Railway Track Circuit Signal State Check Using Object Detection

Bin Li*, Suining Wu†, Zhixin Wang†, Xiaoguang Chen†, Long Shi†, and Shulin Tan†

†Institute of Intellective Technology, CRSCD, Beijing, Beijing, 100071, China
*Corresponding author’s e-mail: libin3345@crscd.com.cn

Abstract. The track circuit is widely used in the railway network for safety, and the corresponding track circuit equipment is laid almost every railway station, which provides an extremely important information service for the safe driving of the train. In this paper, we propose a two-phase detection algorithm for the state of the signal equipment of the track circuit, which is used for automatic safety detection of the status of the track circuit signal equipment in the railway signal room. Compared with the traditional inspection method of inspection workers, our proposed detection algorithm of signal state greatly improves the efficiency of security inspection, and in the experimental results, the proposed method outperforms other methods in different situations.

1. Introduction

Nowadays, railway networks are more and more busy. Railway track circuit is very important in the automatic control system for railway networks. Railway track circuit provide free and occupancy information of track section, and serve as the information transmission channel between ground equipment and on-board equipment [1]. Transmitter, receiver and attenuator are three important parts of track circuit, so it is very important to monitor their working state.

In this paper, we proposed a method to monitor the signal lights of attenuator. Signal lights on the attenuator denotes the working state of circuits. For instance, light turns off when circuit problem occurs, as shown in Figure 1. So far, signal light monitor are visually monitored by skilled workers [2]. However, manual detection is lack of efficiency. Signal light flickers in high frequency, which makes worker uncomfortable in eyes. Visual detection has the advantages of high-accuracy, high-speed and automation. Hence, visual detection is a good solution.

In recent years, deep learning has aroused an upsurge in academia and industry. There are many excellent research and applications based on convolutional neural networks (CNN) [3]. Detection methods based on CNN have a strong ability of learning and generalization. Thereafter, CNN based methods are far more accurate and popular than traditional methods. With ingenious design, CNN based methods are able to solve complex computer vision tasks, for instance, automatic driving, face recognition and so on.

Based on the above analysis, we proposed a CNN based detection method to monitor the signal lights of attenuators. CNN based detection methods can be divided into two categories [4]: one-stage methods and two-stage methods. Two-stage methods consist of two parts, where the first part generates a determined quantity of proposal regions, then the second part determine the accurate bounding box and class label according to the aligned feature maps of proposal regions. One-stage methods simultaneously predict the bounding box and class label using convolutional networks.
Generally, one-stage methods are significantly fast than two-stage methods. However, accuracy of one-stage detectors trails that of two-stage detectors. Recently, numerous effective techniques are proposed to further improve the accuracy of one-stage detectors. Some of them are comparable to state-of-art two-stage detectors.

For the detection of attenuator, detection accuracy and speed are quite important. YOLOv3 [5] is one of the most successful one-stage detectors. The architecture of YOLOv3 is concise, it is convenient to modify YOLOv3 according to our demands. Detail works of this paper are as follows:

- We improved the neural network architecture in YOLOv3 to improve the detection performance.
- We proposed a two-phase detection method to improve the performance of YOLOv3 to detect quite small objects.

2. Algorithm

Redmon et al. [5, 6, 7] proposed three versions of YOLO. The end-to-end YOLO networks avoid using region proposal methods. Unlike one-stage detectors, YOLO networks utilized raw images as the input and output the bounding boxes and classifications of the objects. The detection time of YOLO networks is significantly less than two-stage detectors. YOLO detectors perform fast and accurate at the same time. Moreover, several tiny YOLO networks are proposed to realize real-time and high-accuracy detections. YOLO detectors are popular in industrial applications, such as automatic driving, face recognition, robotics and so on.

However, localization errors occur when YOLO performs classification and localization simultaneously. YOLOv1 [6] and YOLOv2 [7] inference the class and bounding boxes via the highest feature maps. However, the last convolutional layer is often spatially course, which is not sufficient for localization. Inspired by FPN [8], residual network and multi-feature map detection are used in YOLOv3 to improve the detection accuracy. The detection accuracy for small and medium size objects is improved by YOLOv3. While, the detection accuracy for large size objects is slightly less than the previous versions.

Another challenge to improve YOLO detectors is foreground-background class imbalance at training. Lin et al. [9] devised focal loss to solve foreground-background class imbalance problem. With this function, detection network focus on more difficult negative examples. However, Redmon report that focal loss cannot improve YOLOv3. Derakhshani et al. [10] proposed Assisted Excitation method to help the detection network train better with the manually excited activations.
YOLOv3 is an end-to-end detection network. The backbone of YOLOv3 is a deep convolutional network of 53 convolutional layers. YOLOv3 use $3 \times 3$ and $1 \times 1$ convolutional layers. YOLOv3 predict the bounding box and class label via three different scale feature maps, which is able to detect objects of different size accurately. With the specially devised anchor boxes, accuracy of YOLOv3 is improved than previous versions.

Inspired by FPN, YOLOv3 upsamples the feature maps of high level to concatenate with feature maps of low level. Naturally, the feature maps of low level are supplanted with semantic information. Thereafter, the accuracy and generalization are further improved. It is sufficient that we propose the improved method based on YOLOv3.

2.1 Boosted Neural Network Architecture
In order to ensure detection accuracy and efficiency, we chose to improve based on YOLOv3 with faster speed and higher accuracy. The traditional image feature-based recognition algorithm has made some attempts to find that it is more susceptible to recognize the features when identifying text, and the robustness of the algorithm is not high enough. Two-stage algorithm in deep learning, like Faster RCNN [11], has higher accuracy but slower speed, so we use YOLOv3, which has a better trade-off between accuracy and speed, and is optimized on one-stage algorithm.

In order to further improve the accurate detection of small-sized signal states, we divide the signal state recognition into two phases. The first phase monitors the appearance detection of all track circuit equipment. As we know, the appearance of a track circuit equipment has a larger scale and a more obvious feature than the background, the accuracy of the detection is generally relatively high. From the perspective of demand, in order to conveniently determine which equipment the subsequent signal lamp corresponds to, it is more reliable to disassemble track circuit equipment for detection at first.

The second phase identifies the status of the signal, which is the next step in the detection of the equipment identified in the first stage. As it is well known, the YOLO algorithm is not ideal for the recognition of small objects. In order to improve the accuracy, we have optimized the network structure based on YOLOv3 and added feature extractors of different sizes, as shown in Figure 3.

| Layers | Channels | Size & Types | Output |
|--------|----------|--------------|--------|
| Convolutional | 32 | $3 \times 3$ conv, stride 2 | 256x256 |
| Convolutional | 64 | $3 \times 3$ conv, stride 2 | 128x128 |
| Convolutional | 32 | $1 \times 1$ conv | 128x128 |
| Convolutional | 64 | $3 \times 3$ conv | 128x128 |
| Residual | | | 128x128 |
| Convolutional | 128 | $3 \times 3$ conv, stride 2 | 64x64 |
| Convolutional | 64 | $1 \times 1$ conv | 64x64 |
| Convolutional | 128 | $3 \times 3$ conv | 64x64 |
| Residual | | | 64x64 |
| Convolutional | 256 | $3 \times 3$ conv, stride 2 | 32x32 |
| Convolutional | 128 | $1 \times 1$ conv | 32x32 |
| Convolutional | 256 | $3 \times 3$ conv | 32x32 |
| Residual | | | 32x32 |
| Convolutional | 512 | $3 \times 3$ conv, stride 2 | 16x16 |
| Convolutional | 512 | $3 \times 3$ conv | 16x16 |
| Residual | | | 16x16 |
| Convolutional | 1024 | $3 \times 3$ conv, stride 2 | 8x8 |
| Convolutional | 512 | $1 \times 1$ conv | 8x8 |
| Convolutional | 1024 | $3 \times 3$ conv | 8x8 |
| Residual | | | 8x8 |

![Figure 2. Network architecture for first phase of detection.](image)
2.2 Two Phase Detection

In the first phase of the classification and detection process, we used the infrastructure of the original YOLOv3 algorithm. Because the appearance of the device is more critical, we have added some layer detector re-adjustment to improve the accuracy of the original YOLOv3 model. The feature map is shown below in Figure 2.

The first stage of training is not much different from the traditional YOLOv3. First, the classifier is trained with $224 \times 224$ input data, and the classifier is retuned with $448 \times 448$ input data to improve the classification accuracy. In order to improve the localization accuracy of pictures taken at different angles and different distances, we have also added a retuning process to the input of different scales for the training of the detector. First, the detector is trained on the input data of $224 \times 224$, and then retuned on the basis of $448 \times 448$. By optimizing the classifier and detector, the algorithm further enhances the recognition of the appearance of the track circuit state equipment.

For the second phase of signal detection, in order to improve the accuracy of small object detection, we added a layer to detect the small object based on the original YOLOv3 algorithm, which can increase the detection performance to a certain extent. First, we found 12 kinds of prior boxes by k-means clustering algorithm, $(10 \times 10), (13 \times 13), (16 \times 16), (23 \times 23), (33 \times 33), (45 \times 45), (61 \times 61), (90 \times 90), (119 \times 119), (156 \times 156), (198 \times 198), (326 \times 326)$, and divided the 12 kinds of prior boxes into 4 groups, and correspondingly assigned to the feature maps at different scales for box regression.

As shown in Figure 3, we have added a new branch in the shallow part of the network structure, mainly for the detection of small objects, in order to improve the detection accuracy of the small signal.
state in the picture. From the final experimental results, after the shallow detection branch is added, the performance of signal state detection for the track circuit have been improved.

3. Experiments
In the experimental part, we firstly use the traditional way to detect the track circuit equipment, but we found that the text recognition and positioning effect is relatively poor, because there are many other text interferences on the track circuit status device panel. Then we use the proposed two-phase algorithm for identification and detection, and we use the original YOLOv3 for comparative experiments.

3.1 Traditional Computer Vision Method
At the beginning, we tried to detect the text of the semaphore using text recognition and the traditional OCR method, and then used the color extraction algorithm to identify the semaphore area. Due to the collected pictures, there are many words and they are easily disturbed. The accuracy of text recognition is not very high, and there are often some cases of missed detection. It can be seen that the combination of detecting the recognition text and the color extraction of the signal light is not very good for detecting the signal of the track circuit state equipment, so we have abandoned this method and adopted the object detection algorithm based on deep learning.

3.2 Two Phase Detection
The two-phase algorithm divides the identification and detection into two parts. The first part mainly identifies the appearance of the track circuit equipment for the collected data. We use the first phase of the appearance detection algorithm to detect the appearances, as shown in Figure 4, the algorithm can be well recognized regardless of whether the appearance of the device is intact or not, and the confidence score is relatively high.

Figure 4. Detect track circuit in Phase 1.

For the second part, based on the appearance of the equipment identified in the first part, we use the network model of the second phase to identify the status of the signal. As shown in Figure 5, for the leftmost target signal detection area, the algorithm for detecting the signal light can accurately identify the correct state of each lamp position and has a very high confidence score.
Figure 5. Detect signal state in Phase 2.

We integrate two phases of detection algorithms to enable fast signal status recognition in actual use. Compared to only YOLOv3 algorithm is used to detect the appearance and signal state at the same time, the integrated algorithm has a higher performance regardless of appearance or signal state recognition. In addition, our algorithm can achieve better results for different angles captured during data acquisition and in the case of different number of track circuit equipment.

Table 1. Results of Detectors.

| Phases | Real-Time Detectors | mAP | FPS |
|-------|---------------------|-----|-----|
| Phase 1 | Ours | 67.4 | 44 |
| YOLOv3 | 64.1 | 45 |
| Phase 2 | Ours | 65.5 | 43 |
| YOLOv3 | 63.4 | 44 |
| Ensemble | Ours | 66.3 | 38 |
| YOLOv3 | 63.8 | 44 |

As shown in Table 1, in the first phase, although the speed of our algorithm is not fast as directly used YOLOv3, but the accuracy has been improved. Similarly, in the second phase of the recognition task, our algorithm accuracy has also been improved, and there is no loss in speed.

3.3 Bad Case Analysis
As far as the experimental results are concerned, the detection performance is better regardless of the shooting angle and the intensity of the light. There is a very small probability of missed detection, the main reason for the missed detection is that the captured picture device is cropped, and the text is too blurred, resulting in an algorithm that does not have a good way to extract features. This result can be improved by adding some fuzzy images and cut pictures to fine-tune the model.

Occasionally, the problem of incomplete detection of the equipment appears when the signal light is recognized. This problem is also a problem of insufficient data volume. We only have 200 pictures for appearance recognition. After the amount of data is sufficient, the problem can be solved.

4. Conclusion
In this paper, we proposed a two-stage detection algorithm for identifying track circuit signal state equipment. The first phase detects the appearance of the equipment panel, and the second phase detects the state of the signal state in the equipment. In this paper, two dedicated network structures are used to extract the corresponding features for the two phases. In the first phase, we optimize the network structure for detecting the appearance of the equipment. In the second phase, we add a new branch for different size object detection, it improved the accuracy of the smaller signal state detection. Experiments show that the proposed algorithm is better than a single target detection algorithm, and
the accuracy and recall rate are higher. The algorithm can be installed in an environment based on a surveillance camera or an automatic inspection robot for automatic and real-time security detection of track circuit equipment.

**Acknowledgments**

This work is supported by China Railway Signal & Communication Research & Design Institute Group Co., Ltd.

**References**

[1] Chen, J., Roberts, C., Weston, P. (2008) Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems. J. Control Engineering Practice., 16(5):585-596.

[2] De Bruin, T., Verbert, K., Babuska, R. (2016) Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks. J. IEEE Transactions on Neural Networks and Learning Systems., 1-11.

[3] LeCun, Y., Bengio, Y., et al. (1995) Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. The MIT Press. Cambridge.

[4] Liu, W., Anguelov, D., Erhan, D., et al. (2016) SSD: Single Shot Multibox Detetor. European Conference on Computer Vision. Amsterdam. pp. 21-37.

[5] Redmon, J., Farhadi, A. (2018) YOLOv3: An Incremental Improvement. https://arxiv.org/abs/1804.02767.

[6] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway. pp. 779-788.

[7] Redmon, J., Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu. pp. 6517-6525.

[8] Lin, T. -Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017) Feature pyramid networks for object detection. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu. pp. 936-944.

[9] Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2017) Focal loss for dense object detection. J. IEEE Transactions on Pattern Analysis & Machine Intelligence., 99: 2999-3007.

[10] Derakhshani, M. M., Masoudnia, S., Shaker, A. H., Mersa, O., Araabi, B. N. (2019) Assisted excitation of activations: A learning technique to improve object detectors. IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles. pp. 9201-9210.

[11] Girshick, R. (2015) Fast R-CNN. IEEE International Conference on Computer Vision. Santiago. pp. 1440-1448