

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354967459>

Real-time object detection technology in railway operations

Article in HKIE Transactions · September 2021

DOI: 10.33430/V28N3THIE-2020-0028

CITATIONS

0

READS

329

4 authors, including:



Zhangyu Wang

Beihang University (BUAA)

30 PUBLICATIONS 168 CITATIONS

SEE PROFILE

Real-time object detection technology in railway operations

Rock K C Ho¹, Zhangyu Wang², Simon S C Tang¹ and Qiang Zhang³

¹ MTR Corporation Limited, Hong Kong, People's Republic of China

² School of Transportation Science and Engineering, Beihang University, Beijing, People's Republic of China

³ Traffic Control Technology Corporation Limited, Beijing, People's Republic of China

ABSTRACT

Development of new technology to enhance train operability, in particular during manual driving by real-time object detection on track, is one of the rising trends in the railway industry. The function of object detection can provide train operators with reminder alerts whenever there is an object detected close to a train, e.g. a defined distance from the train. In this paper, a two-stage vision-based method is proposed to achieve this goal. At first, the Targets Generation Stage focuses on extracting all potential targets by identifying the centre points of targets. Meanwhile, the Targets Reconfirmation Stage is further adopted to re-analyse the potential targets from the previous stage to filter out incorrect potential targets in the output. The experiment and evaluation result shows that the proposed method achieved an Average Precision (AP) of 0.876 and 0.526 respectively under typical scenario sub-groups and extreme scenario sub-groups of the data set collected from a real railway environment at the methodological level. Furthermore, at the application level, high performance with the False Alarm Rate (FAR) of 0.01% and Missed Detection Rate (MDR) of 0.94%, which is capable of practical application, was achieved during the operation in the Tsuen Wan Line (TWL) in Hong Kong.

KEYWORDS Object detection; railway safety; signalling system; convolutional neural network; classification

CONTACT Rock K C Ho  kincho@mtr.com.hk

Received 27 March 2020

1. Introduction

The signalling system has been playing an essential role in the modern railway industry. For railway lines which are equipped with signalling system, train traffic regulations can be guaranteed and reliable automatic train operation can be achieved. Despite the robustness of the signalling system, there are constraints which the system cannot fully cover. For example, a typical design of the signalling system consists of equipment installation on trains as well as tracksides, and the coverage of the signalling system is bound by the pre-defined sections in the railway network. Therefore, trains operating outside mainline sections (e.g. the siding area, or the depot area) may not benefit. Besides, when critical equipment of the signalling system is under failure, degraded operation will result. In these situations, conventional methods of manual train operation which relies heavily on the human line-of-sight judgement will result. In the light of the situation, the development of new technologies to assist train operators in enhancing the operability and reliability in manual mode driving is active in the industry. Real-time object detection in the railway has been one of the hottest topics (Ye et al., 2018).

Real-time object detection has been studied and achieved by different approaches and designs (Berg et al., 2015). proposed adopting train-mounted thermal cameras to detect obstacles on or near the rails in front of the train; Karaduman (2017) proposed adopting image processing and laser obstruction to identify obstacles in front of an operating train by installing cameras, laser distance meters and position sensors on-board with interface connection with an emergency braking system. Li et al. (2020) adopted

a method to collect images of the railway scene from the Unmanned Aerial Vehicle (UAV) and divided them into sub-blocks. For each sub-block, a target detection module was adopted for object detection in order to allow detection of objects with smaller dimensions in a railway environment. Amaral et al. installed lidar sensors at the railway level crossings and carried out background modelling for railway-level crossings. In the real-time detection stage, the method adopted an algorithm to detect obstacles as occupied space contrasting with the background model (Amaral et al., 2016). Lüy et al. (2018) installed a multi-line laser radar on the tram, and achieved object recognition and tracking with three sub-module segmentation, classification, and Kalman Filter tracking.

This paper describes how real-time object detection based on convolutional neural network (CNN) was explained and experimented. The proposed method aims to achieve accurate and robust real-time object detection by applying targets reconfirmation after objects generation. Besides, the existence of incorrect potential targets in the outcome can also be minimised. The rest of the paper will illustrate the methodology, the experiment results, and the conclusion.

2. Methodology

2.1. Overview of the framework

The framework of the proposed method is divided into two stages: Targets Generation Stage and Targets Reconfirmation Stage (see Figure 1).

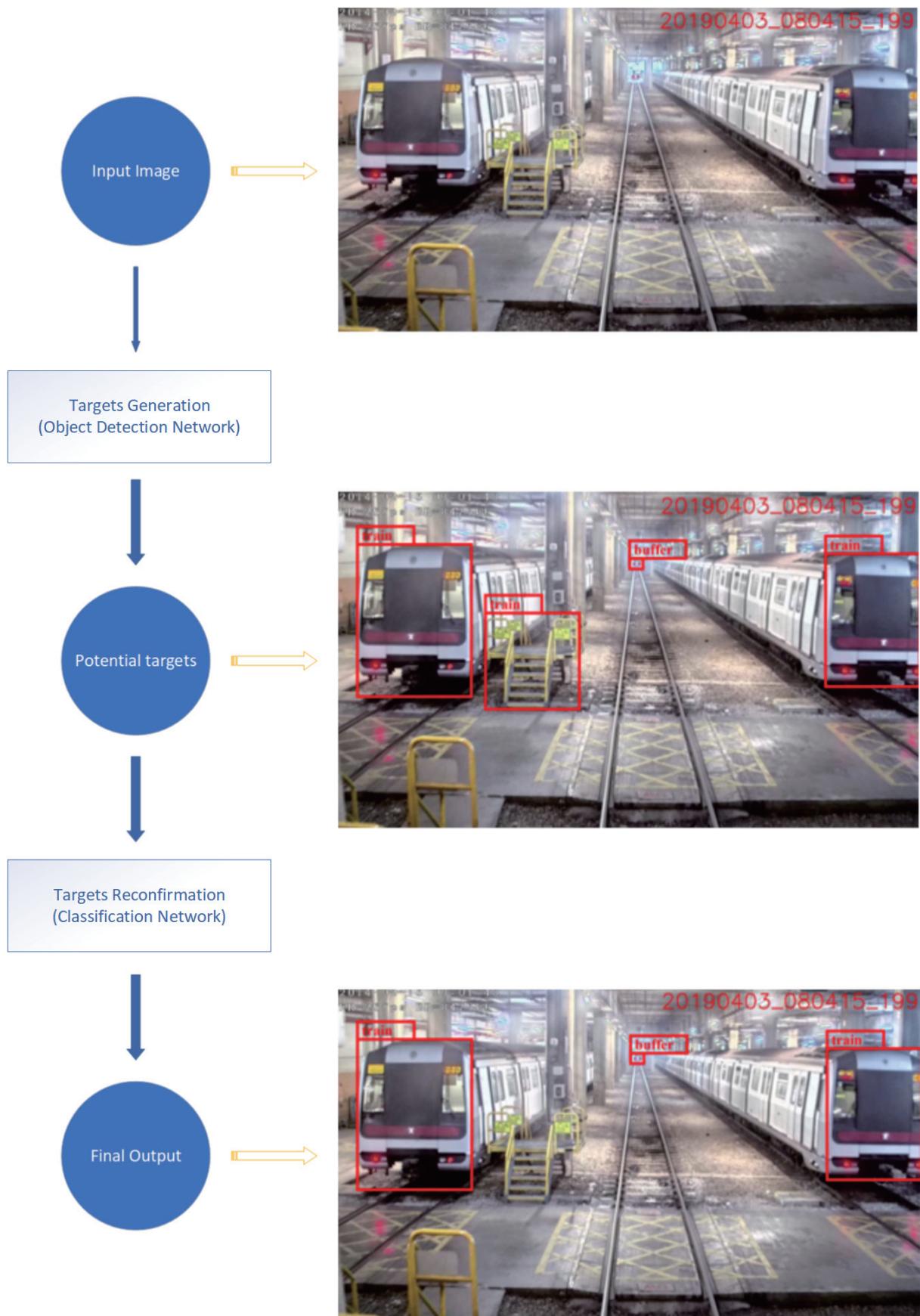


Figure 1. Overview of the framework.

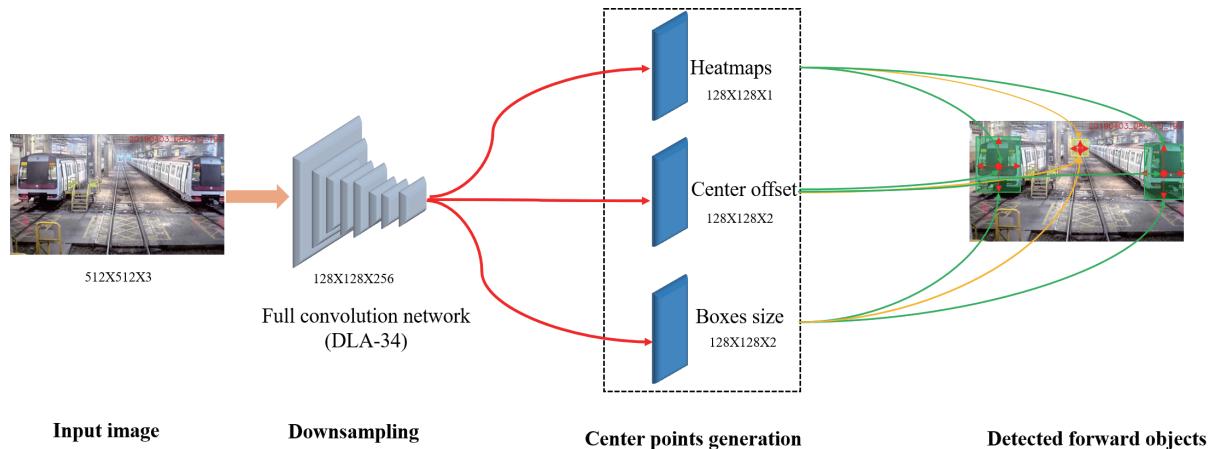


Figure 2. The architecture of the Target Generation Stage.

At the beginning, object targets shall be defined. For this research project, trains and buffer stops were chosen as the object targets for proof of the concept of the proposed methodology of real-time object detection in the railway network.

2.2. Targets generation stage

In this stage, CenterNet (Zhou, Wang and Krähenbühl, 2019), which is a novel object detection framework for object detection, is applied. When an image is processed in this stage, all targets in the image will be extracted. Unlike the method of detecting targets by directly regressing a series of bounding boxes which is widely adopted by one- (Redmon and Farhadi, 2018) or two-stage (Ren et al., 2015) detectors, key point estimation which focuses on identifying the centre points of the targets is adopted in the proposed method.

As illustrated in Figure 2, there are two steps for targets generation: down-sampling and centre points generation. In the down-sampling step, an input image is down-sampled using DLA-34 (Yu, et al. 2018), which is a full convolution network, to result in a feature map with one fourth in resolution of the input image. Then the feature map are to be processed for the centre points generation. During the process, there are three independent network branches responsible for the prediction of the bounding box of the targets, including the heatmaps generation branch, centre offset regression branch, and bounding box size regression branch.

2.3. Heatmaps generation branch

The purpose of the heatmaps generation branch is to generate the centre points of the targets. The number of heatmaps channel shall be defined depending on the number of categories of detection targets. In this project, two channels were adopted for the detection of trains and buffer stops. In this branch, the heatmaps of the down-sampled

feature map are obtained firstly by applying the two-layer convolution method. Then, the peaks of the targets' centre are obtained after applying a 3×3 max pooling operation to the heatmaps. The first 100 of the peaks extracted from the network were adopted as the centre points. Furthermore, the peaks are processed to eliminate those with a value lower than a pre-set threshold. The final heatmaps of the targets are obtained afterwards.

The heatmaps can be expressed as follows:

$$\hat{Y} \in [0,1]^{\frac{W}{R} \times \frac{H}{R} \times C}, \quad (1)$$

where W and H are the width and height of the input image respectively, R is the down-sampling factor, i.e., 4 in the proposed method, and C is the number of categories of the object targets, i.e., 2 in the proposed method.

In the heatmaps, the prediction value, $\hat{Y}_{x,y}$, represents whether it is a target or not, where x and y are the x- and y-coordinates of a respective centre point. When $\hat{Y}_{x_1,y_1} = 1$, the prediction on coordinate x_1, y_1 is identified as a target by the network, and vice versa.

When the network is under training, a gaussian kernel is applied to match the key points to the feature map, i.e.

$$Y_{xyc} = \exp\left(-\frac{(x-\tilde{x}_c)^2 + (y-\tilde{y}_c)^2}{2\sigma_p^2}\right), \quad (2)$$

where \tilde{x}_c and \tilde{y}_c are the x- and y- coordinates of the centre in the ground truth, and σ is the standard deviation to the size of the target.

The loss function of the heatmaps can be expressed as follows:

$$L_k = \frac{1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^\alpha \log(\hat{Y}_{xyc}) & \text{if } Y_{xyc} = 1 \\ (1 - Y_{xyc})^\beta (\hat{Y}_{xyc})^\alpha \log(1 - \hat{Y}_{xyc}) & \text{otherwise;} \end{cases}, \quad (3)$$

where α and β are the hyper-parameter of the focal loss, N is the number of key points.

2.4. Centre offset regression branch

The purpose of the centre offset regression branch is to correct the deviation of the centre point of the target between the predicted value from the heatmaps generation branch and the one in the ground truth. As the feature map is down-sampled to one fourth in resolution of the input image, the centre points will be deviated from the ground truth if the feature map is rebuilt to image with the original resolution. To address this issue, an additional local offset is applied on every centre point. In the centre offset regression branch, a two-layer convolution is applied to the feature map in order to obtain the offset feature map. The feature map can be expressed as follows:

$$\hat{O} \in R^{\frac{W}{R} \times \frac{H}{R} \times 2}, \quad (4)$$

and the deviation of the centre point (x_1, y_1) from the ground truth can be predicted from the feature maps.

In the centre offset regression branch, the loss function can be expressed as follows:

$$L_{off} = \frac{1}{N} \sum_p |\hat{O}_{\tilde{p}} - (\frac{p}{R} - \tilde{p})|, \quad (5)$$

where $\hat{O}_{\tilde{p}}$ is the predicted offset, and $\frac{p}{R} - \tilde{p}$ is the offset between the centre points in the down-sampled feature image and the input image.

2.5. Bounding box size regression branch

The purpose of the bounding box size regression branch is to predict the width and height of the targets. In the bounding box size regression branch, a two-layer convolution is applied to obtain the box size of the feature map. The feature map can be expressed as follows:

$$\hat{S} \in R^{\frac{W}{R} \times \frac{H}{R} \times 2}, \quad (6)$$

and the width and the height of the targets are predicted from the two feature maps.

The loss function can be expressed as follows:

$$L_{size} = \frac{1}{N} \sum_{k=1}^N |\hat{S}_{p_k} - s_k|, \quad (7)$$

where \hat{S}_{p_k} is the predicted bounding box size and s_k is the bounding box size in the ground truth.

In the proposed method, the loss function in the training stage can be expressed as follows:

$$L_{size} = L_k + \gamma_{off} L_{off} + \gamma_{size} L_{size}, \quad (8)$$

where γ_{off} and γ_{size} are the hyper-parameter of the loss.

2.6. Targets reconfirmation stage

After the Target Generation Stage, it is expected that incorrect targets, which contain similar features as the targets, might exist in the output mistakenly. Therefore, a second stage called Targets Reconfirmation Stage is introduced to the proposed method in order to eliminate incorrect targets. In this stage, MobileNetV2 (Sandler et al, 2018), which is a rapid objects classification network, is applied to re-analyse all targets from the previous stage and filter out incorrect targets. The network of the Targets Reconfirmation Stage is shown in Figure 3.

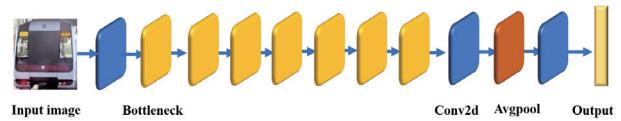


Figure 3. The architecture of the Targets Reconfirmation Stage.

The targets from the previous stage are firstly resized and down-sampled through a convolutional layer to a resolution of $224 \times 224 \times 3$. Then it is processed through 7 bottleneck structures with different strides in each structure to extract the features of the image. It is followed by a convolutional layer with the stride 1×1 to increase the number of feature maps. Furthermore, an average pooling layer is applied to the maps to obtain a 1280-dimensional vector. Finally, a convolutional layer with the stride 1×1 is applied to the vector in order to obtain the final output. The network architecture is illustrated as follows:

Table 1. Network architecture.

Layer	Input	Operator	c	n	s
1	224×3	Conv2d	32	1	2
2	112×32	Bottleneck	16	1	1
3~4	112×16	Bottleneck	24	2	2
5~7	56×24	Bottleneck	32	3	2
8~11	28×32	Bottleneck	64	4	2
12~14	28×64	Bottleneck	96	3	1
15~17	14×96	Bottleneck	160	3	2
18	72×160	Bottleneck	320	1	1
19	72×320	Conv2d 1×1	1280	1	1
20	72×1280	Avgpool 7×7	-	1	-
21	12×3	Conv2d 1×1	3	-	-

where c is the number of output channels, n is the number of times the module repeats, and s is the stride where the module repeats for the first time.

The bottleneck block used in the proposed method is in inverted residuals structure. Unlike the residuals block, the inverted residuals block contains a 1×1 convolutional layer to firstly increase the number of feature maps. Therefore, the features of the image can be wider. Then, a depth-wise convolution is applied to extract the features.

Finally, a convolutional layer is applied to compress the size of the feature map. The structure of a bottleneck is illustrated in Figure 4.

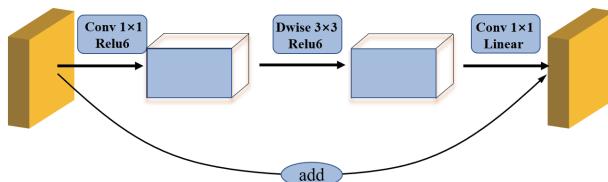


Figure 4. Structure of a bottleneck.

3. Experiment and evaluation

To evaluate the effectiveness of the proposed method, the Train Intelligent Detection System (TIDS) was developed for data acquisition in a real railway environment and validation of the methodology. With the great support and railway research experience from MTR Corporation Limited in Hong Kong, TIDS was installed on five sets of Modernisation Train (M-Train) running in the Tsuen Wan Line (TWL) for proving the concept. TWL was chosen for its singularity of environmental characteristics. With a maximum train operating speed of 80 km/h, TWL contains a lot of steep curves and abrupt gradients with a minimum horizontal curve of 300 m radius and a maximum gradient of 3.2%. Furthermore, the operating environment of the TWL is mostly tunnels without lighting during operating hours. TIDS could be challenged for its performance and reliability under different environmental constraints and operating scenarios in TWL.

TIDS consists of one TIDS host computer for data integration and processing, two high-definition cameras to record videos in front of an operating train, two laser radars for detection of the existence and features of the objects in front of an operating train, and a mini-metre wave radar for detection of the speed of the operating train. The process was conducted on a Titan Xp GPU, Pytorch 1.1.0, CUDA 9.0, and CUDNN 7.1. The evaluation was divided into methodological and application levels.

3.1. Evaluation at methodological level

At the methodological level, TIDS collected a large number of videos with resolution of 1280×720 in a railway mainline and in depot area that were recorded from trains running in the Tsuen Wan Line and Light Rail System. At the beginning, the recorded video data were reviewed in order to shortlist videos or images for identifying trains and buffer stops. The images were then extracted and grouped into a data set. In total, the data set contained 9,500 extracted frames for system evaluation, where 7,500 frames of them were used for training purposes, and the other 2,000 frames were used for testing purposes.

Furthermore, the data set was divided into two sub-groups: the typical scenarios sub-group and the extreme scenarios sub-group. The typical scenarios sub-group contained test data recorded under normal environmental conditions, e.g. clear vision, sunny day, daytime, etc., while the extreme scenarios sub-group contained test data recorded under stringent environmental conditions, e.g. dim lighting, lighting contrast, rainy days, tunnel sections, and night time.

In this experiment, the Average Precision (AP), the Precision and Recall at the confidence of 0.5 were used to evaluate the performance of the proposed method at methodological level. The detection results are shown in Table 2.

Table 2. Detection results.

	Precision	Recall	AP
Typical scenarios	96.51%	94.28%	0.876
Extreme scenarios	75.12%	69.42%	0.526

As shown in Table 2, the AP, Precision and Recall were relatively high in the typical scenarios sub-group. However, the results degraded in the extreme scenarios sub-groups. A review was conducted and it concluded that the features of the targets were sometimes unnoticeable under stringent environmental conditions. Thus, degraded performance resulted.

Furthermore, the effectiveness of the proposed method was also evaluated and compared against the method with only the Targets Recognition Stage, as well as the method with the Single Shot MultiBox Detector (SSD) (Liu et al, 2018). The typical scenarios sub-group were used for the comparison and the results are shown in Table 3.

Table 3. Comparison table with different methods.

	Targets recognition and targets reconfirmation	Targets recognition Only	SSD
Precision – Typical	96.51%	85.70%	81.55%
Precision – Extreme	75.12%	67.20%	62.30%
Recall – Typical	94.28%	95.10%	93.14%
Recall – Extreme	69.42%	70.10%	68.28%
AP – Typical	0.876	0.823	0.784
AP – Extreme	0.526	0.437	0.401

The results in Table 3 show that the proposed method achieved the highest AP in both typical and extreme scenarios sub-groups. With the targets reconfirmation applied, the AP – Typical, and AP – Extreme improved by 0.053 and 0.089 respectively. This improvement was due to the effectiveness of filtering out misidentified targets from the previous stage by the target reconfirmation process. In addition, the AP – Typical, and AP – Extreme of the method proposed in this experiment were 0.092 and 0.125 higher

than the method with SSD respectively. It indicates that the proposed method could achieve better performance in a railway environment. Some detection images which demonstrated the detection of targets are shown in Figures 5 and 6.

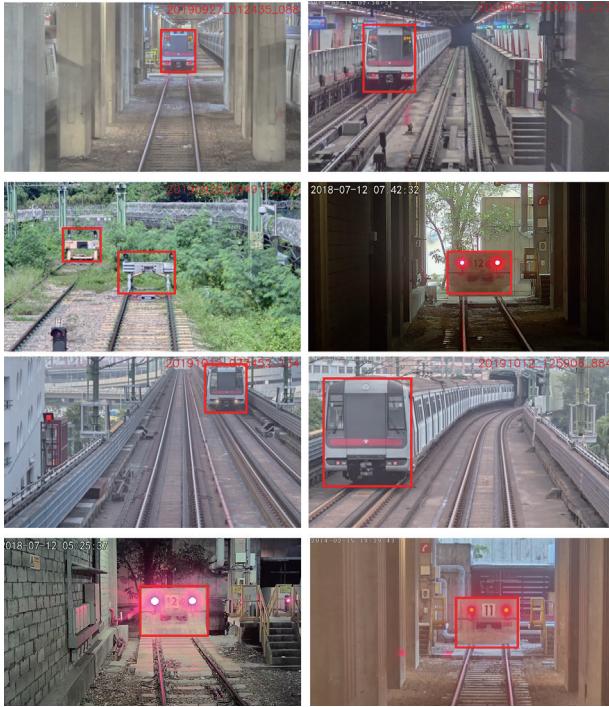


Figure 5. Images of target detection under typical scenarios.

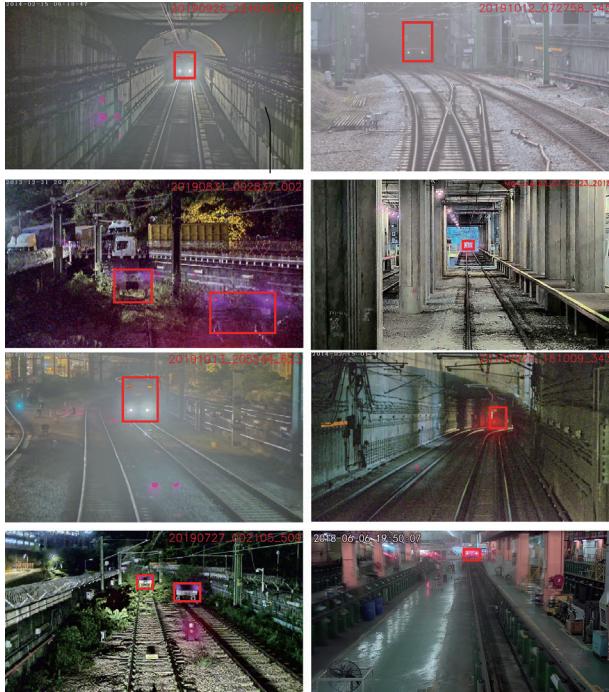


Figure 6. Images of Target Detection under Extreme Scenarios.

3.2. Evaluation at application level

At the application level, the detection function of TIDS was evaluated in real time through continuously data collection and data processing simultaneously as a shadow system on the five sets of M-train under four-month trial in TWL. During the shadow running, data captured by TIDS under different operation modes (e.g. auto mode, coded-manual mode, manual mode) with speed up to 80km/h were continuously collected. The configuration of TIDS installed on an M-Train is shown in Figure 7.

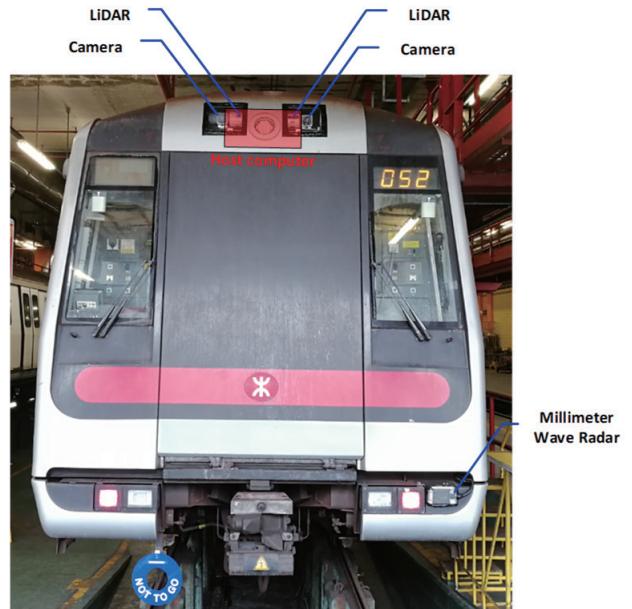


Figure 7. Configuration of TIDS.

From April to July 2019, more than 230 hours of experimental data from TIDS equivalent to 8,000 km operating mileage were reviewed for performance verification.

Several features were implemented in TIDS to strengthen the performance of real-time object detection in TWL. To strengthen the reliability of the system detection, time correlation of consecutive frames was applied to the method at the application level. Four consecutive frames as a minimum were implemented to TIDS to offset the impact of the missed detection in intermediate frames. Besides, in real-time train operation, only the trains or buffer stop in front of the operating train were to be concerned. Therefore, TIDS implemented an additional track envelop identification function to eliminate the nuisance alarms caused by the targets outside the track envelop (Wang et al., 2018). Furthermore, TIDS interfaced with the RS485 communication link from the existing electronic system of the train to acquire useful location-related information, including the travelling route, current and next stations, etc., of the operating train. The information was stored in TIDS to strengthen the track envelop identification function by

allowing the characteristics of the train upcoming track to be known in advance. As there is no interface connection between TIDS and the braking system of M-Train in the proof-of-concept stage, calculation as below was made to identify whether a reminder alert should be provided for the train operator when a train or buffer stop was detected in front of an operating train,

$$s_o \geq \frac{(u_i+u_m)^2}{2a} + (t_s + t_h + t_b) * (u_i + u_m) + s_l, \quad (9)$$

where

- s_o = the measured distance between the operating train and the detected train or buffer stop in front by the TIDS laser radar
- s_l = the maximum tolerance of the measured distance by the TIDS laser radar
- u_i = the measured instantaneous speed of the operating train by the TIDS mini-metre wave radar
- u_t = the maximum tolerance of the measured speed by the TIDS mini-metre wave radar
- a = the minimum brake rate of the emergency braking system of M-Train
- t_s = the maximum delayed time of the TIDS host computer
- t_h = the assumption of human response time
- t_b = the maximum braking system response time required to achieve the minimum brake rate of the emergency braking system of M-Train

When TIDS detected a train or buffer stop in front of the operating train and concluded that its distance was close to the calculated braking distance, TIDS reminder alerts would be prompted to alert the train operators to apply the emergency brake to stop the train.

TIDS data were sampled in the rate of 4 frames per second and a total of 3,445,748 frames were extracted for performance evaluation and the results are shown in Table 4.

Table 4. Performance summary of TIDS.

	Duration (no. of frames)
Positive (P)	47,912
Negative (N)	3,397,836
False positive (FP)	488
False negative (FN)	448

TIDS achieved False Alarm Rate (FAR) and Missed Detection Rate (MDR) of 0.01% and 0.94% respectively. After further review of the missed detection cases, intermediate detection loss between frames, instead of completely missed detection cases, was observed. At the functional level, the results show that all occurrences of

the trains in front could be detected by TIDS. The results demonstrate the system's capability to provide reliable reminder function to train operators. The calculation of FAR and MDR is shown below:

$$FAR = \frac{FP}{P+N}, \quad MDR = \frac{FN}{P}, \quad (10)$$

where P is the number of frames in which trains or buffer stops are present in front of an operating train, N is the number of frames in which there is no train or buffer stop in front of an operating train, FP is the number of frames in which TIDS wrongly reports the detection of a train or a buffer stop, FN is the number of frames in which TIDS fails to detect a train or a buffer stop.

Furthermore, the data consisted of successful target detections under different typical and extreme track scenarios and target detection ranges (e.g. 240 m on straight track, 100 m for maximum gradient of 3.2%, and 70 m for minimum horizontal curve of 300 m radius) which provided confident proof of the capability of TIDS for real application in railway. Some samples are shown in Figure 8 to demonstrate the result of the train detection.

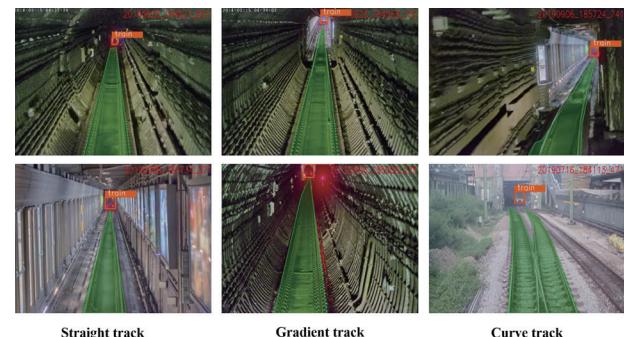


Figure 8. Images of train detection of TIDS.

4. Summary and conclusion

In this project, a CNN-based real-time object detection methodology by applying targets generation and targets reconfirmation was proposed for railway specific application. The two-stage methodology provided capability to overcome different environmental constraints and operating scenarios in railway. The targets generation ensured all potential objects with similar features of the detection targets in the input image to be extracted. Meanwhile, the targets reconfirmation filtered out incorrect objects from the first stage. At the methodological level, the experiment and evaluation results show that the proposed method achieved a 0.876 AP under typical scenarios sub-group and 0.526 AP under extreme scenarios sub-group of the data set collected from real railway line. In addition, at the application level, high performances with FAR of 0.01% and MDR of 0.94% were achieved during the operation in

TWL in Hong Kong. The result provides confident proof that the proposed method satisfactorily detects the desirable targets with a high reliability and its application is adaptable to real railway network.

Sustainable development has been in progress to utilise the capability of the proposed method in railway environment applications. Apart from strengthening the accuracy of the proposed methodology by allowing TIDS to collect data from the operating line daily for intensive training purpose, several new heatmaps channels have also been included in the network to enrich the categories of the detection targets. At the moment, the network can detect objects in significantly smaller size such as humans, signal light and toolboxes with a minimum size of 300 mm x 300 mm x 200 mm under the maximum train operation of 80km/h. According to the field test conducted in Hong Kong, detection could be achieved by the network at distances up to 50 m and 240 m for the toolboxes and train respectively. The enhanced version of the methodology can provide detection and warning of human intrusion on track, improper train operation under red signal as well as leftover equipment on track by workers. The system is also capable for object detection function in other operation modes, e.g. manual driving mode in light rail system, and automatic mode in metro system.

The promising performance of the proposed methodology and developed TIDS in TWL provides a foundation for further development of object detection to different application aspects in railway, e.g. a backup system to the conventional signalling system or novel train-to-train communication-based signalling system to support manual driving under degraded mode, a traffic monitoring system with a behaviour prediction function at railroad junctions in the light rail environment to provide an early alarm of traffic abnormality for light rail operators, an intrusion detection system for fully automatic operation (FAO) railway and the real-time train speed profile optimisation system for traction energy reduction.

Acknowledgments

This work was not funded or supported by any funding agencies or grant-awarding bodies.

Notes on contributors



Mr Rock Ho obtained his M.Sc. degree in Electrical and Electronic Engineering in 2019 from The University of Hong Kong. He has been working in the railway industry in the discipline of rolling stock engineering, including technical design review and commissioning supervision. He is currently Design Support Engineer – Rolling Stock in MTR Corporation Limited.



Mr Zhangyu Wang obtained his M.Sc. degree in Transportation Information Engineering and Control in 2018 from Beihang University. He is working on his Ph.D. degree at the School of Transportation Science and Engineering, Beihang University, Beijing, People's Republic of China. His research focuses on intelligent vehicles and computer vision.



Ir Simon Tang obtained his M.Sc. degree in Electrical and Electronic Engineering in 2002 from The University of Hong Kong. He has been working in the railway industry in the discipline of rolling stock engineering, including asset replacement, design management, system integration and maintenance management. He is currently Lead Design Manager – RS Electrical in MTR Corporation Limited.



Mr Qiang Zhang obtained his M.Sc. degree in Transportation Information Engineering and Control in 2008 from Beijing Jiaotong University. He is working on his Ph.D. degree at the School of electronic information engineering, Beijing Jiaotong University, Beijing, China. His research focuses on automatic train and intelligent control.

References

- [1] Ye T, Wang B, Song P and Li J (2018). Automatic railway traffic object detection system using feature fusion refine neural network under shunting mode. *Sensors*, 18(6), p. 1916.
- [2] Berg A, Öfjäll K, Ahlberg J and Felsberg M (2015). Detecting rails and obstacles using a train-mounted thermal camera. Paulsen R., Pedersen K. (eds) *Image Analysis. SCIA 2015. Lecture Notes in Computer Science*, vol 9127. Springer, Cham
- [3] Karaduman M (2017). *Image processing-based obstacle detection with laser measurement in railways*. 10th International Conference on Electrical and Electronics Engineering (ELECO), IEEE, pp. 899-903.
- [4] Li Y, Dong, H, Li, H, Zhang X, Zhang B and Xiao Z (2020). Multi-block SSD based on small object detection for UAV railway scene surveillance. *Chinese Journal of Aeronautics*, 33(6), pp. 1747–1755.
- [5] Amaral V, Marques F, Lourenço A, Barata J and Santana P (2016). *Laser-based obstacle detection at railway level crossings*. *Journal of Sensors*, 2016, pp. 1–11.

- [6] Lüy M, Çam E, Ulamış F, Uzun İ and Akin S (2018). Initial results of testing a multilayer laser scanner in a collision avoidance system for light rail vehicles. *Applied Sciences*, 8(4), pp. 475.
- [7] Zhou X, Wang D and Krähenbühl P (2019). *Objects as Points*. [online]. arXiv.org. Available at: <<https://arxiv.org/abs/1904.07850>>.
- [8] Redmon J and Farhadi A (2018). YOLOv3: An incremental improvement. [online]. arXiv.org. Available at: <<https://arxiv.org/abs/1804.02767>>.
- [9] Ren S, He K, Girshick R and Sun J (2015). *Faster r-cnn: towards real-time object detection with region proposal networks*. Advances in neural information processing systems, pp. 91-99.
- [10] Yu F, Wang D, Shelhamer E and Darrell T (2018). Deep layer aggregation. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2403-2412.
- [11] Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L C (2018). *Mobilenetv2: inverted residuals and linear bottlenecks*. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520.
- [12] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C (2016). 'SSD: single shot multibox detector'. European conference on computer vision, Springer, Cham, pp. 21-37.
- [13] Wang Z, Wu X, Yu G and Li M (2018). 'Efficient rail area detection using convolutional neural network', IEEE Access, no.6, pp. 77656-77664.