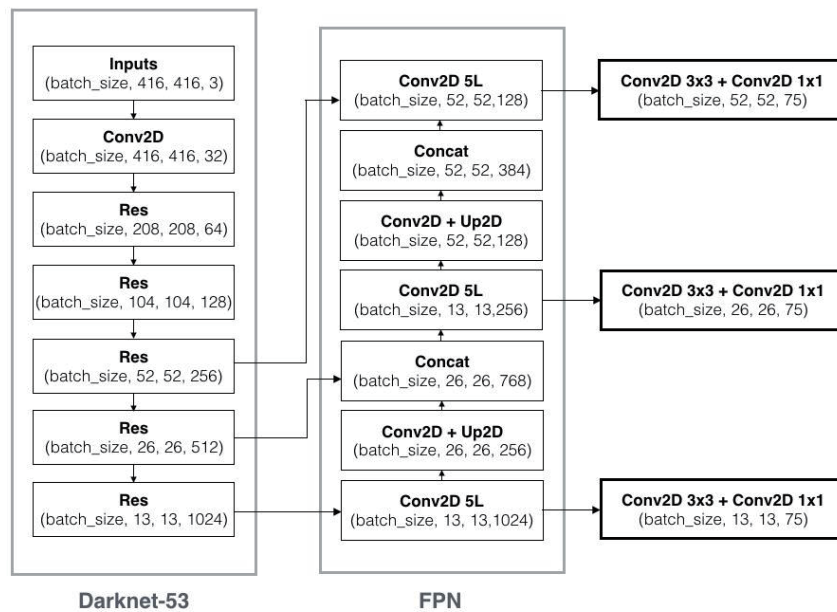


Fig. 4. Yolo v3 neural network model.⁽¹⁹⁾

Our first task is to encapsulate all the required classification-related modules into a Python class. In this study, the pretrained Yolo v3 model is serialized into a pickle object. Thus, in the initial method of the Python class, we load the pretrained model. This object can then be used for prediction via a web application.

Our second task is to adopt Flask to expose “classify_fastener” and “train_fastener” as REST APIs.⁽²⁰⁾ Both are mapped to invoke the classifier class and the appropriate methods within it. The returned objects from both APIs are JavaScript Object Notation (JSON) objects of the results from the machine learning model. These REST APIs that we created can be wired into our web application.

3.5 Defective fastener positioning

The most commonly used positioning method is GPS. However, there are SN interference code problems with the commercial GPS signal, so the positioning error is generally 15–20 m. Although the positioning error can be reduced to about 10 m through the AGPS, the distance between the two tracks of the railway is 1.067 m, so there will be considerable error owing to the positioning accuracy of the GPS.

A photograph or video taken with GoPro contains GPS information that can be parsed directly using GoPro Mobile apps. The GPS data is embedded in an MP4 file called GPX,⁽²¹⁾ as shown in Fig. 5(a), which is currently the most commonly used format. The so-called GPX format stores GPS information in XML format. Using GoPro-captured images, the resulting GPS data, with an error of about 5 m, has almost sufficient accuracy for practical usage.

We propose the use of a hectometer stake as the center of the VDC to decrease the positioning error. At present, there are such stakes at 100 m intervals along the railway. When

the maintenance vehicle passes through these VDCs, its GPS location can be determined. Using the GPS position (x_f, y_f) of the fault fastener, the nearest 100 m stake (x_h, y_h) can be found and then track workers move by a distance of $(x_f - x_h, y_f - y_h)$ to the fastener. Figure 5(b) shows an image of a detected defective e-clip in the upper-left part, which is displayed on the map as a blue pin, where the gray pins are hectometer stakes. Workers can easily find the defective e-clip via the VDCs as shown in red circles.

4. Experimental Results

4.1 GoPro setup and testing

After fixing GoPro on the flat car, the latter is driven by a power locomotive to and from the side line near Dajia Station, Taichung City, Taiwan. The maximum speed of the flat car is 50 km/h. During driving, we observed that GoPro vibrated and we were concerned about the quality of the image. However, the images viewed after the shooting were clearer than expected. After the initial testing, it was confirmed that GoPro could be used to obtain the clear images needed for the study. After that, the Dajia branch of Taiwan Rail Company carried out a long-term field test to determine the stability of GoPro. It was found that GoPro could record continuously for more than 7 h without shutting down due to overheating.

4.2 Fastener types and image collection

A large number of samples are needed for an AI identification system. With the assistance of staff at Dajia Station, fastener image datasets near equilateral turnouts, double-opening reverse switches, and bifurcation sides of left-opening articulated turnouts of the railway were provided. All fastener types could be found near these areas. The images of fasteners were provided by Dajia Station; then, our research team labeled them for training and testing. The experts at Dajia Station were asked to review the status of the fastener classification and to continuously update and collect more data. Finally, six normal and four defective fastener types were defined for inspection as shown in Table 1. There were no images for some fault fastener types owing to their rare occurrence.



Fig. 5. (Color online) (a) Example of GPX file. (b) GoPro-recorded track video file displayed using Google Maps.

AI model evaluation

To prepare Yolo’s training data, it was necessary to mark the bounding box of the fastener, which is used to locate the target object in the image, and to label its corresponding type, which is the annotation of the image. For a trained Yolo model, the evaluation method mainly adopts common metrics of object detection, such as intersection over union (IoU) and mean average precision (mAP).

IoU

As given by Eq. (1), the intersection of the predicted result and the ground truth over the union is the most commonly used indicator for predicting an object. If, for example, $IoU > 0.5$, the prediction of a bounded box is a success if its IoU is more than a half.

$$IoU(A, B) = (A \cap B) / (A \cup B) \tag{1}$$

mAP

The accuracy of each category is calculated from Eq. (2) using IoU as the criterion, usually $IoU = 0.5$. $TP(c)$ is a true positive in class c , which means that the predicted proposal is consistent with the ground truth (the species is correct or the overlap is sufficiently high). On the other hand, $FP(c)$ is a false positive in class c , which means that the predicted proposal is not consistent with the ground truth (there is a type error or the overlap is insufficient). Then, mAP , the average of all calculated accuracies over all classes, is calculated using Eq. (3). For example, $mAP@[0.5: 0.95]$ denotes the integration of all accuracies at intervals of 0.05 from 0.5 to 0.95.








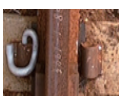


$$Accuracy(c) = TP(c) / (TP(c) + FP(c)) \tag{2}$$

$$mAP = \frac{1}{|classes|} \sum_{c \in classes} \frac{TP(c)}{(TP(c) + FP(c))} \tag{3}$$

Field test on lateral line

For the field test, a short distance of railway track (near Dajia Station) within the jurisdiction of Taichung Railway Engineering Department was selected, and the images captured by GoPro Hero7 Black were used for fastener identification. When a defective fastener was found, GoPro GPX was used to mark its location. After training the Yolo v3 model with 16677 objects, we

Table 1
(Color online) Fastener types for inspection.

	e_clip on wood crossie	e_clip on concrete crossie	Spike	Fishplate	Slide-bed plate	Guard rail plate
Normal						
Defective					NA	NA

tested it using 3294 objects. The results showed a *TP* for 3114 positive samples, an *FP* for 378 negative samples, and an *FN* for 180 positive samples. The precision is therefore 0.89% and the recall rate is 0.95 as shown in Table 2.

Field test of main line

The main line could not be accessed at night so we recorded videos of 70 km of track with lighting for this test. A total of 38 MP4 files were recorded, each about 4 GB. The labeling of all the videos took about one month and involved a total of 25 persons. However, photographs that were too vague or unclear were skipped. The filtered and labeled images were saved as our dataset. Table 3 shows the testing results, which proved the feasibility of our approach. Note that there were no defective spikes or defective slide-bed plates found. Some recognition results are shown in Fig. 6.

Table 4 shows a comparison with some state-of-the-art systems described previously. Most systems restrict their applications to specific rails without ballasts to gain higher accuracy. They also need to locate the fastener in the image first and then classification could be performed. The main contribution of our study is the development of a fastener classification system with no limits on the rail tracks. The collected fastener data include ballasted, no-ballasted, and covered, on daytime, nighttime, and rainy days. The proposed sensing method is not only

Table 2
Field test of lateral line.

Fastener type	Precision
e_clip on wood crosstie (normal/defective)	84.82%/NA
e_clip on concrete crosstie (normal/defective)	88.27%/89.91%
e_clip covered	70.68%
Spike (normal/defective)	79.52%/NA
Fishplate (normal/defective)	84.82%/NA
Slide-bed plate (normal/defective)	86.03%/NA
Guardrail plate (normal/defective)	86.18%/NA

Table 3
Field test of main line.

Fastener type	Average precision (%)	<i>TP</i>	<i>FP</i>
e_clip on wood crosstie (defective)	86.84	3481	425
e_clip on wood crosstie (defective)	85.21	184	19
e_clip on concrete crosstie (normal)	83.94	35192	7799
e_clip on concrete crosstie (defective)	72.87	737	126
e_clip covered	76.18	247	59
Spike (normal)	88.44	31	1
Spike (defective)	NA	NA	NA
Fishplate (normal)	80.92	41	13
Fishplate (defective)	53.65	34	28
Slide-bed plate (normal)	85.10	525	56
Slide-bed plate (defective)	NA	NA	NA
Guardrail plate (normal)	85.72	261	47
Guardrail plate (defective)	74.61	6	1

Detection count = 94522, unique truth count = 45954, precision = 0.83, recall = 0.89, F1 score = 0.86, *FN* = 2850, *TP* = 40724, *FP* = 8569, *FN* = 5230, average *IoU* = 70.27%, and (*mAP@0.50*) = 79.40%.



Fig. 6. (Color online) Some test images. (a) Normal e_clips on concrete cross-tie. (b) Left e_clip is missing while the e_clip on the right is normal.

Table 4
Comparison with some state-of-the-art systems for track inspection.

Source	Fastener types #	Sensor type /speed	Recognition kernel	Accuracy (%)	Restrictions	Recognition speed
Ref. 8	2	Line scan/NA	Traditional CV	NA	Clean track	Not real time
Ref. 9	9	Laser/180 km/h	NA	99.28	Clean track	Real time
Ref. 10	2	<10 km/h	Traditional CV	>80	Clean track	Near real time
Ref. 11	6	HR Cam/262 km/h	Probability mode	95.2	Align and clean	Real time
Ref. 15	13	HR Cam/NA	CNN + SVM	99.91	Align and clean	NA
Ref. 16	4	HR Cam/NA	Faster R-CNN	97.9	Align and clean	0.23 s/frame
Ours	13	SportCam/30 km/h	Yolo v3	83	None	Real time

innovative but also of low cost. Furthermore, we provided an inspection process that positions the detected defective fasteners to assist workers repairing the fasteners.

5. Conclusions

This study was to conducted implement a flat track cart suitable for capturing images of Taiwan's railway system and to use a high-speed video recording device for rail fastener inspection. An expandable database for different types of rail fastener and their corresponding defects has been established, and the classification system is now in operation as a cloud service. The feasibility of the system has been verified in practice. In future works, we hope to integrate front-end image sensing, image processing, and back-end fastener recognition together as an automatic system, which is expected to be deployed for the inspection of railway lines in Taiwan. The final objectives of this study include the (1) rapid deployment, (2) lower cost, and (3) automatic safety inspection of railways to reduce human labor.

Acknowledgments

This work was supported by grants from the Harbor and Marine Technology Center (HMTC), Institute of Transportation (IOT), Ministry of Transportation and Communications (MOTC) under Grant No. MOTC-IOT-108-H1DB001i.

References

- 1 Wikipedia: Track (rail transport), [https://en.wikipedia.org/wiki/Track_\(rail_transport\)](https://en.wikipedia.org/wiki/Track_(rail_transport)) (accessed June 2020).
- 2 Wikipedia: https://zh.wikipedia.org/wiki/Taiwan_Railway_south_loop_continuous_destruction_event (accessed January 2019).
- 3 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg: Computer Vision – ECCV 2016. Lecture Notes in Computer Science, 9905. Springer, Cham (ECCV, 2016) 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
- 4 S. Ren, K. He, R. Girshick, and J. Sun: Extended tech report, arXiv e-prints (2015). <https://ui.adsabs.harvard.edu/abs/2015arXiv150601497R>
- 5 T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar: arXiv:1708.02002 (2017). <https://arxiv.org/pdf/1708.02002.pdf%E3%80%82>
- 6 Y. Lin, C. Hsieh, W. Huang, S. Hsieh, and W. Hung: IEEE Eurasia Conf. IoT, Communication and Engineering (ECICE, 2019) 187–190. <https://doi.org/10.1109/ECICE47484.2019.8942769>.
- 7 Y. Liu: Master Thesis of the Dept. of Electric Engineering, National Taiwan Science and Technology University (2010). <https://hdl.handle.net/11296/q8w3e7>
- 8 N. Chen, S. Shieh, and Y. Liu: Grant No. NSC 97-2221-E011-151-MY3, Ministry of Science and Technology, Taiwan (2008).
- 9 R. Fox-Ivey, T. Nguyen, and J. Laurent: Department of Transportation, U.S. Federal Railroad Administration, Office of Research, Development and Technology, Washington, Technical Report, <https://railroads.dot.gov/sites/fra.dot.gov/files/2020-04/Pavemetrics%20Phase%20202.pdf> (accessed July 2020).
- 10 L. Molina Camargo, E. Resendiz, J. Hart, J. Edwards, N. Ahuja, and C. Barkan: University of Illinois at Urbana Champaign, Grant No. DTRT0-G-005 of the U.S. Department of Transportation, Final Report (2011).
- 11 H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen: IEEE Trans. Instrum. Meas. **63** (2014) 877. <https://doi.org/10.1109/TIM.2013.2283741>
- 12 A. Ladola, C. Parekh, D. Patel, and H. Bhagatwala: Int. Res. J. Eng. **5** (2018) 1543.
- 13 S. Ritika, S. Mittal, and D. Rao: arXiv-prints (2017). <https://ui.adsabs.harvard.edu/abs/2017arXiv171206107R>
- 14 M. Karakose, O. Yaman, M. Baygin, K. Murat, and E. Akin: Int. J. Mech. Eng. Robot. Res. **6** (2017) 22.
- 15 X. Gibert, V. Patel, and R. Chellapp: IEEE Trans. Intell. Transport. Syst. **18** (2017).
- 16 X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li: Eng. Appl. Artif. Intell. **80** (2019) 66. <https://doi.org/10.1016/j.engappai.2019.01.008>
- 17 J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han: IEEE Trans. Instrum. Meas. **67** (2018) 257. <https://doi.org/10.1109/TIM.2017.2775345>
- 18 J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: arXiv:1506.02640 (2015). <https://ui.adsabs.harvard.edu/abs/2015arXiv150602640R>
- 19 J. Redmon and A. Farhadi: arXiv:1804.02767 (2018). <https://arxiv.org/abs/1804.02767v1>
- 20 REST API Tutorial: <https://restfulapi.net/rest-api-design-tutorial-with-example/> (accessed June 2020).
- 21 GPS Exchange Format: https://en.wikipedia.org/wiki/GPS_Exchange_Forma (accessed June 2020).

About the Authors



Chen-Chiung Hsieh received his B.S., M.S., and Ph.D. degrees from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1986, 1988, and 1992, respectively. From 1992 to 2004, he was with the Institute for Information Industry (III) as a vice director. From 2004 to 2006, he was a senior director at Acer Inc. He is presently a full professor in the Department of Computer Science and Engineering at Tatung University, Taipei, Taiwan. His research is mainly focused on image and multimedia processing.
(cchsieh@gm.ttu.edu.tw)



Ya-Wen Lin received her M.S. degree from National Taiwan University of Science and Technology, Taiwan, in 1999. From 2002 to 2006, she was an assistant researcher at the Institute of Transportation, Taiwan. From 2006 to 2011, she was a deputy researcher at the Institute of Transportation. Since 2011, she has been a researcher at the Institute of Transportation. Her research interests are in safety issues related to bridges, rails, and slopes.

(yawen@mail.ihmt.gov.tw)



Li-Hung Tsai received his B.S., M.S., and Ph.D. degrees from the Department of Hydraulic and Ocean Engineering, National Cheng Kung University, Tainan, Taiwan, in 1989, 1991, and 2003, respectively. He is presently a director at the Institute of Transportation. His research is mainly focused on ocean engineering, harbor engineering, and navigation safety.

(ali@mail.ihmt.gov.tw)



Wei-Hsin Huang received his B.S. degree from the Department of Mechanical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan, in 1995 and his M.S. and Ph.D. degrees from the Department of Mechanical Engineering, National Central University, Taoyuan, Taiwan, in 1997 and 2002, respectively. He is presently a full professor in the Department of Media Design, and the dean of research and development at Tatung University, Taipei, Taiwan. His research is mainly focused on system integration, including software and hardware. (wshuang@gm.ttu.edu.tw)



Shang-Lin Hsieh received his B.S. degree from National Chiao Tung University and his M.S. and Ph.D. degrees from National Taiwan University, Taiwan, in 1985, 1991, and 2000, respectively. From 1987 to 1989, he was an assistant researcher at the Industrial Technology Research Institute (ITRI), Hsinchu, Taiwan. Currently, he is an associate professor in the Department of Computer Science and Engineering at Tatung University, Taipei, Taiwan. His research interests are multimedia technology, human posture recognition, and fall detection. (slhsieh@ttu.edu.tw)



Wei-Hung Hung received his B.S. and M.S. degrees from the Department of Computer Science and Engineering, Tatung University, in 2018 and 2020, respectively. His research interests include image processing and deep learning. (whhungbill@gmail.com)