

# An innovative application of pantograph recognition system based on deep learning

**Handong Li<sup>1,4</sup>, Roberto Palacin<sup>2,5</sup> and Satnam Dlay<sup>2,6</sup>**

<sup>1</sup>Mathematics, Physics and Electrical Engineering, Northumbria University Newcastle, Newcastle upon Tyne, UK, NE1 8ST

<sup>2</sup>Electrical and Electronic Engineering, School of Engineering, Newcastle University, Newcastle upon Tyne, UK, NE1 7RU

<sup>4</sup>handong.li@northumbria.ac.uk

<sup>5</sup>roberto.palacin@ncl.ac.uk

<sup>6</sup>satnam.dlay@ncl.ac.uk

**Abstract.** One of the most significant aspects for the correct operation of a modern railway electrification system is the health of the pantograph and overhead line. The application of machine vision technology to monitor the status of pantographs in real time can reduce pantographs and catenary accidents caused by unpredictable events. It is challenging to achieve real-time and accurate criteria with present pantograph detecting technologies. Therefore, this methodology collects pantograph images through high-definition cameras and transmits them to the cloud through 5G, use the Mask R-CNN algorithm to process and analyse the images., This technology can assist railway technicians in judging the status of the pantograph. Mask R-CNN employs the Resnet network for feature extraction. Resnet has the characteristics of cross-layer connection, which avoids the problem of network degradation due to the deep learning network being too deep, and greatly improves the training efficiency. The recognition matching degree of pantographs are greater than 0.975, enabling pantograph recognition in a variety of environmental conditions. The use of 5G connection increases transmission speed and allows for real-time detection of pantograph status, which is critical for the railway's automated operation.

**Keywords:** 5G, Mask-RCNN, Pantograph detection, Machine learning, Transport, Railway

## 1. Introduction

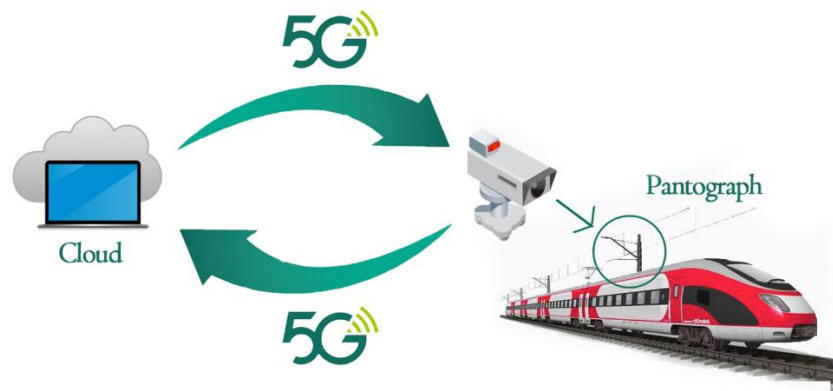
The pantograph is a key device for urban rail transport to obtain electrical energy from the power grid, and its quality determines the power transmission and operation safety of the train [1]. According to the data released by the researchers, the failure caused by the power supply system during train operation are one third of the railway failures [2]. Therefore, the real-time and accurate monitoring of the pantograph can reduce the occurrence of railway accidents to a certain extent [3], and the recognition accuracy of the acquired image directly determines the reliability of the safety warning.

Recent efforts have been made on optimizing and improving the architectural performance of instance segmentation algorithms. Fully Convolutional Network (FCN) [4][5] achieves simple, efficient, and end-to-end semantic segmentation, but FCN cannot solve the task of instance segmentation due to the

translation invariance of convolution and the inability to distinguish individual object instances. The subsequently proposed Instance FCN [6] and fully convolutional instance-aware semantic segmentation (FCIS) [7] solved this problem. Instance FCN is an architecture for improving local pixels of an image, and an instance-sensitive score map is proposed, which solves the problem of the same pixel having varied responses in different regions. However, it is not an end-to-end network architecture because it requires a scoring network for auxiliary discrimination after generating the full-instance graph. FCIS is the first fully convolutional end-to-end instance segmentation model based on Instance FCN. While using the instance-sensitive score map, it adds an internal/external score map that distinguishes the positional relationship of the same pixel in the target instance. Regions of Interest (RoI) [8][9] generated by Region Prediction Network (RPN) [10] work together and aggregate to accomplish the task of simultaneous segmentation and classification.

Based on the above research progress, considering that the train operating environment is complicated, and the segmentation accuracy requirement is high, a simple neural network cannot meet the work requirements. Therefore, this letter firstly uses HD cameras to obtain the pantograph image during the normal operation of the train between stations. Then the Mask R-CNN algorithm is used to segment and detect the pantograph images, and the results show that the algorithm can achieve a positive effect in different environments. Finally, this letter employs 5G signal transmission experiments to test the rate and feasibility when transmit the video data package. The system framework is shown in Fig. 1:

The status of the pantograph is captured using a high-definition camera, and the image is sent to the cloud via 5G communication. The images are processed on a cloud computer and then sent back to the train receiver via 5G communication. The purpose of real-time monitoring is realized under the processing of transmission and high-speed cloud platform.



**Figure 1.** System framework.

## 2. Proposed method

### 2.1. Communication technology

The architecture of a 5G network is built on a service system, with an access network, a bearer network, and a core network. 5G networks have fast speed, high capacity, high dependability, low latency, and low power consumption when compared to previous networks.[11][12] The peak value of its base station is greater than 20 Gb/s, which improves the problem of the limited number of 4G communication connection terminals, and the uplink and downlink delays between the base station and the terminal on the user plane are both 0.5ms. In addition, 5G [13] communication solves the problem of high-power consumption by reducing signaling overhead, so that the device terminal can remain online for a long time, which provides a guarantee for high-speed transmission of data streams such as video and pictures in this letter.

## 2.2. Deep learning background

The goal of picture instance segmentation is to determine the object's category and precisely segment the object's shape and size. Convolutional neural networks use convolutional layers instead of fully linked layers, allowing the network to receive images of any size and retrieve the class to which each pixel belongs using abstract features.

We start by labelling the areas of the dataset that need to be identified and classed, then train the computer with a convolutional neural network to recognise the categorization. Our approach is based on the Mask-RCNN technique, which uses the region proposal network to generate a series of candidate regions before performing classification, localization, and segmentation. This is a two-stage technique for natural picture segmentation; while the convergence time is slower than the single-stage algorithm, the region proposal network has already served as a screening function, alleviating the problem of class imbalance to some extent. The purpose of this work is to use the method to find pantographs in complex natural situations, reliably classify them, and indicate their bounding boxes and masks.

Mask-RCNN is a powerful framework that can handle a wide range of image processing applications, including object detection and semantic segmentation. The backbone network used by Mask-RCNN is deep residual network, which is implemented in this research as the ResNet101+FPN model. It can extract feature maps at various stages from the input picture and uses the Feature Pyramid Network (FPN) for many tasks. The extraction of scale data, as well as its top-down and horizontal link structure, combines features from many scales, resulting in powerful semantic and spatial information. For each pixel on the feature map, a fixed number of anchor boxes is established, and several candidate regions (ROIs) of various sizes are created by computing the intersection between each anchor box and the real box depicted on the picture. Finally, the candidate ROIs are binary classified and bounding regression is performed using the region proposal network (RPN). The calculation of redundant information in the second stage is successfully reduced by filtering out ROIs with low categorization scores.

The ROIs in the original image are mapped to the pixels in the feature map one by one in the second step, and ROIs of various sizes are converted to uniform sizes. Secondly, the RoI Align technology is used to choose the region of interest, and the grayscale size on the pixel point is obtained using the bilinear interpolation approach to extract the critical feature information contained in the ROI, and to complete the classification, regression, and segmentation. The goal is to serialise the feature aggregation process as a whole. Finally, each pixel on each ROI is identified and forecasted by regression, and the final binary mask is generated by adding a segmentation branch to the fully connected layer.

The training objective of Mask-RCNN is to reduce the loss of multitask as shown in equation (1). The cost function is the sum of the errors from the class ( $L_{cls}$ ), the bounding box ( $L_{bbox}$ ) and the mask ( $L_{mask}$ ).

$$\begin{aligned}
 L_{total} &= L_{cls} + L_{bbox} + L_{mask} \\
 L_{cls} &= \text{softmax Cross Entropy} \\
 L_{bbox} &= \text{Regression} \\
 L_{mask} &= \text{Binary Cross Entropy}
 \end{aligned} \tag{1}$$

## 3. Experiments

The experimental environment used in this letter is shown in Table 1, and the parameters in the model training process are shown in Table 2.

In this letter, a high-precision visual camera is used to collect pictures of relevant distribution lines at the uninterrupted operation site to form a data set, with a total of 1000 images in the data set. The dataset is first pre-processed, and the image size is set to 1920 pixels  $\times$  1080 pixels. Then use the labelling tool 'LabelMe' to manually label the data.

**Table 1.** Experimental setup.

Project	Model or Parameter Value
CPU	Intel i9-12900k
Memory	GB&128
GPU	GTX3080Ti
Operating System	Ubuntu16.04
Software Environment	Anaconda3, Cuda9.0, Python3.6
Development tools	PyCharm
Network Framework	TensorFlow

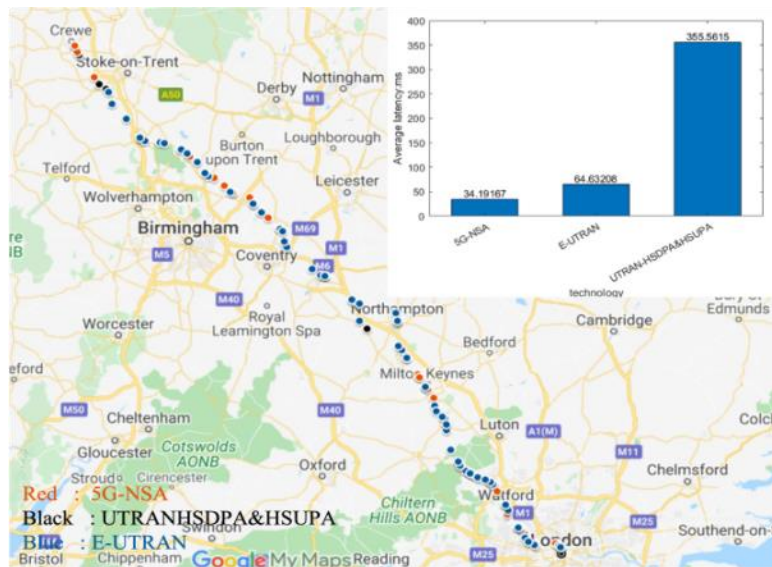
**Table 2.** Training parameters.

Project	Model or Parameter Value
Batch size	10
Weight decay	0.0001
Learning rate	0.001
Number of training iterations	400

## 4. Results

### 4.1. Communication method comparison

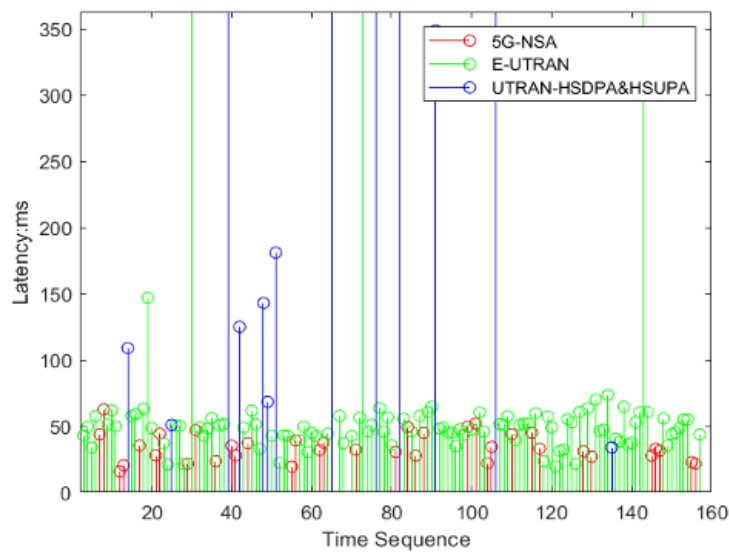
In this letter, a large number of pantograph images (1000) are collected and transmitted to the cloud computer in real time through 3G, 4G, 5G signals. The images are segmented on the cloud computer using the enhanced Mask-RCNN model, and the current pantograph is evaluated. In the British railway environment, the situation illustrates the benefits and drawbacks of each signal transmission.



**Figure 2.** Map of the locations and average time latency of 3 technologies.

The transmission test results of high-speed trains under various technologies are firstly compared in this paper. The train travels from London to Crewe. Every 10 minutes, a video is uploaded and pinged to google.com. The software creates '.srt' files and records the data including details such as latitude, longitude, and latency automatically. From the collection of '.srt' files, the following results were

obtained. After each 'ping,' as shown in Fig. 2 the location information is saved, and Google Map is used to display the latitude and longitude coordinates obtained using various technologies on a map. The E-UTRAN technology receives the most points, as shown in Fig. 2 This demonstrates that 4G technology is still widely used in this road section. 3G has the least number of location points, followed by 5G. It can be speculated that 5G equipment is still under construction in this road section. The average latency of the three technologies is also depicted in Fig.7, and the number of delays at each time interval in time order are shown in Fig.3. Each point corresponds to a location on the map in Fig.2. It is obvious that 5G has the shortest and most stable latency. Although 4G has a lower latency than 3G, there are still significant fluctuations. 3G has the highest average latency and fluctuations, so we use 5G to provide ultrahigh-speed links for HD video streaming and low data-rate speeds for sensor networks on the train in this project.



**Figure 3.** Latency.

#### 4.2. Recognition result

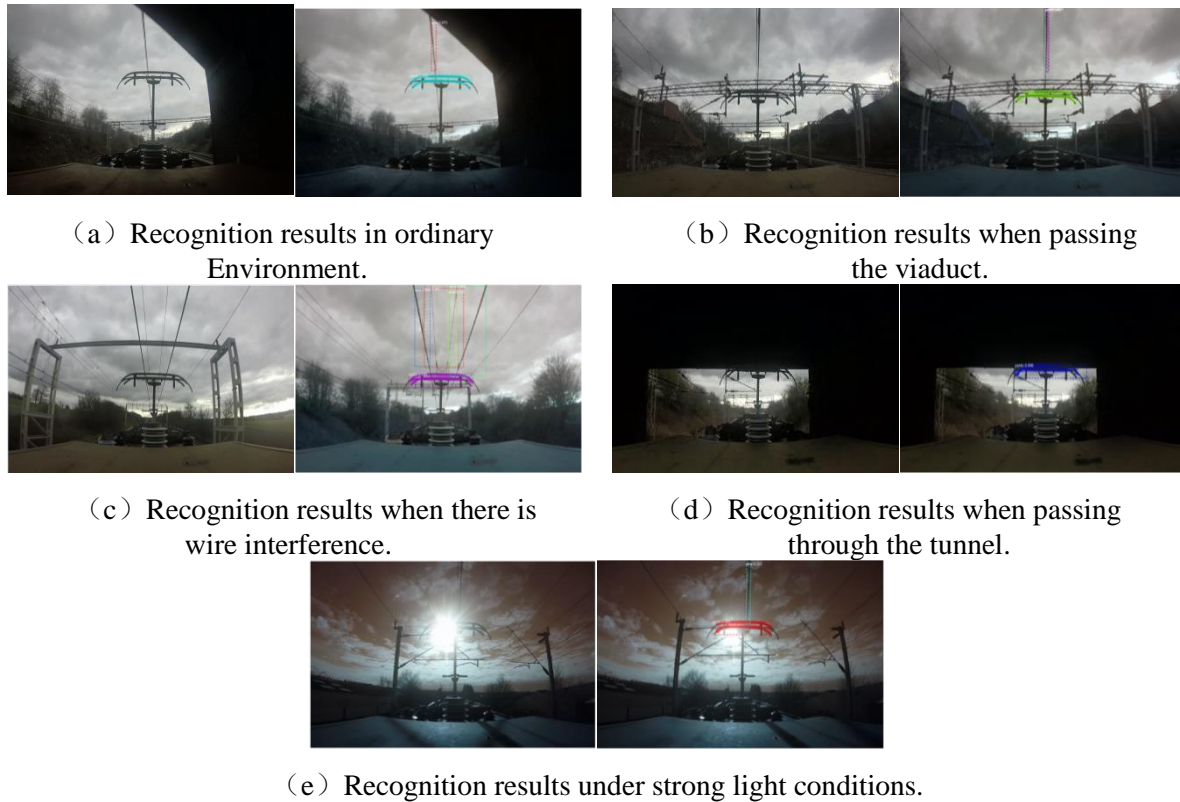
Fig. 4(a) is the pantograph recognition result in the conventional railway environment. The value in the figure represents the recognition matching degree, where the range is [0,1], and the accuracy is one thousandth. The pantograph can be accurately identified without the influence of any environmental factors, and the matching degree reaches 1, while the contact line is also successfully identified with 0.975 matching degree.

During the actual operation of the train, background interference such as load-bearing cables and high-voltage wires as shown in Fig.4(b) and 4(c) may occur. It can be seen from Fig. 4(b) that the complex environmental background will not affect the recognition of the pantograph. The recognition matching degree of the pantograph still reaches 0.999, but the recognition of the contact line in Fig.4(c) is confused.

Fig.4(d) is the train passing through the tunnel, the brightness of the original image is extremely dark due to no lighting conditions. In contrast, Fig.4(e) shows the pantograph over-brightening when the train is facing the sun, in which the pantographs are vaguely visible; But the recognition matching degree of pantograph and contact line at this point still maintaining a high level which is over 0.96, except when the contact line is completely unrecognisable.

After the extensive experiments, we can conclude that the 5G signal performs significantly better than the traditional transmission method in the UK railway environment, with faster speeds and lower packet loss, and the Mask-RCNN algorithm is highly resistant to environmental conditions of extreme

brightness, basically achieves real-time recognition of trains through caves, tunnels, backlighting and night-time passage and other environmental conditions., which ensures the safety of the trains.



**Figure 4.** Recognition results.

## 5. Conclusion

In this paper, the experiments are carried out in real-time communication and data processing for real-time monitoring of pantographs during train operation. The transmission test of high-speed trains under different technical conditions shows that the transmission latency of 5G has been greatly improved from 0.65ms of 4G to 0.34ms. The received pantograph and contact line images are segmented by the Mask-RCNN algorithm. The results show that the algorithm has universal applicability, the matching degree of pantograph recognition under complex background conditions and different brightness conditions reaches more than 0.975. However, in the case of large environmental interference factors, there may be a slight deviation in the identification of the contact line. In the future, this design can be employed in online recognition once the UK's 5G infrastructure is complete.

## Acknowledgements

The research was supported by Innovate UK “Holistic Pantograph and Overhead Line Monitoring System (HPOMS)”

The authors would also like to thank Transmission Dynamics and Angel Trains for their support and provision of data.

## References

- [1] R. Girshick, “Fast r-cnn”, Proceedings of the IEEE International Conference on Computer Vision, pp.1440 – 1448, 2015.

- [2] Q. Zhang, X. Chang and S. B. Bian, "Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN," in *IEEE Access*, vol. 8, pp. 6997-7004, 2020
- [3] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017
- [4] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta and P. Popovski, "Five disruptive technology directions for 5G," in *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74-80, February 2014
- [5] S. Kumar, A. S. Dixit, R. R. Malekar, H. D. Raut and L. K. Shevada, "Fifth Generation Antennas: A Comprehensive Review of Design and Performance Enhancement Techniques," in *IEEE Access*, vol. 8, pp. 163568-163593, 2020
- [6] A. Tusha, S. Doğan and H. Arslan, "A Hybrid Downlink NOMA With OFDM and OFDM-IM for Beyond 5G Wireless Networks," in *IEEE Signal Processing Letters*, vol. 27, pp. 491-495, 2020
- [7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017
- [8] Y. Zhang, J. H. Han, Y. W. Kwon and Y. S. Moon, "A New Architecture of Feature Pyramid Network for Object Detection," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1224-1228
- [9] K. He, X. Zhang, S. Ren, et al. "Deep residual learning for image recognition," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770-778, 2016.
- [10] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936-944,
- [11] M. Overgaard Lauersen, B. Köylü, B. Haddock and J. A. Sorensen, "Kidney segmentation for quantitative analysis applying MaskRCNN architecture," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 2021, pp. 1-6
- [12] Songhui, S. Mingming and H. Chufeng, "Objects detection and location based on mask RCNN and stereo vision," 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 2019, pp. 369-373
- [13] X. Siheng et al., "Power Equipment Recognition Method based on Mask R-CNN and Bayesian Context Network," 2020 IEEE Power & Energy Society General Meeting (PESGM), 2020, pp. 1-5