Object detection at level crossings using deep learning techniques

AUTHOR: Muhammad A B Fayyaz

Muhammad has recently submitted his PhD thesis, which considers how to integrate Artificial Intelligence (AI) into existing sensing systems at level crossings, thereby improving their Safety Integrity Levels (SILs). Muhammad is a PWI Ambassador. He has a mechanical engineering background and expertise in different fields of AI. He is enthusiastic about learning and sharing his expertise using different platforms such as YouTube and Kaggle. He is currently exploring Computer Vision applications within the rail industry.

THE PWI YOUNG ACHIEVER AWARD

This article is an updated version of a paper that was awarded "Highly Commended" status by the judges of the PWI 2020 Young Achiever Award (YAA) competition.

This award is open to all involved in rail infrastructure engineering projects; engineers, project & maintenance staff, graduates, apprentices etc. See page 59 for details of the 2021 award.

AUTHOR: Dr. Christopher Johnson

Dr Christopher Johnson is a Systems Engineering Team Leader at Park Signalling. Previously he operated as a programme manager in Industry 4.0 at Manchester Metropolitan University and has a distinguished track record in delivering impact through industrial partnerships as a recognised project manager (APM-PMQ, SCQF Level 7). Working closely with the rail supply industry, Dr Johnson's background work has included synergistic, adaptive prognostic monitoring of rail assets, bespoke image analysis techniques in the classification of surface born rail defects and pioneering the use of adaptive learning techniques in the analysis of rail vehicle and track components.

ABSTRACT

Level crossings pose a significant risk, and their use or misuse can lead to serious accidents between railway vehicles and road and footpath users who may sometimes behave unpredictably. Currently, Great Britain has 7,500 level crossings and the Rail Safety and Standards Board (RSSB) reported 3 fatalities and 385 near misses in the year 2015-16. The Office of Rail and Road (ORR) and RSSB have recommended the upgrading of the automated technologies used at level crossings. For this purpose, automated safety systems remain a key area for investigation for inclusion in interlocking systems at level crossings. In this study, we propose a solution to automate the operational cycle at level crossings using deep learning technologies. The proposed solution will add another layer of resilience to the safety system without the use of manual operators at level crossings and ensure an interlocking system with a safety integrity level (SIL) of 3-4.

The paper discusses the current technologies utilised at level crossings along with the algorithms required for post-processing the data. The proposed model integrates deep learning technology with the current vision system, CCTV, to detect and localise an object at a level crossing. Two approaches are discussed to train the neural network: from scratch or using transfer learning techniques. Neural networks along with their associated accuracy, which represent the percentage of true predictions on the test dataset is mentioned as well. Classification, detection and segmentation models are used to classify, detect and localise objects like vehicles, bicycles and pedestrians at level crossings. Neural networks trained from scratch achieved an accuracy of 75%, networks trained for object detection using transfer learning techniques achieved an accuracy from 58 to 82%, and the image segmentation model, Mask RCNN, achieved an accuracy of 95% on test data. Finally, this paper discusses some future work to improve the network's accuracy and another application of convolutional neural networks (CNN) using radar sensing.

INTRODUCTION

Railway level crossings represent some critical safety issues for the railway industry and are therefore of great interest to authorities charged with improving safety eg RSSB, Network Rail and Office of Rail and Road (ORR) (Office for Rail and Road, 2017)[1]. Level crossings account for nearly half of the potentially higher risk incidents on British railways, with misuse from road users accounting for nearly 90% of the risk at level crossings over the last five years. For Great Britain, there were three fatalities and 385 near misses at level crossings in 2015–2016. Furthermore, in its annual safety report, the Rail Safety and Standards Board (RSSB) (TSSA 2014)[2] highlighted the risk of incidents at level crossings during 2016/17 with a further six fatalities at level crossings including four pedestrians and two road vehicles (Office for Rail and Road, 2016)[3].

Traditional intrusive sensors are installed either inside or on rail lines (Darlington, 2017)[4]. Intrusive sensors are costly and disrupt the rail system during their installation and maintenance, which further increases the cost and shortens the product lifecycle (Petrov, 2011; Darlington, 2017)[5,4]. Intrusive sensors were replaced with non-intrusive sensors (Ohta, 2005; Valera and A. Velastin, 2005; Kim et al., 2012; Govoni et al., 2015; Horne et al., 2016; Leddar, 2018) [6,7,8,9,10,11] which are installed outside rail lines and do not affect the rail system during installation and maintenance. Therefore, the product is comparatively cheap with a longer product-lifecycle. Some post-processing techniques are required to process the information acquired from any of the above-mentioned sensors to analyse and differentiate between obstacles present at the level crossings. Most of these traditional algorithms detect the foreground by comparing pixel values with background pixels. However, the environment at level crossings is complex and dynamic with growing vegetation and the presence of many harmless objects.

Traditional algorithms cannot detect complex texture, adapt to a dynamic background or avoid detection of unnecessary harmless objects. To avoid these problems, the proposed work utilises "deep learning" technology (LeCun, Bengio and Hinton, 2015)[12] integrated with the proposed vision system; CCTV. Deep learning technology can learn representations from labelled pixels; hence it does not depend on background pixels. Deep learning technology can classify, detect and localise objects in the level crossing area. It can classify and differentiate between a child and a small inanimate object, which was impossible with traditional algorithms. The system can detect an object regardless of its position, orientation and scale without any additional training because it learns representation from the data and not rely on background pixels. In the proposed system, deep learning technology is integrated with the existing vision system installed at level crossings, hence implementation cost is significantly reduced as well.

This paper provides a summary of existing sensing systems and algorithms used at level crossings before discussing deep learning technology and its integration at level crossings with the video sensing system. The paper uses different models, which are trained using deep learning technology; from scratch and transfer learning techniques along with their accuracy and results at real-site deployment. Finally, the paper gives some concluding remarks and suggestions for future work.

METHODOLOGY

To select the most appropriate sensing system, a summary of sensors is presented in Table 1, which are currently used or have the potential for their applicability at level crossings. The work in (Fayyaz et al., 2020)[13] discusses the limitations and preferability of each sensor. Table 1 strongly advocates the preferability of CCTV for its application at level crossings because of its low cost and maintenance.

The traditional algorithms rely on calculating the background pixel values to model the background, which then confirms the presence of the foreground using a predefined threshold value. This method does not depend on the object features or recognise

	Equipment cost	Maintenance	Product life cycle	False positive	Impact on weather
Inductive Loops	5	5	4	4	4
Strain Gauge/Piezometers	5	5	4	4	3
CCTV	2	3	2	3	3
Stereo Camera	3	4	2	3	3
Thermal Camera	4	2	2	2	2
SONAR	2	4	3	3	2
Millimeter-Wavelength Beam Interference	4	4	1	3	4
Laser Range Finder Beam (LRBF)	3	4	2	5	4
Light Detection and Ranging (LiDAR)	5	4	2	2	2
Radar	3	2	1	1	2

Table 1: A brief comparison of different sensors available at level crossings, where five represents the worst choice and one represents the best choice for the given parameter

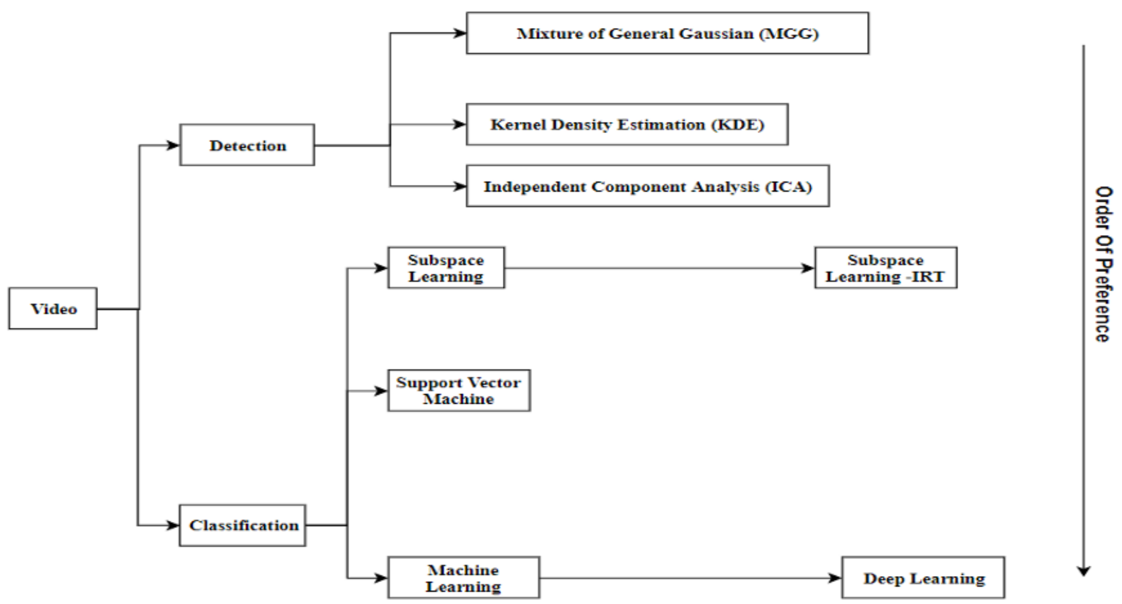

Figure 1: A decision tree to demonstrate which sensor is prefered for its applicability at level crossings

objects from their specific properties, rather on the variation within the background scene. This further suggests that any variation in background scene due to change in lighting conditions or dynamic environment eg, growing or moving vegetation would give false positives. These false positives would pose a significant threat to road or rail users. Also, the recovery time for most of these mentioned algorithms is slow, which means more chances of failure, which is undesirable for use at level crossings. To overcome these problems, the work in (LeCun, Bengio and Hinton, 2015)[12] proposed the use of the deep learning technique. Deep learning is a subset of machine learning, which can automatically learn the representation from the images and classify the given categories without any supervision. It consists of multiple processing layers and can learn representation using backpropagation algorithms. With enough representations, very complex functions are learned and used for different applications such as classification and detection. The deep learning model learns features for the general-purpose avoiding the need for human engineering, for example, the classification model learns small features like edges, orientation and location to more complex representations such as motifs or parts of objects with assembled features. From such learned models, it can detect and classify objects very effectively.

The model does not require any supervision or regular increment of data to update the existing system. Another advantage of deep learning is its ability to integrate with other existing sensing systems eg a video system. A decision tree in Figure 1 is used to demonstrate the preferability of these sensors for its applicability at level crossings.

This paper aims to use deep learning technology, which is integrated with the vision sensing system at a level crossing to classify and detect an obstacle. The objects that may pose a threat to a train or other users at a level crossing are vehicles, pedestrians (adult and child) and bicycles, hence the model is trained to detect and classify these particular objects during the operation cycle at a level crossing. Convolutional neural networks (CNN), which are used for image classification (Krizhevsky and Hinton, 2012)[14], detection (Karpathy and Leung, 2014)[15] and segmentation (Huang, Pedoeem and Chen, no date; Redmon et al., 2016)[16,17] are trained using two different methods; from scratch and using transfer learning techniques. Images are obtained for each required category from an open source eg ImageNet and split into training and validation. The training set of images are required to train the model, whereas the validation dataset is used for evaluation of the model before its deployment at the real site.

A traditional neural network is used as a base model for image classification when training the network from scratch, where different depths of convolutional neural network and regularisation techniques are used to improve the accuracy of the model. To better understand the effect of the convolution layer, often the visualisation of such representations during the training process is used. Afterwards, the transfer learning technique is used to train different pre-trained neural networks along with their associated weights to retrain the model on new datasets and its application at level crossings. The pretrained neural network used are RetinaNet (Lin, Ai and Doll, 2018)[18], YOLO (Redmon et al., 2016)[17], tiny-YOLO (Huang, Pedoeem and Chen, no date)[16] and Mask-RCNN (Ren, He and Girshick, 2015)[19].

Trail #	Change	Network's Summary	Accuracy
1	Traditional Neural Network (base)	1 Flatten Layer 2 Dense Layer	55%
2	Add Convolution Layers	2 Convolution Layer 2 Max. Pooling Layer 2 Dense Layer	68%
3	Add Regularisation	4 Convolution Layer 4 Max Pooling Layer 2 Dense Layer 1 Dropout Layer	75%

Table 2: Details of CNN's architecture used to train the model for classification using a traditional neural network and a network with convolution layers

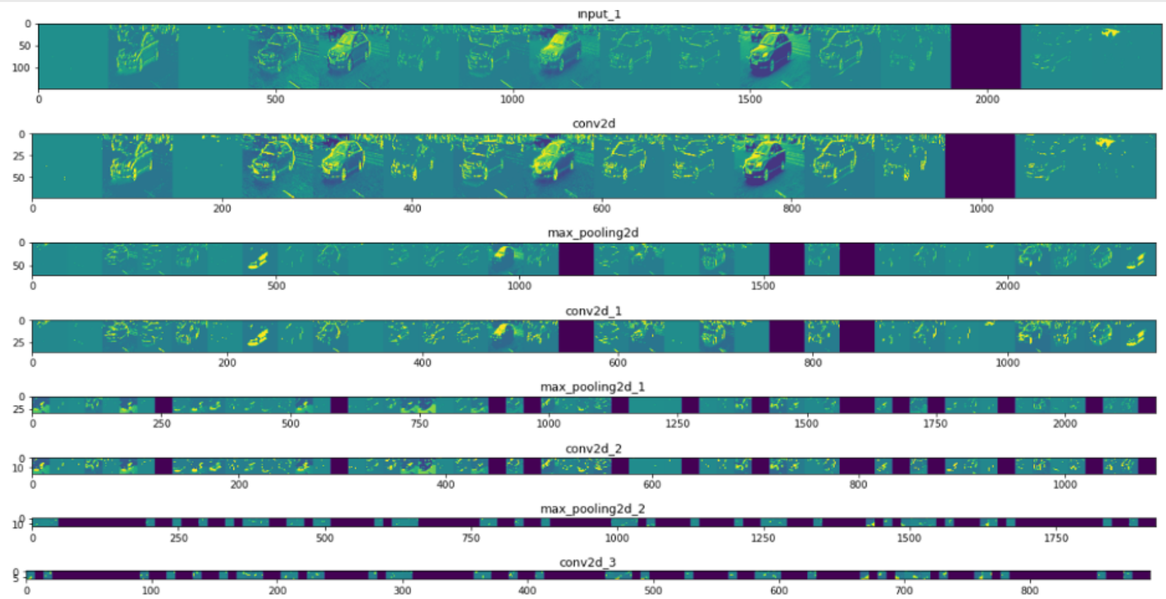

Figure 2: Visual representation of a car at level crossing site. This visual above represents the features learned from successive convolution layers within a neural network

RESULTS

CONVOLUTIONAL NEURAL NETWORK TRAINED FROM SCRATCH

A public dataset from ImageNet is used to train and evaluate the CNN model. To understand the effect of the convolution layer (Dumoulin and Visin, 2016; Karpathy, 2018) [20,15] a traditional neural network with only Flatten and dense layer is trained on the given dataset, which achieved an accuracy of about 55%. With the addition of two convolution layers and two max pooling layers (Ricco, 2017)[21] the accuracy increased from 55% to 68%. Where different regularisation techniques are also used, an increase in the model's accuracy to 75% is obtained. The accuracy metrics represent the percentage of true predictions on the test-dataset. The details of these models along with their architecture and accuracy metrics are summarised in Table 2.

To understand the convolution layer, a common practice is to visualise the representations learned during the training process. An image of a vehicle (car) present in the dynamic environment of a level crossing is fed to the network and representation from each subsequent layer is visualised as shown in Figure 2.

In the first layer, as shown in Figure 2, the activation from the neural network has retained all the information from the given image (eg car). It contains some edges or collection of apparent edges that is interpretable by human eyes as well. The activations or representations become more complex in deeper layers of the neural network. They are more abstract and less interpretable. These complex features correspond to the specific class rather than a visual interpretation of the generic object. The black boxes in the subsequent layers of the neural network demonstrate the sparsity of the network. These black boxes represent that the pattern encoded by the particular filter is not found in this particular location. As the convolution neural network is fed with raw data containing specific objects, the subsequent layers within the CNN learn relevant information specific to the class and leaves irrelevant information eg the visual appearance of the image. For this reason, any object similar to the given object in the image regardless of its orientation and size will be predicted accurately. The whole idea of retaining generic information in the subsequent layers gave rise to “transfer learning” techniques, where a notable model trained on millions of images is used. The generic pattern learned on these models will be used along with specific features learned from a small dataset fed to the CNN.

CONVOLUTIONAL NEURAL NETWORK TRAINED USING TRANSFER LEARNING TECHNIQUES

Neural networks, which are trained on millions of images, are used to retrain the network on custom classes or images for a specific problem. Transfer learning is achieved in three possible scenarios; freeze all the layers except the last final layer, freeze first few layers and retrain the remaining one and retrain the whole network with new weight initialisation.

A common practice is to use notable models, which are trained on millions of images and achieve state-of-the-art results and to feed them with a custom dataset for specific classes. Table 3 shows two notable models along with the summary of its parameters and achieved accuracy. The accuracy achieved on a new custom dataset using transfer learning is mentioned as well.

Many different data augmentations and regularisation techniques could be used to further improve the accuracy of the networks. However, accuracy beyond 75% is acceptable to validate and test the new data for its application at a level crossing. The models can predict at a rate of 15 frames per second (fps) with low computational power, which suggests the efficiency and reliability of these models for real-time application. The natural progression from the classification is the localisation and detection of these objects at level crossings. A summary of the models used for transfer learning is mentioned in Table 4.

Once trained, the model is deployed at a real site for real-time predictions. The results as shown in Figure 3 demonstrate the effectiveness and reliability of these models for its applicability at level crossings.

The accuracy and confidence of the YOLO network is higher than that of the RetinaNet, which is roughly 82% ± 10% but it did not classify every single object, whereas the RetinaNet can predict all objects within two distinct environments with an accuracy of 70% ± 8%. Other networks eg, Mask-RCNN can detect and localise the objects to the pixel-level precision compared with the bounding boxes. Pixel-wise detection is called image segmentation. The accuracy of Mask RCNN eg, 95% ± 7% is highest compared to all other networks used for its application at a level crossing as shown in Figure 3.

CONCLUSION

This paper addresses the safety issue at level crossings and requires an automated system for classification and detection to automate the operational cycle. A brief introduction to sensors, which are used or have the potential to be used at a level crossing is mentioned along with some associated algorithms used for post-processing the data acquired for real-time detection.

The paper demonstrates the effect of convolution layers on traditional neural network and the representations learned during the training of the network. The accuracy achieved from a CNN trained from scratch is further increased using transfer learning techniques. Particularly, the YOLO and Mask-RCNN can precisely classify and localise objects in a dynamic background environment.

The network can correctly classify and predict objects in real-time, which demonstrates its effectiveness and accessibility to replace the manual operator at level crossings increasing the resilience of the safety system. Furthermore, the model could be used for objective risk analysis. The classification from the network is useful data to analyse which particular objects frequently misuse the level crossings and cause the greatest threat. Such information will allow authorities to perform risk assessments more effectively. For example, the real-world classification for level crossings demonstrates that cyclists more frequently misuse level crossings, posing a threat to themselves and the train driver. From the data, the authorities can take precautionary measures and actions to avoid such incidents and improve the safety of bicycle users.

The proposed network should be trained on a larger dataset gathered from the same distribution (sites which strongly represent the site of deployment). This will ensure that the network is not overfitting (a state where a model precisely learns on training examples but does not perform well on the new dataset). The dataset will need to be from Great Britain only because other countries deploy different technologies at level crossings or do not use them, particularly in underdeveloped countries. The dataset from different locations from within Great Britain will allow the user to gather a large variance of images to avoid our network's bias towards one particular ethnicity.

As an alternative to using CCTV sensors (as described in this paper), CNN could also be trained on images obtained from radar. Radar imaging is a diverse application, which provides sufficient information for CNN layers to learn representation and classify objects. One such application is to use micro-Doppler signatures from radar to train a new CNN model.

REFERENCES

1. Office of Rail and Road (2017) Annual safety performance report. A reference guide to safety trends on GB railways 2016/17. doi: 10.1210/jc.2013-2845.
2. TSSA (2014) RSSB Annual Safety Performance Report 2016-17.
3. Office of Rail and Road (2016) ORR's annual health and safety report of performance on Britain's railways: 2015-16. Available at: https://orr.gov.uk/__data/assets/pdf_file/0020/22457/annual-health-and-safety-report-july-2016.pdf.
4. Darlington, P. (2017) Obstacle detection for level crossings, Rail News. Available at: https://www.railengineer.co.uk/2017/05/30/

Name	Summary	Accuracy
InceptionV3	The network achieves 21.2% top-1 and 5.6% top-5 error for single frame evaluation with a computational cost of 5 billion multiply-adds per inference and with using less than 25 million parameters (Inception, 2019).	76%
MobileNet	Same pre-processing as Inception with an accuracy 66.51% top-1 and 87.09% top-5 on ImageNet-2012 Validation dataset. The network has about 3.3 million parameters (MobileNet, 2019).	82%

Table 3: Some notable models along with their summary and achieved accuracy on a new custom dataset using transfer learning

Name	Summary	Accuracy % (based on test data)
RetinaNet	• ResNet + FPN as backbone for feature extraction. • Two task-specific subnetworks for classification and bounding box regression (Lin, Ai and Doll, 2018).	70
Yolo	• A Neural Network is applied to the whole image. • The Image is divided into grids for predicting bounding boxes and probabilities. • These predictions and probabilities are weighted to give final output (Redmon et al., 2016).	82
Tiny-Yolo	• A smaller version of YOLO network • Design for computers with no GPUs (Huang, Pedoeem and Chen, no date).	58
Mask RCNN	• Trained on CocoDataset • Extends Faster R-CNN by adding a branch for predicting an object mask in parallel with an existing branch of bounding box prediction (Ren, He and Girshick, 2015).	95

Table 4: Models used for detection and classification using transfer learning techniques. The model's summary and results are mentioned as well

obstacle-detection-for-level-crossings/ (Accessed: 26 February 2020).
5. Petrov, S. (2011) 'Loop detectors in active level crossing applications', AusRAIL PLUS, pp. 22–24. Available at: http://railknowledgebank.com/Presto/content/GetDoc.axd?ctID=MTk4MTRjNDUtNWQ0My00OTBmLTllYWUtZWFjM2U2OTE0ZDY3&rID=MjE3OA==&pID=Nzkx&attchmnt=True&uSesDM=False&rIdx=MTcwOA==&rCFU=.
6. Ohta, M. (2005) Level crossings obstacle detection system using stereo cameras, Quarterly report of RTRI. doi: 10.2219/rtriqr.46.110.
7. Valera, M. and A. Velastin, S. (2005) 'Intelligent distributed surveillance systems: a review', in IEEE Proceedings-Vision, Image and Signal Processing, pp. 192–204.
8. Kim, G. et al. (2012) 'Design of safety equipment for railroad level crossings using laser range finder', in 9th International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, pp. 2909–2913. doi: 10.1109/FSKD.2012.6234334.
9. Govoni, M. et al. (2015) 'Ultra-wide bandwidth systems for the surveillance of railway crossing areas', IEEE Communications Magazine, pp. 117–123. Available at: https://ieeexplore.ieee.org/abstract/document/7295472.
10. Horne, D. et al. (2016) 'Evaluation of radar vehicle detection at four quadrant gate rail crossings', Journal of Rail Transport Planning & Management, 6(2), pp. 149–162. Available at: https://pdf.sciencedirectassets.com/280481/1-s2.0-S2210970616X00042/1-s2.0-S2210970616300063/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEN3%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJGMEQCIF3bGKYpsAroj8%2B8wg7FK0OyRQNvG7ijPnyjGICAHe4hAiBFItZ4RTqv5b.
11. Leddar, T. (2018) Solid-state LiDARS : Enabling the automotive industry towards autonomous driving, Leddar Tech. Available at: https://www.tu-auto.com/intelligence/solid-state-lidars-enabling-the-automotive-industry-towards-autonomous-driving/.
12. LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', nature, 521(7553), pp. 436–444. Available at: https://www.nature.com/articles/nature14539.
13. Fayyaz, M. A. B. et al. (2020) 'A review of the technological developments for interlocking at level crossing', Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit. doi: 10.1177/0954409720941726.
14. Krizhevsky, A. and Hinton, G. E. (2012) 'ImageNet classification with deep convolutional neural networks', Advances in neural information processing systems, pp. 1097–1105.
15. Karpathy, A. and Leung, T. (2014) 'Large-scale video classification with convolutional neural networks', Proceedings of the IEEE conference on Computer Vision and Pattern Recognition.
16. Huang, R., Pedoeem, J. and Chen, C. (no date) 'YOLO-LITE : A real-time object detection algorithm optimized for non-GPU computers'.
17. Redmon, J. et al. (2016) 'You only look once: Unified, real-time object detection', in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.
18. Lin, T., Ai, F. and Doll, P. (2018) 'Focal Loss for dense object detection', computer vision and pattern recognition.
19. Ren, S., He, K. and Girshick, R. (2015) 'Faster R-CNN : Towards real-time object detection with region proposal networks', Advances in neural information processing systems, pp. 91–99.
20. Dumoulin, V. and Visin, F. (2016) 'A guide to convolution arithmetic for deep learning'. Available at: http://arxiv.org/abs/1603.07285.
21. Ricco, J. (2017) What is max pooling in convolutional neural networks?, Quora. Available at: https://www.quora.com/What-is-max-pooling-in-convolutional-neural-networks (Accessed: 29 April 2020).
22. Inception (2019) Inception in Tensorflow.
23. MobileNet (2019) MobileNet.

Figure 3: Predictions from: A) Yolo | B) RetinaNet | C) Mask-RCNN at two distinct environments

TECHNICAL ARTICLE

AS PUBLISHED IN

The PWI Journal

April 2021

VOLUME 139 PART 2

#thepwi

W thepwi.org

PermanentWayInstitution

@PermWayInstit

Permanent Way Institution

@the_pwi

If you would like to reproduce this article, please contact:

Kerrie Illsley
CREATIVE MANAGER
Permanent Way Institution
kerrie.illsley@thepwi.org

PERMANENT WAY INSTITUTION
THE PROFESSIONAL COMMUNITY FOR RAIL INFRASTRUCTURE ENGINEERING