

# An optimized railway fastener detection method based on modified Faster R-CNN

Tangbo Bai <sup>a,b</sup>, Jianwei Yang <sup>a,b,\*</sup>, Guiyang Xu <sup>a,b</sup>, Dechen Yao <sup>a,b</sup>

<sup>a</sup> School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

<sup>b</sup> Beijing Key Laboratory of Performance Guarantee on Urban Rail Transit Vehicles, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

## ARTICLE INFO

### Keywords:

Railway  
Fastener detection  
Image Processing  
Faster R-CNN

## ABSTRACT

Accurate fastener positioning and state detection form the prerequisite for ensuring the safe operation of rail track. The demands for intelligent, fast and accurate detection cannot be satisfied by traditional methods using image processing and fastener classification. In view of this, a two-stage classification model based on the modified Faster Region-based Convolution Neural Network (Faster R-CNN) and the Support Vector Data Description (SVDD) algorithms is proposed in the paper for fastener detection. Firstly, the data set of detection images is built with the images being labeled, and the classification and detection model based on Faster R-CNN is constructed according to the characteristics of practical fastener images. The anchor box optimization function is established by labeled data set to optimize the box of region proposal network in the model, to enhance the detection rate and accuracy of detection. Then, according to the detection result by Faster R-CNN, the SVDD algorithm is applied for the second stage classification of deviated fasteners, which avoids inaccurate classification caused by different deviated angles of fasteners. Through the verification and analysis of practical detection case, it is verified that the proposed method can improve the efficiency and precision of fastener detection with higher detection rates and accuracy in comparison with other baseline detection methods, making it suitable for fast and accurate detection of fastener states.

## 1. Introduction

Railroad fasteners are an important component of rail track systems, and their status directly affects the safety of railway train operations [1]. Some events such as fastener deviation, missing and fracture are likely to occur, due to the large number, poor working environment and complex forces to bear. With the rapid development of railway transportation, it brings more lines, longer mileage and more maintenance difficulty. In such cases, intelligent, fast, and accurate fastener detection techniques are needed urgently [2].

In recent years, extensive research of track detection have been carried out, including methods based on vibration signals [3,4], acoustic signals [5,6] and machine vision [7,8]. The detection methods using vibration and acoustic signals have good performance in tracking surface damage, while difficult to effectively detect failures such as fastener deviation and fracture, which often cause minor vibration and noise. Therefore, the research on fastener detection methods is mainly focused on image acquisition and processing.

Among the research works on fastener detection methods based on image processing, commonly utilized tools include image enhancement, feature extraction, and so on. For example, Fan et al. [9] proposed a line local binary pattern encoding method to represent the key components of fasteners, and the method had a good performance on detecting the fully and partly missing fasteners. Ma et al. [10] extracted fastener edge feature by using a switching median filter and an improved Canny edge detection method, and realized real-time detection of the missing fastener by template matching based on curve feature projection. Ou et al. [11] established a Bayesian hierarchical model of fastener feature words and structure labels, represented the fastener image by a topic distribution, and classified the fasteners using the support vector machine (SVM). The advantage of these methods is high computational efficiency. However, they have low adaptability when dealing with different track line conditions and fastener types, and thus appropriate correction methods are needed for these algorithms.

With the progress of image detection technology, some object detection algorithms based on deep neural network have been applied

\* Corresponding author at: School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China.

E-mail address: [yangjianwei@bucea.edu.cn](mailto:yangjianwei@bucea.edu.cn) (J. Yang).

for track detection [12–16]. In terms of fastener detection, Cui et al. [14] segmented fastener images into different parts to avoid the interference of debris on fasteners, and tested segmentation models in the real-time deep learning module. Wei et al. [8] proposed an improved “*You Only Look Once*” (YOLO) v3 model named Track Line Multi-target Defect Detection Network (TLMDNet) by integrating scale reduction and feature concatenation in order to enhance detection accuracy and efficiency. Xavier et al. [15] introduced a new algorithm for fastener detection based on Convolutional Neural Networks (CNN) and SVM, and the result showed that detection performance could be improved by combining multiple detectors under a multitask learning framework. Wei et al. [16] proposed a novel fastener detection method using Dense-Scale Invariant Feature Transform (Dense-SIFT) and Faster R-CNN to improve detection rate and efficiency. Such methods can directly detect the target region of the image without feature extraction, and thus have strong adaptability.

As the most widely used deep learning network in the field of image processing, CNN has been applied to target detection due to its superior performance, and has formed the algorithm Regions with CNN (R-CNN) [17], which firstly utilizes algorithms of Selective Search [18] or Edge Boxes [19] to generate candidate regions, and then feeds each region into CNN for training and classification. By virtue of the specialty of the target detection problem, 2000 region proposals are needed for CNN training in the R-CNN algorithm, which results in low detection speed and large resource occupation. In order to solve this problem, Fast R-CNN [20] and Faster R-CNN [21] have been proposed, where the former one firstly executed the convolution on the whole image to get feature maps, and then dealt with the candidate regions, thus the times of convolution was reduced. Compared to R-CNN, it can improve the efficiency and realize the end-to-end training, but the calculation speed and accuracy was affected because of the existence of a large number of invalid candidate regions. To deal with this problem, the Region Proposal Networks (RPN) method was applied in Faster R-CNN to generate candidate regions, which replaced the previous segmentation algorithm, thus it improved the efficiency of detection window generation. The Mask R-CNN [22] and Cascade R-CNN [23] were proposed based on Fast R-CNN to address instance segmentation problem and improve detection accuracy [24].

Faster R-CNN, as an end-to-end target detection method, has higher detection accuracy, but it suffers from intensive and inefficient computation when directly used. Methods such as Single Shot MultiBox Detector (SSD) [25], YOLOv3 [26,27], EfficientDet [28] and YOLO v4 [29] can reduce the computation time and improve the efficiency by converting the object detection task into a regression problem, but the accuracy is still lower compared to end-to-end algorithms. One problem in fasteners detection is that there are sufficient data under normal conditions while limited data under faulty conditions are available. The accuracy of the traditional pattern recognition method depends on the sample number and the proportion of all kinds of samples. When training samples are small and the sample number of different states varies greatly, the classification performance is poor. SVDD is a one-class classification method [30–32], which can judge whether a fault exists in a case that only signals under normal working conditions are available.

The above methods have good performance when dealing with specific detection and classification problems. However, practical applications are often mixed with multiple types of cases, and a single method cannot be effective for all cases. In view of this situation, two-stage based methods are applied to solve this problem and show superior performance [33,34]. Jiang et al. [35] proposed a two-stage strategy for fast crack detection of high-resolution images and improved the detection accuracy of blurred cracks, where YOLOv4 was employed to filter out images without cracks and to generate coarse region proposals, then hybrid dilated convolutional block (HDCB) net was used to detect pixel-level cracks from the coarse region proposals. Ge et al. [36] incorporated R-CNN Gradient Annealing (RGA) and Parallel R-CNN Modules (PRM) as a two-stage strategy, to enhance the impact of

positive proposals and solve the imbalance issue in object detection models. These methods provide ideas for solving complex detection problems.

In view of this, a fastener detection algorithm based on the modified Faster R-CNN model is proposed in this paper. Region proposal network of Faster R-CNN is optimized by characteristic information of fastener images. The algorithm can simultaneously improve the computational accuracy and the operation speed, which lay a foundation for further real-time state detection of fasteners. Then, a two-stage classification model based on Faster R-CNN and SVDD algorithm is established. The Faster R-CNN model is firstly used to recognize fastener states (such as fracture, missing, etc.) that have remarkable differences, and to distinguish normal state and deviation state that can be easily confused, second stage classification is performed, so as to improve the detection accuracy of fastener states. The effectiveness of the proposed method in fastener defect detection is verified by the result analysis of practical detection images of railway fastener, and the corresponding accuracy is shown to be higher than traditional methods.

The main contributions of this paper are as follows: 1) An optimized RPN network of Faster R-CNN model is proposed for fastener detection, where labeled data set is used to construct the anchor box optimization function. It reduces the sample number and overlaps of anchor boxes, and improves the computation efficiency and detection accuracy of Faster R-CNN. 2) Considering the detection difficulty of deviated fasteners, a two-stage classification model based on Faster R-CNN and SVDD algorithm is established, where fasteners can be recognized into two types, incomplete and complete ones, and they are sequentially dealt with respectively. It avoids the interference of deviated angles to the training model at the state of fastener deviation, and enhances the efficiency and accuracy of fastener identification, which is of practical significance in ensuring the safe operation of the rail transit vehicles.

The rest of this paper is organized as follows. The theoretical background of Faster R-CNN and SVDD is first discussed in Section II. The theoretical framework of the proposed method is then presented in Section III. Next, the effectiveness of the proposed method is demonstrated in Section IV by practical application. Finally, the conclusions are drawn in Section V.

## 2. Theoretical background

### 2.1. Theoretical model of Faster R-CNN

The scheme of Faster R-CNN [21] is shown in Fig. 1, which mainly contains three parts: feature extraction, candidate box extraction, and boundary box regression and classification. When doing object detection, convolution layers are first applied to extract feature maps. Through the RPN network, the context attribute of anchor points is judged by Softmax, and then bounding box regression is used to correct the anchor point, so as to generate a relatively accurate candidate region. Finally, the region of interest (ROI) pooling is performed on the feature maps extracted by convolution layers and the candidate region to extract the candidate feature maps, and further classification and bounding box regression are carried out to output the information of classification and location.

It can be seen from Fig. 1 that RPN plays a key role in the Faster R-CNN structure, and the number and accuracy of generated candidate regions have a direct influence on the computational efficiency and detection accuracy of the entire network. RPN is the major modification of the previous Faster R-CNN algorithms. Its role is to use the feature maps extracted by CNN from the original image as the input and to output high-quality candidate regions with different sizes and length-width ratios.

As shown in Fig. 2, the RPN generates candidate regions based on sliding window mechanism instead of selective searching. Through the sliding window, the convolution feature layer gets a 256-dimensional eigenvector. When it slides on the convolution feature map, the center

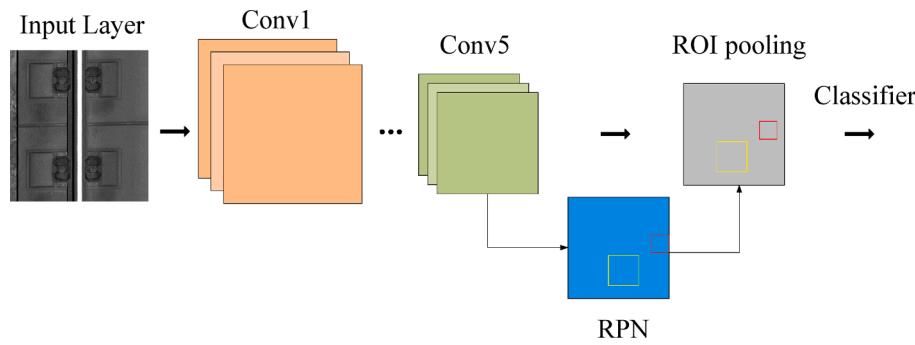


Fig. 1. Framework of Faster R-CNN.

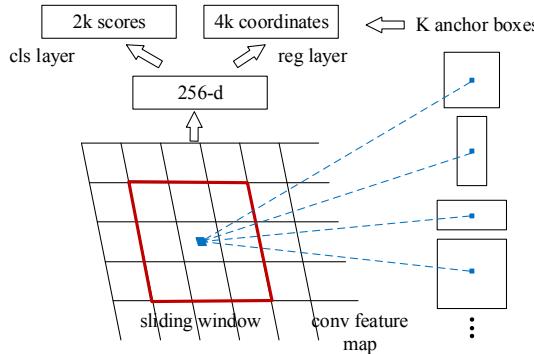


Fig. 2. Framework of region proposal network [21]

point is seen as an anchor, and  $k$  anchor boxes are obtained at each position. Then, through two full connections one can get 2 k scores and 4 k coordinates where the scores are used to evaluate whether detection targets are contained in the candidate region, the coordinates including the length  $w$  and width  $h$  of the candidate region, and the central position coordinates  $(x, y)$ . The setting of the number, the size, the length and width of anchor boxes can affect the output performance of the entire network. In the original algorithm [21],  $k$  was taken as 9, namely, 9 anchor boxes were determined at each anchor point. In this work, the corresponding sizes are selected as  $128 \times 128$ ,  $256 \times 256$  and  $512 \times 512$  respectively, and the height-width ratios are set as 1:1, 1:2 and 2:1, respectively.

The RPN loss function  $L$  is composed of bounding box classification loss  $L_{cls}$  and bounding box regression loss  $L_{reg}$ , and is defined as Eq. (1):

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \frac{1}{N_{reg}} \sum_i t_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

$$L_{cls}(p_i, p_i^*) = -\log[(p_i, p_i^*) + (1 - p_i)(1 - p_i^*)] \quad (2)$$

$$L_{reg}(t_i, t_i^*) = \sum_i \text{smooth}_{L1}(t_i, t_i^*) \quad (3)$$

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & |x| \geq 1 \end{cases} \quad (4)$$

where,  $L_{cls}$  and  $L_{reg}$  are calculated as Eq. (2) and (3),  $\text{smooth}_{L1}$  in Eq. (3) denotes the loss function, which is calculated by Eq. (4).  $N_{cls}$  is the minimum number of batch images input to the network, and  $N_{reg}$  is the total number of anchor coordinates, both of which are normalized weight parameters.  $p_i$  represents the probability of the  $i$ -th anchor box that is the target, and  $p_i^*$  is the sample label.  $t_i$  and  $t_i^*$  are the predicted and real bounding box coordinate respectively, and the corresponding calculation formulas are shown as Eq. (5):

$$\begin{cases} t_x = \frac{(x - x_r)}{w_r}, t_y = \frac{(y - y_r)}{h_r} \\ t_w = \log\left(\frac{w}{w_r}\right), t_h = \log\left(\frac{h}{h_r}\right) \\ t_x^* = \frac{(x^* - x_r)}{w_r}, t_y^* = \frac{(y^* - y_r)}{h_r} \\ t_w^* = \log\left(\frac{w^*}{w_r}\right), t_h^* = \log\left(\frac{h^*}{h_r}\right) \end{cases} \quad (5)$$

where,  $(x, y)$  is the predicted bounding box coordinate,  $(x_r, y_r)$  is the candidate box coordinate, and  $(x^*, y^*)$  is the bounding box coordinate of real target. Accordingly,  $w$  and  $h$  are the width and height of the box.

## 2.2. SVdd

As a One-Class Classification (OCC) method, SVDD was proposed by D.M.J.Tax and R.P.W.Duin [37]. The theoretical basis of SVDD is derived from SVM. The basic idea of SVDD is: to regard the target objects as a whole, and to construct a closed compact region  $\Omega$ , which contains all target objects or as many as possible, and non-target objects are excluded from  $\Omega$  or few of them are included in it. Let us assume that a dataset  $S$  containing  $n$  target objects  $\{x_i\}$ ,  $i = 1, 2, \dots, n$ , where  $S \subseteq X, X \subseteq R^d$ .  $\Phi: X \rightarrow F$  is the mapping from original space  $X$  to the high-dimensional feature space  $F$ . In practice, the dimension of  $F$  may be very high, and kernel function of SVM can be utilized to solve such problem subtly. According to relevant functional theory, as long as one kernel function  $K(x_i, x_j)$  meets Mercer condition, it is corresponding to a certain inner product in transformed space,  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ . Therefore, in high dimension space, only inner product calculation is required, and the inner product calculation can be realized by the function in original space instead of knowing the specific form of  $\Phi(x)$ . It aims to find a hyper-sphere with smallest volume, which contains all  $\Phi(x_i)$ . The hyper-sphere can be represented by its center  $a$  and radius  $R$ , and it satisfies:

$$\text{mine}(R, a, \xi) = R^2 + C \sum_{i=1}^n \xi_i, \quad i = 1, \dots, n \quad (6)$$

with corresponding constraint:

$$\|\varphi(x_i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \quad (7)$$

where,  $\xi_i \geq 0$ ,  $i = 1, \dots, n$  is slack variable which ensures the classification robustness.  $C$  is a given constant, which serves as the trade-off between model simplicity and the number of target objects being rejected. After the above problem is transformed into Lagrange extremal problem, it can be represented by the Lagrange polynomial dual form [37]:

$$L(R, a, \xi, \alpha, \gamma) = \sum_{i=1}^n \alpha_i K(x_i, x_i) - \sum_{i=1, j=1}^n \alpha_i \alpha_j K(x_i, x_j) \quad (8)$$

where,  $a_i \geq 0$  is Lagrange coefficient, and kernel function  $K(x_i, x_j)$  can be used to represent inner product calculation  $\Phi(x_i) \cdot \Phi(x_j)$ . In practice, most  $a_i$  is equal to 0 and only few  $a_i$  is positive.  $a_i$  not equal to 0 is a support vector, which can be used to determine the values of  $a$  and  $R$ . The non-support vectors that correspond to  $a_i$  being 0 are neglected within the calculation, so such method possesses high calculation efficiency. The radius of the hyper-sphere  $R$  can be calculated by any support vector  $x_k$  using Eq. (9).

$$R^2 = K(x_k, x_k) - 2 \sum_{i=1}^n a_i K(x_i, x_k) + \sum_{i=1, j=1}^n a_i a_j K(x_i, x_j) \quad (9)$$

Discriminant function can be used to make a judgment on whether a new sample  $z$  belongs to target sample, and it is shown as follows:

$$f(z) = \|\varphi(z) - a\|^2 = K(z, z) - 2 \sum_{i=1}^n a_i K(z, x_i) + \sum_{i=1, j=1}^n a_i a_j K(x_i, x_j) \quad (10)$$

For Eq. (10), if  $f(z) \leq R^2$ ,  $z$  belongs to a target sample and it is accepted; otherwise, it is not a target sample and is rejected.

### 3. Proposed method

#### 3.1. Framework of the proposed approach

The key problem of fastener detection lies in the accurate fastener positioning and effective classification. To realize it, a two-stage classification detection algorithm based on Faster R-CNN model is proposed in this paper. Anchor box scale optimization model is built based on the characteristics of fastener images in order to improve the accuracy and efficiency of fastener detection. Then a classification model is constructed by combining Faster R-CNN and SVDD method, and the incomplete and complete fasteners are respectively dealt with to enhance the classification accuracy. The framework of the method is shown in Fig. 3. It is composed of several modules, including data acquisition, model optimization, and classification for incomplete fasteners and deviated fasteners.

As illustrated in Fig. 3, track image detection is first conducted in this paper, where the images are grouped and labeled, and data sets are built. Then, the RPN structure is optimized according to the labeled information, and a detection model based on Faster R-CNN is established to realize accurate and fast fastener positioning and detection of incomplete fasteners. Finally, extract image features of complete fasteners output from Faster R-CNN, and distinguish them as deviated and normal ones by SVDD classification model, so as to improve the classification accuracy of deviated fasteners.

#### Algorithm 1: The main steps of the proposed method

Input: Images of the fasteners, labels of the images

Output: The classification results of the fasteners

1 Initialize the parameters and send the image to the backbone network Resnet101.

(continued on next column)

(continued)

#### Algorithm 1: The main steps of the proposed method

- 2 The output feature map of conv4\_x in Resnet101 is shared by RPN and RoI Pooling.
- 3 Compute the relative scale  $s$  and length-width ratio  $r$  using Eq. (11) and 12, and generate optimization candidate boxes.
- 4 Compute the RPN loss function using Eq. (1) to update parameter of the bounding box.
- 5 Collect output proposal boxes of RPN by Pooling layer, calculate proposal feature maps and send them into the follow-up network.
- 6 Average pooling is performed after conv5\_x to get 2048-dimension feature for classification and regression.
- 7 Three states of fasteners are obtained through the classifier: complete fasteners, fracture fasteners and missing fasteners.
- 8 Train the SVDD classifier by samples of normal fasteners using Eq. (6) ~ (10), to identify normal fasteners and deviation fasteners from complete ones in step 7.

The structure of the proposed Faster R-CNN network is shown in Fig. 4. The Resnet101 is used as the backbone network of Faster R-CNN network, and it is divided in five scales, which are conv1, conv2\_x, conv3\_x, conv4\_x and conv5\_x. The output of conv4\_x is shared by RPN and RoI Pooling, and conv5\_x is applied to the feature map of dimension  $14 \times 14 \times 1024$  by RoI Pooling, which satisfies the input demand of conv5\_x. Average pooling is performed after conv5\_x to get 2048-dimension feature for classification and regression. The modified RPN network and SVDD classifier are described in Section 3.2 and 3.3. The main steps of this algorithm are described in Algorithm 1.

#### 3.2. RPN optimization

When doing fastener positioning based on the track detection images, the ratio between images and fasteners remains consistent. So, being different with traditional target detection problems, it is not necessary to generate a large number of anchor boxes of different sizes and ratios at each anchor point when RPN is applied for candidate region selection. Instead, the sizes, ratios and overlaps of anchor boxes can

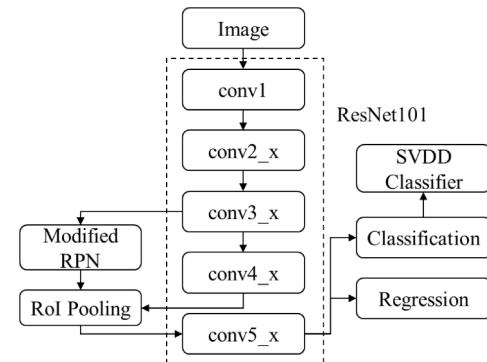


Fig. 4. The structure of the proposed Faster R-CNN.

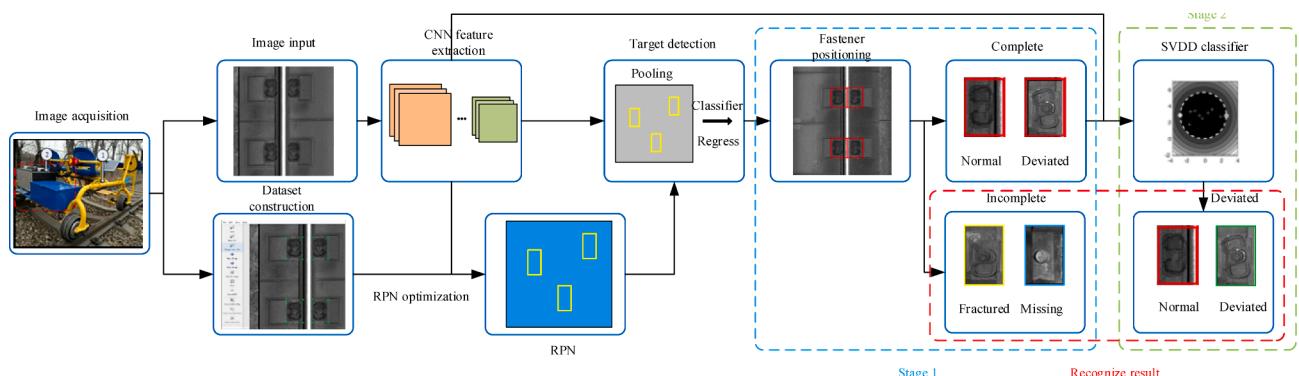


Fig. 3. Framework of the proposed method.

be optimized according to the location and distribution information of the fastener in the image. Since the sizes of images and fasteners are different under different acquisition parameters and scaling ratios, relative coordinates are used instead of absolute ones to optimize the anchor box scale in this paper.

Suppose that  $n$  fasteners are labeled in the training data set, among which the candidate region of the  $i$ -th fastener is of length  $w_i$ , width  $h_i$ , and central position coordinates  $(x_i, y_i)$ , respectively. The overall length and width of the image are  $w_0$  and  $h_0$ . Since the relative scale  $s$  and length-width ratio  $r$  are consistent between fasteners and images, they are defined as follows in this paper:

$$s = \text{round} \left( \frac{1}{n} \sum_i \frac{w_i h_i}{w_0 h_0} \right) \quad (11)$$

$$r = \frac{1}{n} \sum_i \frac{w_i}{h_i} \quad (12)$$

During the generation of candidate boxes, there may be overlaps. In this paper, an non-maximum suppression (NMS) method is adopted for candidate box deletion. Since there is no overlap for practical fastener, the threshold value is set to 0 in this paper, and the calculation procedures are as follows:

- (1) Choose the candidate box with the highest score;
- (2) Calculate the value of Intersection over Union (IoU) between each of the remaining boxes and the one chosen in step (1). If the value is above the set threshold, the corresponding candidate box is considered invalid and will be deleted.
- (3) Continue to choose the box with highest score among those not yet deleted, and repeat the above steps (1)-(2) until there is no candidate box to delete.

At the same time, since the parameters in the model updating speed are determined by vectors, to avoid the experience given by vectors in running models without best performance, adaptive vectors are used in the experiment to increase the speed of model optimization. The initial value is set to 0.001. During the process of training, after each epoch, the loss and precision of the current model are evaluated in the validation set. The changes of loss value are observed at every other epoch, and when it is less than 0.0001,  $lr$  is adjusted by Eq. (13):

$$lr^* = lr^* \gamma \quad (13)$$

where,  $\gamma$  represents the adjustment coefficient, and it is set to 0.1 in the paper.

### 3.3. A two-stage classification model based on SVDD

The types of incomplete and complete fasteners and their corresponding box coordinates are provided by the Faster R-CNN. Fig. 5 shows two classes of complete fasteners, including the normal ones and deviated ones. Since the deviation degrees vary for deviated fasteners, the difference between deviated samples may be greater than that between the normal fastener and a deviated sample with low degree deviation, when a traditional two-class classification algorithm is applied for training. If the Faster R-CNN is directly used for classification, it needs to introduce multilevel information of deviation angles, which

greatly increases the network complexity and the computation time. Therefore, in this paper, the SVDD algorithm is used for two-stage classification, where only the samples of normal fasteners are used for training, and then, to judge whether the fasteners are normal or not by the model. If the judgment requirement is not satisfied, it is regarded as a deviated fastener. It can avoid the interference by introducing the angle information of deviated fasteners.

After the training process is carried out, the image feature can be extracted first. Feature extraction is performed on the regional images of complete fasteners generated by the Faster R-CNN in this paper. The image histogram information can effectively represent the distribution and change of objects in the image, and it has been successfully applied to equipment state detection [38,39]. Thus, histogram information is chosen in this paper for feature extraction. And gray histogram information is calculated by Eq. (14) and the Equations of histogram features are listed in Table 1.

$$P(g) = \frac{N(g)}{M}, g = 0, 1, \dots, L \quad (14)$$

where,  $g$  represents the gray level,  $N(g)$  is the number of pixels with gray level  $g$  in the image,  $M$  denotes the total pixels of the image, and  $L$  is the maximum value of gray level in the image.

When the SVDD algorithm is used for fastener classification, the images of normal fasteners in the training data set are first processed to obtain image features, and the SVDD model is trained to obtain the sphere radius  $R$ . Then the features of images to be tested are calculated and fed into the trained SVDD model, and the distances between the test sample and center point are calculated. If the distance is smaller than or equal to the sphere radius  $R$ , it is regarded as a normal fastener; otherwise, it is considered as a deviated fastener.

## 4. Case studies

### 4.1. Image acquisition

A new intelligent track inspection instrument designed and manufactured by a company in Beijing is used in this paper for subway track detection. This instrument contains two parts: an electric inspection vehicle and a track status inspection system, as shown in Fig. 6. The electric inspection vehicle is composed of four small rail wheels, a car body and two seats. The track status inspection system consists of a host

Table 1

The expression of histogram features of the image.

Index name	Expression
Mean	$HF_M = \sum_{g=0}^{L-1} g P(g)$
Standard deviation	$HF_{SD} = \sqrt{\sum_{g=0}^{L-1} (g - HF_M)^2 P(g)}$
Skewness	$HF_S = \frac{1}{\sigma_g^3} \sum_{g=0}^{L-1} (g - HF_M)^3 P(g)$
Kurtosis	$HF_k = \frac{\sum_{g=0}^{L-1} (g - HF_M)^4}{\sigma_g^4}$
Energy	$HF_{EG} = \sum_{g=0}^{L-1} [P(g)]^2$
Entropy	$HF_{EP} = - \sum_{g=0}^{L-1} P(g) \log_2 [P(g)]$

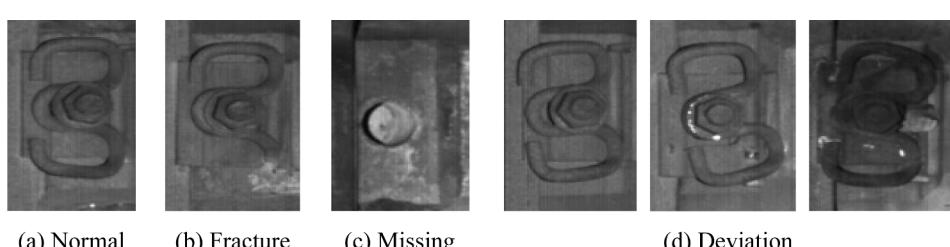
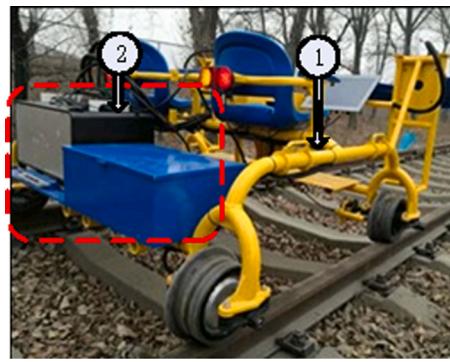


Fig. 5. Different states of fasteners.



**Fig. 6.** New intelligent track inspection instrument 1) electric inspection vehicle 2) track status inspection system.

and two high-definition line array image scanning modules.

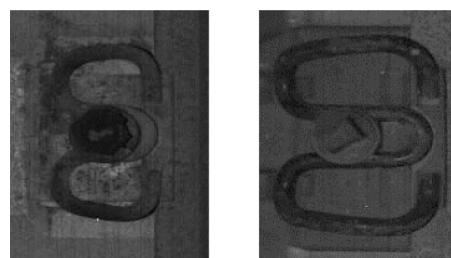
When the inspection instrument is at work, 1–2 operators shall ride and operate the vehicle forward to the inspection area and start to work. The acquisition speed of the inspection system is 20 km/h and the total weight is about 300 kg. In addition, the inspection system has the following characteristics: 1) it can dynamically collect track surface data, and can cover the entire track structure surface; 2) it is equipped with accurate detection technology, with the mileage positioning precision of 0.3 m; 3) it has a high-definition image acquisition sub-system, with the pixel resolution up to 0.3 mm; 4) it is equipped with lightweight and structured design, which is simple and convenient to operate, and is quickly assembled/disassembled.

In this paper, the test is performed on the main lines of three railways, and the testing length is 30 km. Data from four types of fasteners are collected to verify the algorithm, as shown in Fig. 7. The images are cropped with the same resolution  $160 \times 200$  from the original ones, and the relative size of different fasteners is consistent with that of the real ones. The four types of fasteners are marked as Type I ~ IV in the paper. Type I and II are similar in appearance and different in size, and they are more widely used; Type III and IV are mainly used at railway turnouts and are used less.

#### 4.2. Data set construction

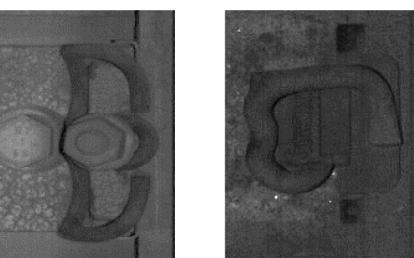
From the data obtained in the test, 4200 images are randomly chosen in this paper, and the number from Type I ~ IV is 1500, 1500, 600 and 600 respectively. Half the images of each fastener are taken as the training image data set, and the rest as the test image data set.

When constructing the training data set, the object detection labeling is an important step in image detection. LabelImg software is applied for image labeling in the test and the effective fastener in the single image (e.g. Fig. 8) is considered as object detection region, as shown in Fig. 9. Fasteners can be labeled as complete, fractured and missing, and complete fasteners include normal and deviated ones. After labeling, the coordinates of ROI for fastener target in the detection image can be



(a) Type I

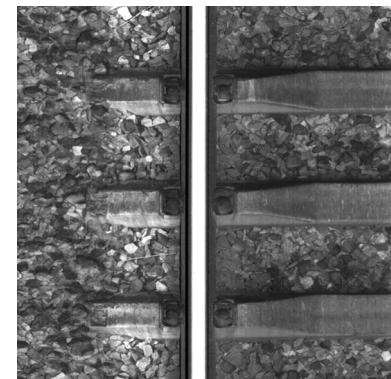
(b) Type II



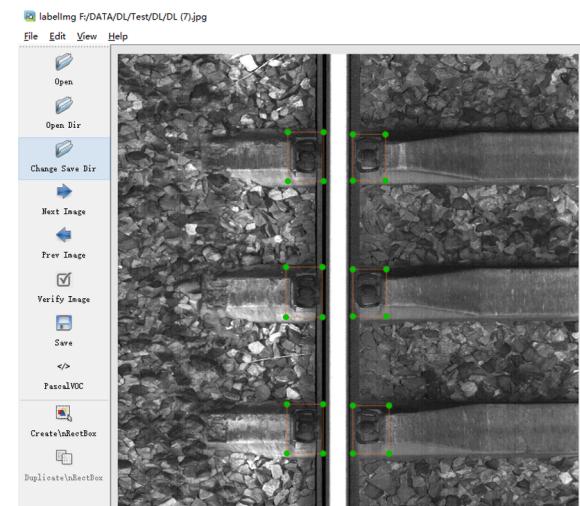
(c) Type III

(d) Type IV

**Fig. 7.** Four types of fasteners.



**Fig. 8.** Track detection image.



**Fig. 9.** Image annotation.

obtained in order to generate the coordinate data set. Network optimization, training and testing can be carried out based on the data set.

#### 4.3. Result analysis

##### 4.3.1. Effect of Faster R-CNN optimization

In this paper, the fastener detection model operates under the hardware environment with GPU of NVIDIA RTX2080, CPU of Intel i7 9700, and memory of 32 GB. The software environment is based on Windows 10, Tensorflow, Python 3.6, CUDA 9.0 and cuDNN 7.3.

The iteration number is set to 4000 times in the training process in this paper. After model training, the test data set of track detection image is fed into the training model, to obtain the detection result of fastener. Fig. 10 illustrates the results of fastener detection by the

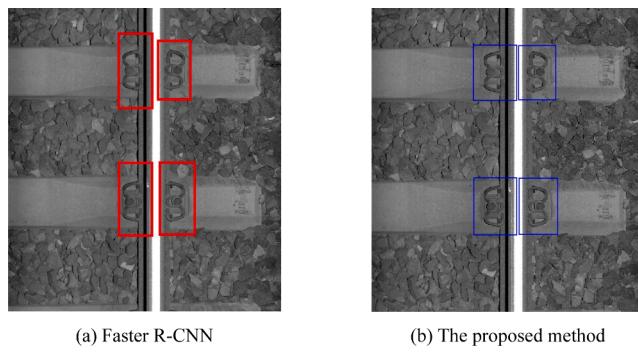


Fig. 10. The result of fastener detection.

original Faster R-CNN method and the method proposed in this paper.

It can be seen from Fig. 10 that fastener detection can be achieved by the original Faster R-CNN method, but due to the randomness of RPN network generation, the sizes of output positioning boxes are different, which is not convenient for subsequent processing. To compare the two methods more precisely, the evaluation indicator IoU commonly used in target detection tasks is applied in this paper, and IoU of single target to be detected is calculated by Eq. (15):

$$\text{IoU} = \frac{A_I}{A_U} \quad (15)$$

where  $A_I$  is the intersection area between bounding box and ground truth, and  $A_U$  is the union area between them. IoU can be used to evaluate the accuracy of target detection, and the larger value means the better performance of the method. The average of IoUs of all detection targets is applied for evaluation in this paper.

The average values of IoUs for 4 types of fasteners on their respective datasets are shown in Table 2. It can be seen that the proposed method has higher average value of IoU, and it performs better for every type of fastener. Through the modified network, the positioning for fastener is more accurate, and the sizes of output boxes are more uniform, which is convenient for subsequent processing.

#### 4.3.2. Effect analysis of fastener classification

In order to evaluate the results of fastener detection, five indicators are selected: detection rate  $R_d$  and accuracy  $R_a$  for single type of fasteners, detection rate  $R_{do}$  and accuracy  $R_{ao}$  for the overall types of fasteners, and the detection speed frame per second  $FPS$ .  $R_d$  denotes the proportion of the fasteners correctly detected to all abnormal fasteners, and  $R_a$  refers to the proportion of the fasteners correctly detected to all detected abnormal fasteners. They can be calculated by Eq. (16) ~ (19):

$$R_d = \frac{F_d}{F_n + F_d} \quad (16)$$

$$R_a = \frac{F_d}{F_w + F_d} \quad (17)$$

$$R_{do} = \frac{\sum N_i R_{di}}{\sum N_i} \quad (18)$$

$$R_{ao} = \frac{\sum N_i R_{ai}}{\sum N_i} \quad (19)$$

where,  $F_d$  represents the number of abnormal fasteners correctly

Table 2

The average of IoUs of 4 types of fasteners.

Method	Type I	Type II	Type III	Type IV
Faster R-CNN	0.726	0.733	0.664	0.677
The proposed method	0.871	0.882	0.829	0.843

detected,  $F_n$  denotes the number of undetected abnormal fasteners,  $F_w$  represents the number of abnormal fasteners incorrectly detected,  $N_i$ ,  $R_{di}$  and  $R_{ai}$  represent the number,  $R_d$  and  $R_a$  of fasteners of each type respectively.

When the Faster R-CNN is directly used for fastener detection, the detection performance can be guaranteed for fractured and missing fasteners, since both fault types have more obvious distinctions with normal ones. However, for deviated fasteners, misclassification and aliasing are likely to happen, as shown in Fig. 11. After classification by the modified Faster R-CNN, the SVDD model is applied to complete the classification of deviated fasteners and to avoid the interference to classification model caused by different deviated angles of fasteners. The corresponding performance is more satisfying, as shown in Fig. 12. The classification accuracy by the above two methods is shown in Fig. 13, including four fastener states of normal, fractured, missing and incomplete. Since YOLO V4 is currently the state-of-the-art method for object detection, it is used to process this same data set as a contrast and the classification accuracy is shown in Fig. 13 too.

It can be observed from Fig. 13 that, YOLO V4 takes into account both efficiency and precision, its accuracy rate is close to Faster R-CNN when the speed is far ahead, showing good performance, but the accuracy is low in the classification of deviated fasteners which is concerned in this paper. There is little difference between the two Faster R-CNN based methods in the classification of fractured and missing fasteners, and the accuracy of the proposed method in this paper is improved to some degree because it can locate the fasteners more precisely. In the classification of deviated fasteners, partially deviated fasteners are confused with normal ones by the Faster R-CNN method, while the proposed method using two stage strategy achieves more accurate detection, which verifies the classification effectiveness for deviated fastener by the second stage. For further comparison and analysis, the commonly used SSD, YOLO V4 with backbone CSPResNeXt50 and CSPDarknet53, Mask R-CNN, Cascade R-CNN and the traditional Faster R-CNN algorithm are also used to process the data set in this study, and the results are shown in Table 3.

From the results of fastener detection by various methods in Table 3, it can be concluded that the one-stage methods have higher efficiency but lower detection accuracy than two-stage methods, since the one-stage methods have simpler network structures than the latter. SSD method enhances the calculation efficiency, but the corresponding detection accuracy is decreased. CSPResNeXt50 and CSPDarknet53 based YOLO V4 methods have the fastest detection speed, making them suitable for high speed on-line detection, but the accuracy is lower than proposed method. By using cascading structures, Cascade R-CNN improves the detection accuracy compared with the traditional Faster R-CNN method, but the cascade of multiple subnetworks also increases the

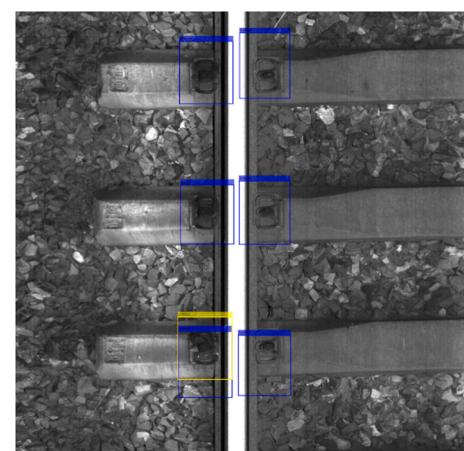


Fig. 11. Detection result of deviated fastener by Faster R-CNN.

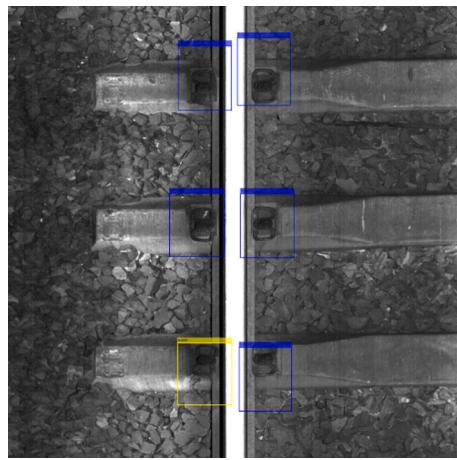


Fig. 12. Detection result of deviated fastener by proposed method.

complexity of the network, which affects the detection speed. Compared with the commonly used baseline methods, this proposed method has higher detection rate and accuracy. Meanwhile, the processing time for single image of the presented method is short and the detection efficiency is high. By optimizing RPN, the proposed method in this paper can improve the detection accuracy and enhance the computational efficiency simultaneously. It makes up the shortcomings of the traditional Faster R-CNN algorithm in rapid detection and improves the efficiency and accuracy in the meantime, which is suitable for fast and accurate detection of fastener state.

In order to further verify the role of SVDD in the two-stage classification, the commonly used methods of SVM, Artificial Neural Networks (ANNS), K-Nearest Neighbor (KNN) and CNN are applied by replacing SVDD, to accomplish the second stage of classification, where normal and deviated fasteners are classified. The first three methods are processed in accordance with the one in Section 3.3. CNN method directly uses the images of normal fasteners and deviated fasteners obtained from the first stage for training and testing. The accuracy results of

deviated fasteners classification are shown in Table 4.

It can be seen from Table 4 that the traditional classification methods are not effective, because the samples of normal fasteners are sufficient and more consistent while the samples of deviated fasteners are insufficient and less consistent since the deviation degrees vary for deviated fasteners. The CNN method avoids the information loss caused by feature extraction and improves the accuracy, but it is still affected by the imbalance of sample size and distribution. The proposed method is free of the above problems and obtain the best classification performance, since SVDD can be trained and classified using only normal samples.

## 5. Conclusions

Fast, accurate, and intelligent detection of fasteners is of great significance to ensure the safe operation of railway vehicles. According to the characteristic of railway fasteners, a new two-stage classification model based on a deep learning network is constructed in this paper, to recognize the fastener state. Through the verification and analysis of practical detection images, the following conclusions are drawn:

(1) The traditional Faster R-CNN method is modified according to the characteristics of object detection for railway fastener. The candidate box is generated by the parameter optimized RPN based on the fastener object image, which can reduce the computation load, speed up the calculation and enhance precision of target positioning.

(2) A two-stage classification model based on the Faster R-CNN is

**Table 4**  
The classification accuracy comparison of different methods in second stage.

Methods	Fastener Type I	Fastener Type II	Fastener Type III	Fastener Type IV	Total
SVM	82.41	85.59	81.7	77.19	82.48
ANN	86.32	82.51	83.14	79.73	83.42
KNN	87.65	83.43	84.69	76.08	83.82
CNN	86.37	87.59	86.75	82.85	86.25
Proposed method	90.04	90.82	89.97	86.75	89.90

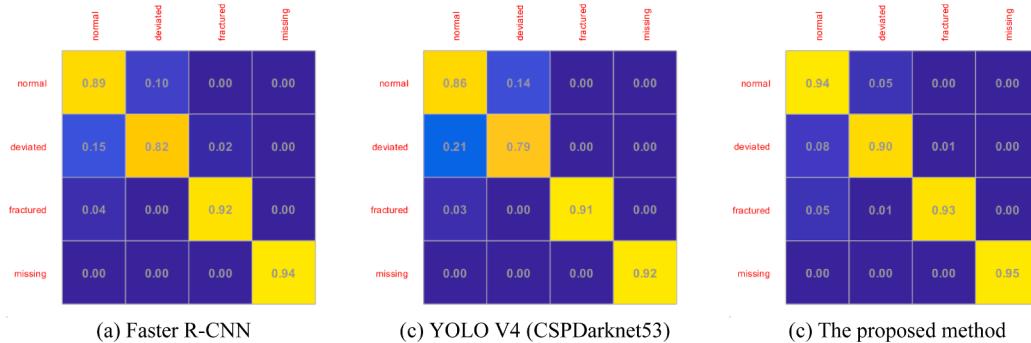


Fig. 13. Comparison of fastener classification results.

**Table 3**

The results comparison of fastener detection.

Methods	Fastener Type I			Fastener Type II			Fastener Type III			Fastener Type IV			$R_{do}$	$R_{ao}$
	$R_d$	$R_a$	$FPS$	$R_d$	$R_a$	$FPS$	$R_d$	$R_a$	$FPS$	$R_d$	$R_a$	$FPS$		
SSD	82.99	85.13	17.54	83.32	85.53	17.69	82.32	84.60	17.22	79.66	81.66	16.78	82.61	84.78
YOLO V4 CSPResNeXt50	84.50	85.65	38.91	84.30	85.78	39.24	83.83	85.18	38.67	80.83	82.32	34.35	83.88	85.22
YOLO V4 CSPDarknet53	85.77	86.93	40.43	85.72	86.67	41.20	85.11	86.94	40.09	82.35	84.83	36.54	85.17	86.54
Faster R-CNN	88.65	89.36	12.34	88.98	89.76	12.53	88.32	89.09	12.10	85.18	87.03	10.82	88.30	89.19
Mask R-CNN	90.33	91.20	7.58	90.86	91.93	8.17	89.93	90.73	7.16	87.93	88.87	6.69	90.18	91.12
Cascade R-CNN	90.73	90.87	8.23	91.13	91.40	8.41	90.53	90.54	8.35	88.73	88.14	7.02	90.61	90.69
Proposed method	92.92	93.07	16.95	93.25	93.47	17.26	92.59	92.94	16.43	89.92	90.47	16.15	92.63	92.88

built, according to abnormal information of defected fasteners. Firstly, the Faster R-CNN is used to classify the fasteners as complete, fractured and missing ones. Then the SVDD classification model is applied to image feature information of complete fasteners to recognize normal and deviated ones. The method avoids the information interference of deviated angles that might be encountered by using a direct classification method, and improves the detection accuracy of deviated fasteners.

(3) The detection test is carried out on railway fasteners. With the collected images, the training and test data sets are established for detection. The results are analyzed and it can be concluded that the proposed method has higher detection accuracy and efficiency compared with traditional detection methods.

In addition to the above conclusions, with the rapid development of detection algorithms, the idea of network structure optimization based on specific detection objects proposed in this paper can be extended to different backbone networks, or other excellent detection methods, such as YOLO V4, Mask R-CNN, Cascade R-CNN, etc., to test whether better performance can be obtained.

#### CRediT authorship contribution statement

**Tangbo Bai:** Conceptualization, Methodology, Software, Validation, Writing - original draft. **Jianwei Yang:** Validation, Investigation. **Guiyang Xu:** Resources, Data curation. **Dechen Yao:** Software.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This research acknowledges the financial support provided by General Project of Scientific Research Program of Beijing Municipal Education Commission (KM202010016003), Beijing Postdoctoral Research Foundation (2019-98), National Natural Science Foundation of China (51975038).

#### References

- [1] Z. Peng, C. Wang, Z. Ma, H. Liu, A multifeature hierarchical locating algorithm for hexagon nut of railway fasteners, *IEEE Trans. Instrum. Meas.* 69 (3) (2020) 693–699.
- [2] J. Liu, Y. Huang, Q. Zou, et al., Learning visual similarity for inspecting defective railway fastener, *IEEE Sens. J.* 19 (16) (2019) 6844–6857.
- [3] C. Chellaswamy, M. Krishnasamy, L. Balaji, A. Dhanalakshmi, R. Ramesh, Optimized railway track health monitoring system based on dynamic differential evolution algorithm, *Measurement* 152 (2020) 107332, <https://doi.org/10.1016/j.measurement.2019.107332>.
- [4] Z. Zhan, H. Sun, X. Yu, J. Yu, Y. Zhao, X. Sha, Y.e. Chen, Q. Huang, W.J. Li, Wireless rail fastener looseness detection based on MEMS accelerometer and vibration entropy, *IEEE Sens. J.* 20 (6) (2020) 3226–3234.
- [5] D. Milne, A. Masoudi, E. Ferro, G. Watson, L. Le Pen, An analysis of railway track behaviour based on distributed optical fibre acoustic sensing, *Mech. Syst. Sig. Process.* 142 (2020) 106769, <https://doi.org/10.1016/j.ymssp.2020.106769>.
- [6] W.-F. Zhu, X.-J. Chen, Z.-W. Li, X.-Z. Meng, G.-P. Fan, W. Shao, H.-Y. Zhang, A SAFT method for the detection of void defect inside a ballastless track structure using ultrasonic array sensors, *Sensors* 19 (21) (2019) 4677, <https://doi.org/10.3390/s19214677>.
- [7] M. Guerrieri, G. Parla, C. Celauro, Digital image analysis technique for measuring railway track defects and ballast gradation, *Measurement* 113 (2018) 137–147.
- [8] X. Wei, D. Wei, D.a. Suo, L. Jia, Y. Li, Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model, *IEEE Access* 8 (2020) 61973–61988.
- [9] H. Fan, P.C. Cosman, Y. Hou, B. Li, High-speed railway fastener detection based on a line local binary pattern, *IEEE Signal Process Lett.* 25 (6) (2018) 788–792.
- [10] H.F. Ma, Y.Z. Min, C. Yin, et al., A real time detection method of track fasteners missing of railway based on machine vision, *International Journal of Performability Engineering* 14 (6) (2018) 1190–1200.
- [11] Y. Ou, J. Luo, B. Li, B. He, A classification model of railway fasteners based on computer vision, *Neural Comput. Appl.* 31 (12) (2019) 9307–9319.
- [12] X. Xu, Y. Lei, F. Yang, Railway subgrade defect automatic recognition method based on improved Faster R-CNN, *Sci. Program.* 2018 (2018) 1–12.
- [13] S. Ma, L. Gao, X. Liu, J. Lin, Deep learning for track quality evaluation of high-speed railway based on vehicle-body vibration prediction, *IEEE Access* 7 (2019) 185099–185107.
- [14] H. Cui, J. Li, Q. Hu, Q. Mao, Real-time inspection system for ballast railway fasteners based on point cloud deep learning, *IEEE Access* 8 (2020) 61604–61614.
- [15] X. Gibert, V.M. Patel, R. Chellappa, Deep multitask learning for railway track inspection, *IEEE Trans. Intell. Transp. Syst.* 18 (1) (2017) 153–164.
- [16] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, Y. Li, Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study, *Eng. Appl. Artif. Intell.* 80 (2019) 66–81.
- [17] R. Girshick, J. Donahue, T. Darrell, et al., Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA (2014) 580–587.
- [18] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders, Selective search for object recognition, *Int. J. Comput. Vision* 104 (2) (2013) 154–171.
- [19] C.L. Zitnick, P. Dollár, Edge boxes: Locating object proposals from edges, *European Conference on Computer Vision*, Springer, Zurich, 2014, pp. 391–405.
- [20] R. Girshick, Fast R-CNN, *Proceedings of IEEE International Conference on Computer Vision*, Santiago, Chile (2015) 1440–1448.
- [21] S. Ren, K.M. He, R. Girshick, R.-C.-N.-N. Faster, et al., in: *Towards real-time object detection with region proposal networks*, MIT Press, 2015, pp. 91–99.
- [22] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, *IEEE International Conference on Computer Vision*, Honolulu, HI, USA (2017) 2980–2988.
- [23] Z. Cai and N. Vasconcelos, Cascade R-CNN: Delving into high quality object detection. *IEEE Conference on Computer Vision and Pattern Recognition*, UT, USA (2018) 6154–6162.
- [24] J. Zhang, Z. Xie, J. Sun, X. Zou, J. Wang, A Cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection, *IEEE Access* 8 (2020) 29742–29754.
- [25] W. Liu, A. Dragomir, E. Dumitru, et al., SSD: Single shot multibox detector, *European Conference on Computer Vision* (2016) 21–37.
- [26] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE (2016) 779–788.
- [27] W. Fang, L. Wang, P. Ren, Tinier-YOLO: A real-time object detection method for constrained environments, *IEEE Access* 8 (2020) 1935–1944.
- [28] M.X. Tan, R.M. Pang, Q. V. Le, Efficientdet: Scalable and efficient object detection, *arXiv* (2019) 1911.09070.
- [29] A. Bochkovskiy, C.Y. Wang, H.Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *arXiv* (2020) 2004.10934.
- [30] D. Tax, A. Ypma, R. Duin, Support vector data description applied to machine vibration analysis, *Proceedings of the Fifth Annual Conference of the Advanced School for Computing and Imaging* 54 (1999) (1999) 15–23.
- [31] Z. Wang, Z. Zhao, S. Weng, C. Zhang, Solving one-class problem with outlier examples by SVM, *Neurocomputing* 149 (2015) 100–105.
- [32] M. Cha, J.S. Kim, J.-G. Baek, Density weighted support vector data description, *Expert Syst. Appl.* 41 (7) (2014) 3343–3350.
- [33] X.W. Yao, J.W. Han, D.W. Zhang, F.P. Nie, Revisiting Co-saliency detection: A novel approach based on two-stage multi-view spectral rotation Co-clustering, *IEEE Trans. Image Process.* 26 (2017) 3196–3209.
- [34] W. Jiang, L. Wu, S. Liu, M. Liu, CNN-based two-stage cell segmentation improves plant cell tracking, *Pattern Recogn. Lett.* 128 (2019) 311–317.
- [35] W.B. Jiang, M. Liu, Y.N. Peng, L.H. Wu, Y.N. Wang, HDCB-Net: A neural network with the Hybrid Dilated Convolution for pixel-level crack detection on concrete bridges, *IEEE Transactions on Industrial Informatics*, 10.1109/TII.2020.3033170.
- [36] Z. Ge, Z. Jie, X. Huang, C. Li, O. Yoshie, Delving deep into the imbalance of positive proposals in two-stage object detection, *Neurocomputing* 425 (2021) 107–116.
- [37] D.M.J. Tax, R.P.W. Duin, Support vector data description, *Machine Learning* 54 (1) (2004) 45–66.
- [38] T. Bai, L. Zhang, L. Duan, J. Wang, NSCT based Infrared Image Enhancement Method for Rotating Machinery Fault Diagnosis, *IEEE Trans. Instrum. Meas.* 65 (10) (2016) 2293–2301.
- [39] L. Duan, M. Yao, J. Wang, T. Bai, L. Zhang, Segmented infrared image analysis for rotating machinery fault diagnosis, *Infrared Phys. Technol.* 77 (2016) 267–276.