

## Research Article

# Semantic-Segmentation-Based Rail Fastener State Recognition Algorithm

**Liming Li** <sup>1,2</sup> **Rui Sun**,<sup>2</sup> **Shuguang Zhao** <sup>1</sup> **Xiaodong Chai**,<sup>2</sup> **Shubin Zheng**,<sup>2</sup>  
and **Ruichao Shen**<sup>2</sup>

<sup>1</sup>*School of Information Science and Technology, Donghua University, Shanghai 201620, China*

<sup>2</sup>*School of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai 201620, China*

Correspondence should be addressed to Shuguang Zhao; [sgzhao@dhu.edu.cn](mailto:sgzhao@dhu.edu.cn)

Received 13 June 2020; Revised 10 September 2020; Accepted 13 February 2021; Published 3 March 2021

Academic Editor: Zheng-zheng Wang

Copyright © 2021 Liming Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rail fastener status recognition and detection are key steps in the inspection of the rail area status and function of real engineering projects. With the development of and widespread interest in image processing techniques and deep learning theory, detection methods that combine the two have yielded promising results in practical detection applications. In this paper, a semantic-segmentation-based algorithm for the state recognition of rail fasteners is proposed. On the one hand, we propose a functional area location and annotation method based on a salient detection model and construct a novel slab-fastclip-type rail fastener dataset. On the other hand, we propose a semantic-segmentation-framework-based model for rail fastener detection, where we detect and classify rail fastener states by combining the pyramid scene analysis network (PSPNet) and vector geometry measurements. Experimental results prove the validity and superiority of the proposed method, which can be introduced into practical engineering projects.

## 1. Introduction

As shown in Figure 1, a rail fastener is a fixed coupling part that prevents horizontal and vertical offsets in rails. Thus, rail fastener detection can be used for maintaining the stability of railway systems and ensuring the safety of trains. Figure 1 shows that the slab fastclip (SFC) type of rail fastener is used for coupling steel rails and sleepers in ballast and ballastless rail scenarios, respectively. Traditional rail fasteners detection requires workers to walk along the railroad to determine the state of the rail fasteners and other functional components; this method has low detection efficiency and precision and is dangerous [1]. Therefore, automatic rail fastener detection has attracted increasing attention from researchers.

To address the limitation of traditional detection methods, many automatic detection methods using machine vision and image processing technology have been proposed and achieved good experimental results [1–11]. Recently, deep learning theory has received increasing attention in

target detection and image segmentation works and has been successfully applied to rail fastener detection [12–14]. However, these efforts are often heavily dependent on time-consuming and expensive manual annotations.

Thus, this paper proposes a semantic-segmentation-based rail fastener state recognition algorithm. Our research was conducted in three main aspects. First is the construction of a functional area marker model for rail fasteners based on significance detection. In this model, the fastclip parts in the rail fastener image are regionally localized by the significance detection model. Then, the different functional parts in the image are semiautomatically labeled by constructing pseudolabels. Finally, a SFC-type rail fastener dataset is constructed based on the labeling results and the true state information. This method can effectively avoid the tedious manual collection and labeling of fastener samples and can compete in fastener detection. Second, we creatively propose a semantic-segmentation-based method for detecting rail fasteners. The functional regions in the input rail fastener image are segmented by a semantic

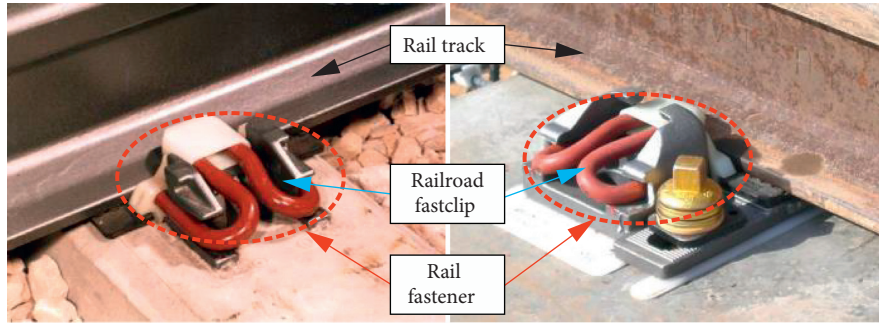


FIGURE 1: SFC-type rail fastener.

segmentation network model, and a fastener state detection method based on vector geometry relationships was designed based on the segmentation results. Third, the overall model of the proposed system is an end-to-end detection model that can detect and complete the classification of the rail fastener status in the raw input image, which can have excellent advantages in actual engineering projects.

The rest of this paper is organized as follows. Section 2 introduces related works on current rail fastener detection. Section 3 describes the overall framework and methods. Section 4 discusses the experimental results and analysis. Finally, the conclusion and future work are presented in Section 5.

## 2. Related Work

With the development of computer vision and image processing technologies, researchers have been committed to the research of rail fastener detection using two-dimensional visual images. In most of the studies, the primary purpose of the rail fastener inspection task is to check for missing fasteners on both sides of a rail. In [1], the top-down detection method is proposed to detect the fastener region and predict its status. In [2], an automatic and configurable real-time vision system is proposed to detect the presence/absence of rail fasteners. In [3], a fastener location and detection method based on the combination of wavelet transform and template matching is proposed; the method can accurately locate fasteners and predict their status. In [4], the authors present a method based on image processing and pattern recognition techniques, which can be customized to detect the absence of fasteners. In [5], wavelet transformation and principal component analysis are combined to detect fasteners. In [6], a method based on line local binary mode coding is proposed; the authors comprehensively considered the correlation between the image center point and its neighborhood nodes. In [7], the from-coarse-to-fine strategy is proposed to detect and recognize broken rail fasteners with a method based on Haar-like features and the Adaboost algorithm. In [8], the probabilistic structure topic model is proposed for simultaneously learning the probabilistic representations of different objects using unlabeled samples. In [9], a fastener detection method

based on the combination of the Shi-Tomasi and Harris-Stephen feature detection algorithms is proposed; the method can successfully detect the presence of fasteners. In [10], an autonomous visual rail fastener inspection system is proposed; in the system, the histogram of oriented gradient features and the linear support vector machine (SVM) classifiers' method are utilized to inspect the defect situation and classify fasteners.

Several laser detection methods have also been proposed. For example, in [15], a real-time rail fastener detection system is proposed using laser ranging, which can effectively reduce the calculation cost. In [16], a fastener detection method based on the light sensor mechanism is proposed; the method uses a decision tree classifier and the centerline extraction method to detect the incomplete state and loose fasteners. In [17], a structured light method based on motion image for the moving object inspection method is proposed, offering a fresh perspective when inspecting missing fastening components on high-speed railways. In [18], the authors proposed a structured-light-based system to evaluate the rail gauge and detect missing rail fasteners.

A rail fastener image, which contains some functional parts, can be further divided into the rail, fastener, and background regions. In [19], unlike in traditional rail fastener detection methods, attention is also given to the location and detection of the hexagonal nut in a rail fastener image.

In recent years, with the increasing application of deep learning technologies, some researchers have also applied deep learning to rail fastener detection. For example, in [12], the authors proposed a template matching classification method to automatically collect and annotate fastener samples and further deployed a similarity-based deep convolutional neural network (DCNN) to estimate the fastener state. In [13], a real-time inspection system for ballast railway fasteners based on point cloud deep learning was developed, demonstrating excellent accuracy and efficiency in field testing on ballastless tracks. In [14], a fastener detection method based on visual rail inspection is proposed using material classification and semantic segmentation with DCNN to, respectively, identify and segment the different functional parts in a rail fastener image. In [20], the authors proposed Yolo v3, which is deployed and trained as a deep learning model for detecting the state of rail fasteners. In

[21], an end-to-end abnormal fastener detection method, which can identify abnormal fasteners from a rail scene image, is proposed.

In summary, we consider the following major problems of previous approaches:

- (1) Although good results can be achieved in rail fastener detection through deep learning frameworks, the detection methods based on existing supervised learning are heavily dependent on the manual pixel-level annotation of image data. Annotating large-scale rail fastener image datasets one by one through manual methods is extremely tedious, time-consuming, and expensive.
- (2) Existing rail fastener detection methods are mainly focused on the missing state of rail fasteners in images. The main body of the detection target generally determines the missing state only for the overall area of the fastener, rather than performing specific state detection based on local functional areas. The positioning results and status of rail fasteners can only be considered from a qualitative analysis perspective, and no unified quantitative evaluation criteria exist for the accurate description and comparison of the effects of the detection method.
- (3) Rail fastener detection is a standard data sample imbalance problem in which the positive sample data images of fasteners taken and collected in actual railway scenes are generally much larger than the negative samples. Unbalanced training and experimental samples can affect the accuracy of experimental results.

To solve the problems of existing approaches, we present a novel semantic-segmentation-based rail fastener state recognition algorithm. The contributions of our work are as follows:

- (1) To reduce the reliance of traditional deep learning methods on manual annotation, we provide a semiautomatic method for locating and annotating rail fasteners based on saliency detection. The experimental results show that the method can accurately locate and segment fastener pop-up regions and generate accurate pixel-level annotations of the rail fastener image, reducing the cost of the manual annotation of functional regions and improving the efficiency by 25 times that of the traditional manual annotation process.
- (2) We further classify the fastener state into five specific situations based on a priori knowledge of the geometric relationships between the different functional regions in the fastener image. Meanwhile, we shift the attention of detection to specific functional structural regions (i.e., the fastener fastclip and rail regions). We propose a semantic-segmentation-based rail fastener detection method and introduce new quantitative

evaluation metrics to describe and evaluate the results of fastener positioning experiments. The experimental results prove that the fastener cartridge positioning effect and state detection results of this method have obvious accuracy and superiority.

- (3) To solve the problem of data imbalance, a new standard dataset based on the SFC-type fastener is constructed by the method in Contribution 1, and the image's negative sample data is reasonably augmented in the construction process. To some extent, the problem of dataset imbalance is alleviated, and the overall performance of the detection method framework is improved.

### 3. Proposed Method

First, we perform localization and semiautomatic labeling of the functional regions in the original images using a salient detection model (SDM). Then, we propose a semantic-segmentation-based rail fastener state detection method, which is based on the semantic information and spatial relationships among the functional parts in the rail fastener image, to achieve the accurate detection and monitoring of the rail fastener state.

#### 3.1. Positioning and Marking of Functional Areas.

Figure 2 describes in detail a specific implementation of the method for locating and labeling rail fasteners based on significance detection, which enables the interactive automatic labeling of the functional areas in the input image. Our overall approach consists of three main modules. First, the fastener region in the input image is positioned by the salient detection model to obtain the original rail saliency map corresponding to each original rail image. Then, the target region segmentation model is used to locate and segment the fastener regions on both sides of the rails in the image to obtain the corresponding position of the rail fastener image, which mainly contains the rails and fastener fastclip regions that we are interested in. Finally, a salient detection model is used to further locate and correct the fastener fastclip areas in the saliency map, and a pseudolabel construction model is used to semiautomatically label the functional areas in the image to generate the corresponding interactive pseudolabels for each rail fastener image. The following section discusses these three main modules in detail. In addition, we constructed a special dataset of SFC rail fasteners using the above method. This dataset includes the rail fastener images obtained by salient detection and target area segmentation, the corresponding functional area pseudolabeling, the ground truth, and the labeled real fastener states.

**3.1.1. Salient Detection Model (SDM).** We first perform the salient detection of the original rail fastener image by setting up a sliding processing window based on the comparison of regional features. As shown in Figure 3, sliding image processing window  $W$  is defined with an adjustable window

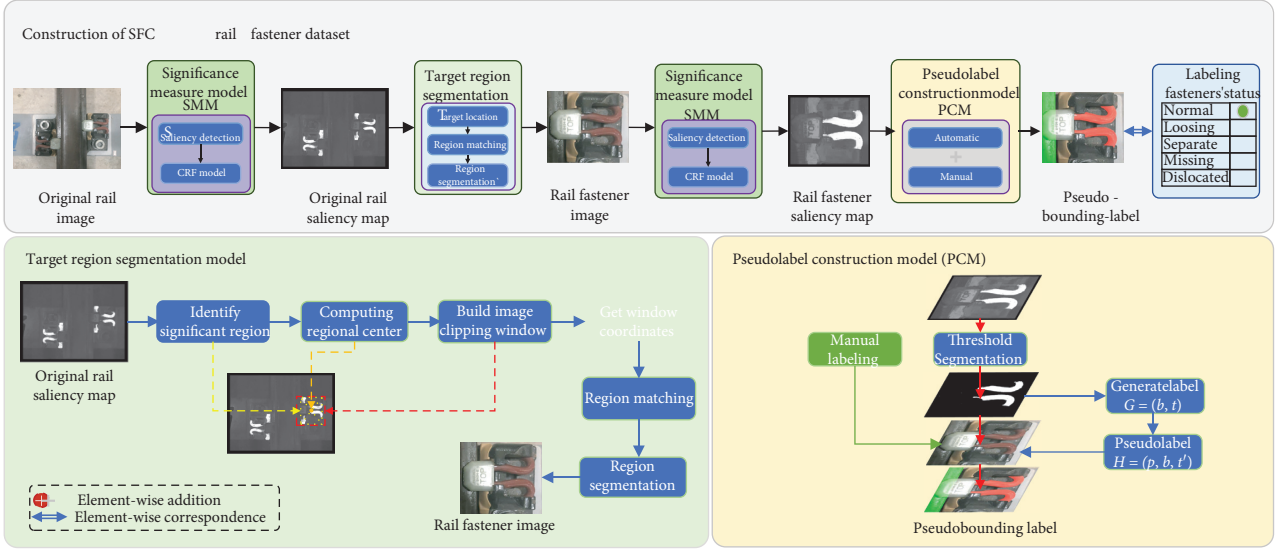


FIGURE 2: Overall framework of the rail functional area positioning and labeling method: (a) details of the target region segmentation model and (b) details of PCM.

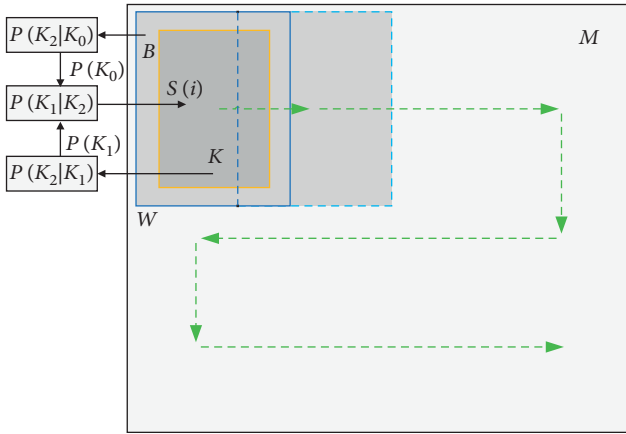


FIGURE 3: Detailed diagram of the sliding processing window in SDM.

scale in the input image, and image processing window  $W$  is divided into processing kernel  $K$  and outer frame  $B$ . Random variable  $i \in \mathbf{R}^2$  represents a pixel node in image processing window  $W$ , and a specific characteristic attribute of the node is defined as  $f(i)$ . The characteristic properties of the described pixel node are measured by computing the intensity and color features of the image to measure the significance of the node. This salient detection mechanism is similarly defined in the literature [22].

For pixel node  $i$  in the sliding image processing frame, window  $W$ , we assume that events  $H_0$  and  $H_1$  indicate that pixel node  $i$  is located in processing kernel  $K$  and outer frame  $B$ , and event  $H_2$  indicates that pixel node  $i$  is a significant pixel node. The significant measure ( $S(i)$ ) of pixel node  $i$  can be expressed by the following Bayesian formula:

$$S(i) = P(H_0|H_2) = \frac{P(H_2|H_0)P(H_0)}{P(H_2|H_0)P(H_0) + P(H_2|H_1)P(H_1)}. \quad (1)$$

In addition, to further calculate the significance metric value ( $S(i)$ ) for each pixel node  $i$  in sliding image processing window  $W$ , we assume that the probability of the existence of the above event ( $H_0$ ) is  $p$ , which ranges from (0,1). To reduce the error interference during the calculation of the saliency measure, we introduce a normalized regular histogram to enhance the robustness and stability of the algorithm. Second, we represent  $P(K_2|K_0)$  and  $P(K_2|K_1)$  in equation (1) by feature histograms  $h_K(i)$  and  $h_B(i)$  and use  $\mathbf{h}_F(i)$  and  $\mathbf{h}_B(i)$  to represent the products of the pixels in the corresponding histograms in the CIELAB color space. Finally, the significant metric value of pixel node  $i$  that meets the requirements of the specific feature attributes in sliding image processing frame window  $W$  is

$$S(i) = \frac{\mathbf{h}_K(i)p}{\mathbf{h}_K(i)p + \mathbf{h}_B(i)(1-p)}. \quad (2)$$

We construct the feature histograms of any window in different position cases by moving different proportions of the sliding image processing frame window in the original image of the rail fastener and measuring the significance of the pixel nodes within the sliding image processing frame window according to equation (2).

On this basis, we label and segment the pixel nodes by the minimization energy function in the conditional random field (CRF) [23] to perform the target-level foreground segmentation of the significant regions in the rail fastener image. Suppose that input image  $M$  contains  $n$  pixel nodes  $i$ ; we first define array  $X = (x_1, x_2, \dots, x_n)$  to represent the set



of all pixel nodes  $i$  in image  $M$ . Then, let array  $Y = (y_1, y_2, \dots, y_n)$  represent the significant metric value of pixel node  $i$ . Finally, we assume that a binary segmentation label that can label pixel node  $i$  exists, where  $y_i \in \{F, B\}$  indicates that pixel node  $i$  is a salient and non-salient label. The mathematical description of the CRF model we constructed based on salient detection is

$$p(Y|X, s) = \frac{1}{Z(X)} \exp\left(-\sum_n E(y, x, s)\right), \quad (3)$$

where  $Z(X)$  is the normalization factor. We estimate the label attribute ( $y_i$ ) of a pixel node in an image by using the energy function in the minimized CRF model, and the mathematical representation is

$$\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n) = \arg \min E(y, x, s). \quad (4)$$

Energy function  $E$  in the CRF model consists of three feature models: two one-dimensional  $U^S$  terms based on saliency feature model  $U^S$ , color feature model  $U^C$  for a particular pixel node, and a two-dimensional term based on spatial relationship model  $Q$  for neighboring pixels. The mathematical expression is

$$E(y, x, s) = \sum_{i=1}^n (w_s U^S(y_i, s_i) + w_c U^C(y_i, x_i)) + \sum_{(h,j) \in X} Q(y_h, y_j, x_h, x_j), \quad (5)$$

where parameters  $w_s$  and  $w_c$  are the weighting factors that control the corresponding feature models. Their specific values are determined by the method described in the literature [24].  $x_h$  and  $x_j$  are adjacent pixel nodes in rail fastener image  $M$ .

For the original rail images in Figure 2, we first optimally estimate the labeling of pixel points by the energy minimization function of the CRF model and label the pixel nodes as significant and nonsignificant feature nodes. The image is then segmented into significant foreground and nonsignificant background areas based on the significant feature contrast of adjacent pixel nodes. In this approach, the fastclip areas on the rail fastener image can be marked and highlighted as prominent foreground areas, and the original rail salient maps can be generated.

**3.1.2. Target Region Segmentation Model.** Considering that the original rail image input contains some irrelevant regions and backgrounds in addition to our fastener regions of interest, we propose a target region segmentation module. In this module, we first divide the localization results of saliency detection into left and right subimages and then segment and extract the corresponding target regions in the subimages.

As shown in Figure 2a, the coordinates of the center pixel of the significant region are first calculated and discriminated based on the spatial geometric prior information of the saliency positioning module results. Subsequently, a specific size image clipping frame is constructed with the central

pixel as the coordinate origin. Then, the boundary coordinates of the clipping frame are mapped and matched with the original rail image by obtaining the clipping frame. Finally, by segmenting the matching region in the original image, we finally obtain our desired rail fastener image.

We first define array  $T = \{N_{1,1}, N_{1,2}, \dots, N_{n,m}\}$  to represent the set of all pixel nodes  $N_{i,j}$  in the original rail saliency map ( $M_s$ ). Then, we perform the binary segmentation of saliency map  $M_s$  and define function  $F_s(N_{i,j})$  to represent the saliency eigenvalue of pixel node  $N_{i,j}$ . The value of the function is 1 when the pixel node is the significant foreground and 0 for the background node. The coordinates of pixel node  $N_{i,j}$  are represented by  $(x_i, y_j)$ . We can calculate the value of the coordinates of the center of the significant region ( $C(X, Y)$ ) as follows:

$$\begin{cases} X = \bar{x} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N x_i F_s(x_i, y_i), \\ Y = \bar{y} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N y_i F_s(x_i, y_i), \end{cases} \quad (6)$$

where the value of function  $F_s(x_i, y_i)$  is 1 and  $M$  and  $N$  are the respective lengths and widths of the significant regions derived by statistical operations.

Based on the coordinates ( $C(X, Y)$ ) of the centroids of the significant regions obtained from the above equation, we can reconstruct an image crop frame ( $Z = (C, V)$ ) containing all significant regions based on our a priori knowledge, where  $C$  is the coordinates of the centroids of the regions and  $V$  represents the size of the reconstructed crop frame. To facilitate the subsequent image computation and processing, the range of the  $V$  values is positioned here as [470, 520] pixels. Through this method of randomly generating crop frames of variable size in the parameter interval, the negative sample data of the input image can be randomly augmented. Consequently, the data samples can be maintained at a relatively balanced level as much as possible. To a certain extent, the imbalance of the dataset of rail fasteners is alleviated and solved, improving the overall performance and accuracy of the algorithm.

Finally, by obtaining the boundary coordinates of the image clipping frame for coordinate mapping and matching with the original rail image, we can obtain an image that contains only the rail fasteners and rail edges of ROI (region of interest), completing the task of segmenting the foreground target area.

**3.1.3. Pseudolabel Construction Module (PCM).** Given the irregular shape characteristics of the fastener bar area in the rail fastener image, the bar area in the image must be manually marked at a high cost when using the traditional supervised learning method to detect a specific bar area in the rail fastener image, whereas the shape characteristics of the rail travel area in the image are usually significant and a regular rectangular area, which can be marked using the manual marking method. Therefore, this work attempts to construct image target-level pseudolabels instead of pixel-level manual annotations for the

rail fasteners' bullet regions and combine the results of manual annotations for the rail walk regions to learn and train the labels of different rail functional regions in the image through weak supervised learning.

Our model for constructing pseudolabels for the automatic labeling of fastener fastclip regions is inspired by the literature [25]. Ultimately, accurate target classification and semantic segmentation can be implemented for the different classes of track-functional fasteners in the image.

As shown in Figure 2b, we first construct an image target-level pseudolabel for the rail functional area in the rail fastener image positioned by the saliency area. First, let  $M'_s \in \mathbb{R}^{w \times h \times 3}$  represent the segmentation result of the significant map, where  $w$  and  $h$  are the height and width of the input image, respectively. Suppose that  $T$  represents the set of different functional area categories in rail fastener image  $M_O$ ; then, the significant map of the rail fasteners generated by the SDM is transformed into a binarized image by threshold segmentation. Assuming the existence of binarized label  $t$  to label the fastclip and background regions, the pixel value of the significant region in the segmentation result is set to 1 and the background pixel to 0. Then, image-level label  $G$  is constructed for the pixels in the fastener fastclip region, which is mathematically expressed as

$$G = (b, t), \quad t \in T, \quad (7)$$

where  $b$  is the boundary segmentation box for a significant foreground region. When the value of  $t$  is 1,  $G$  denotes the set of pixels in the bounding box where the fastener striping region of the figure is calibrated.

Second, we assume that marker vector  $t'$  has a similar definition as  $t$ : 0 and 1 denote the pixel properties of the background and the pop-up bar region of the graph. In addition, the rail region pixel property is set to 2. Then, the mathematical expression for constructing pseudolabel  $H$  is

$$H = \begin{cases} (p, b', t') \leftarrow G(b, t) & t' = t = 1, \\ (b', t'), t' = 2, \end{cases} \quad (8)$$

where parameter  $p$  represents the probability that the bounding box belongs to the fastener fastclip area. According to equation (8), we perform a pixel-level annotation of the rail regions in the rail fastener image manually ( $t' = 2$ ) and jointly construct the pseudolabeling ( $H$ ) of the image functional regions with the fastener bullet image-level labels ( $G$ ) obtained by the above automatic annotation method ( $t' = 1$ ).

This semiautomatic image labeling method is used to generate the proposed pseudolabel in this work. The pseudolabeling structure consists of the target-level labeling of the fastclip region and the pixel-set labeling of the rail region. Consequently, we can finally obtain complete semantic labels for the different functional regions of the rail fastener image. Furthermore, as shown in Figure 2, we construct a new rail fastener image dataset, which includes the cropped rail fastener image, the generated pseudolabels, and the real state of the fastener.

**3.2. Semantic Segmentation and Defect Detection.** In this section, we first describe in detail the overall architecture of a vision-based rail fastener defect detection system. The system detects and outputs the state of the rail fastener in the raw input rail image via an end-to-end detection model. Second, we elaborate and discuss the semantic segmentation-based fastener defect detection algorithm within the system.

**3.2.1. Overall System Architecture.** As shown in Figure 4, the objective of this study is to photograph and collect original rail images in real engineering applications. Our proposed detection system is an end-to-end rail fastener detection system. The acquired original input rail image is first localized regionally, and the target region is segmented based on the SDM to obtain rail fastener images that contain only useful functional region features. Then, a semantic segmentation-based defect detection method for rail fasteners, which utilizes a semantic segmentation model and a state detection method based on vector geometry measurements, is proposed to detect and judge the state of rail fasteners. The results of the rail fastener defect detection and the status classification are finalized and generated.

**3.2.2. Status Detection Method.** Figure 5 presents the overall architecture of our semantic-segmentation-based approach to the rail fastener defect detection proposed in the detection method. In our designed detection method, we first perform an accurate semantic segmentation of the rails and fastener fastclip in the input rail fastener image using a semantic segmentation model. A fastener defect detection algorithm based on vector geometry parameters is then designed based on the segmentation results of the different functional regions to realize the detection and classification of fastener defects and states.

*Functional region segmentation* is an important part of this method. In this part, the input rail fastener image of the rail, the fastener bar, and the background region of our interest are segmented using the semantic segmentation model to obtain the semantic segmentation result of the corresponding region. The semantic segmentation model involves the accurate prediction and segmentation of the different functional regions in the rail fastener image by detecting the semantic segmentation network involved in the system.

For the rail fastener images taken and collected in the actual engineering scene, this paper uses PSPNet [26] to learn the context information and key details between the pseudolabels constructed for the different region categories in the global scene of the input image to finally achieve the semantic segmentation of different functional regions in the image. PSPNet can fuse semantic and detailed information features in different network layers. The diverse positions, shapes, and sizes of different functional areas in the rail fastener image can be used in rail fastener detection.

We use the pseudolabeled rail fastener images of different functional regions as training data, extract the underlying features of the input images, and generate the

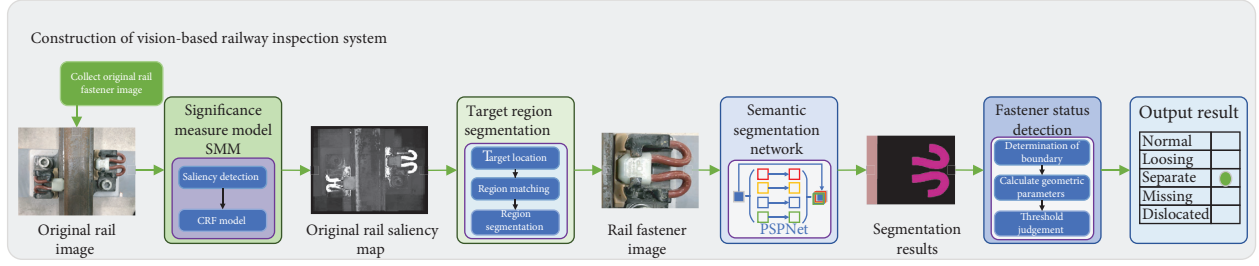


FIGURE 4: Overall architecture of our detection system.

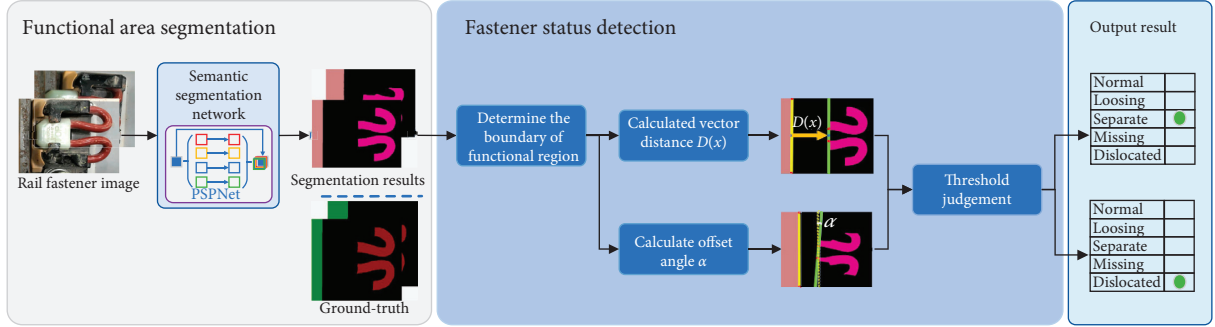


FIGURE 5: Overall architecture of our detection method.

corresponding feature maps using the pretrained deep residual network (ResNet) [27] and the expansion convolution strategies [28, 29] in the PSPNet split network architecture used in this study. ResNet can achieve good network classification and identification by deepening the training network depth; the expansion convolution strategy can expand the size of the receiving field to a certain extent without changing the feature layer scale to obtain more global image information by expansion convolution on the basis of completing the original network structure. Second, we use the pyramid pooling module to learn and acquire contextual information about the different regional category labels in the image based on the full use of the captured global image features. The pyramid pool module in the PSPNet network architecture integrates four pyramid sub-models of different scales to sample and output feature elements of different dimensions from low to high dimensions and then fuses the different dimensions to obtain the global context of the pyramid pool. As shown in the figure, the pyramidal pool module involved in this study contains pyramidal submodels with  $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ , and  $6 \times 6$  dimensions. Finally, the accurate segmentation of the different functional areas in the original rail fastener image is achieved by fusing and convolutionalizing the multidimensional feature elements with the original feature map to generate the final segmentation result map.

**Fastener status detection** is a detection method based on the vector geometry computation proposed based on the a priori knowledge embodied in the semantic segmentation results of the functional regions. From the semantic segmentation results, we can derive the following information. (1) Regardless of the angle of the input image, the edge of the rail region in the image is always a straight line. (2) The

thresholds for the vector pixel distances between different functional regions can be divided by experimentally measuring the images of rail fasteners in different states. (3) If no rail fastener fastclip area is detected in the image prediction results, then the rail fastener is in an unhealthy missing state.

On the basis of the a priori information obtained from the above experimental experience, this paper proposes a method for fastener defect detection and state classification based on vector geometry relationships. The method calculates the vector geometry relationship between the functional regions processed by the semantic segmentation model, determines the threshold of different states according to the specific vector distance, and finally classifies the states according to the defective state of the obtained rail fasteners. The implementation is as follows. First, we determine the linear boundaries of the different functional regions by a least squares linear fitting algorithm. The vector distance and the offset angle from the rail region boundary to the rail fastener region boundary are then calculated. Finally, the defect detection and classification results of the final rail fastener state are obtained by comparing and judging the calculated results with the empirical threshold.

Figure 6 shows two images of rail fasteners in a non-healthy state and the corresponding state detection principle. For the images of unhealthy rail fasteners in the detached and offset states presented in Figures 6(a) and 6(b), we determine the working state of the fasteners in the images by calculating the vector distance and offset angle of the rail and fastclip boundaries, respectively. Let  $M$  represent the semantic segmentation results obtained by the above method, and  $w$  and  $h$  represent the length and width of the image, respectively, with a range of  $[0, 473]$  pixels. Given the regular and smooth rectangle rail region in the image, we

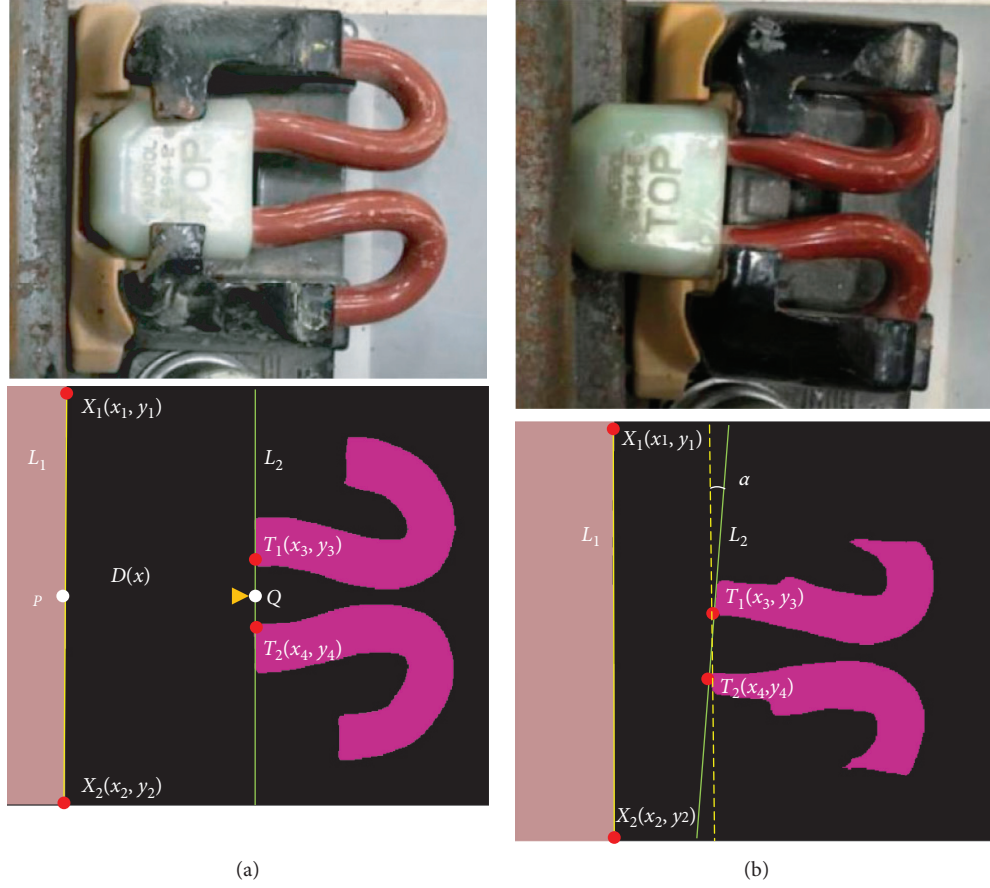


FIGURE 6: Schematic of a state detection method based on vector geometry.

obtain the linear equation ( $L_1: y = ax + b$ ) for the border of the rail region by defining two pixel points ( $X_1(x, 0)$ ) and ( $X_2(x, 473)$ ) as the locus points for the border of the rail region, which are calculated from the coordinates of the two locus points, as shown by the yellow solid line in Figure 6. In addition, we select 30 discrete pixel nodes  $T_i(x, y)$  with the smallest value of transverse coordinate  $x$  in the fastener region as the locator, where  $i \in [1, 30]$ , and then, use the least squares linear fitting algorithm to determine the linear equation ( $L_2: y = cx + d$ ) for the near-rail side boundary of the rail fastener, as shown by the green solid line in Figure 6. The error term is defined as  $S(c, d) = \sum (y_i - (cx + b))^2$ . According to the principle of least squares,  $S$  must be of minimum value, and the condition for obtaining the minimum value of  $S$  is  $(\partial S / \partial c) = (\partial S / \partial d)$ , which gives

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix} \begin{pmatrix} -\sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}. \quad (9)$$

By substituting the coordinates of the location point to the value of parameter  $c$  and  $d$ , the following equations are derived:

$$c = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad (10)$$

$$d = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}.$$

As shown in Figure 6(a), the linear fitting equation ( $L_2$ ) for the near-rail measurement boundary of the rail fastener region is obtained by the above method. Defining parameter  $D(x)$  as the vector distance from the rail region boundary to the fastener region boundary and assuming that  $P(x_{L_1}, y_{L_1})$  and  $Q(x_{L_2}, y_{L_2})$  are the pixel points on rail region line  $L_1$  and fastener region boundary line  $L_2$ , respectively,  $D(x)$  is calculated as

$$D(x) = \sqrt{(x_{L_2} - x_{L_1})^2 + (y_{L_2} - y_{L_1})^2}. \quad (11)$$

To simplify the calculation of the equation, we set  $y$  to 236. Empirically, if the vector distance between the line of the fastener near the pixels on the side of the rail and the rail in the predicted result is less than 100 pixels, then the rail fastener is in the normal state; if the vector distance between the two is in the range of 100–160 pixels, then the rail



fastener is in an unhealthy loose state; if the vector distance between the two is greater than 160 pixels, then the rail fastener is in an unhealthy separate state. If the fastener sling area is not detected, the rail fastener is in the missing state.

In addition, we calculate the degree of deflection of the rail fasteners relative to the rails by calculating the angle of clamping ( $\alpha$ ) of the two boundary lines ( $L_1$  and  $L_2$ ), as shown in Figure 6(b). According to the experimental test, if  $\alpha \in [5^\circ, 90^\circ]$ , then the rail fastener is in the dislocated state; otherwise, it is in the normal state within a reasonable error. Parameter  $\alpha$  is calculated as

$$\alpha = \arccos \left| \frac{1 + ac}{\sqrt{1 + a^2} \sqrt{1 + c^2}} \right|. \quad (12)$$

Calculated vector distance parameter  $D(x)$  and angular offset  $\alpha$  are compared with a predetermined threshold, and the rail fasteners are subjected to defect detection and state classification by threshold judgment. Furthermore, to avoid detection errors due to the simultaneous action of two geometric parameters on the experimental results, the rail fasteners in the unhealthy state satisfy only one of the dislocated or other nonnormal states, prioritizing the influence of the vector distance parameter on the detection results during the experiment.

## 4. Experiments and Analysis

In this section, we first construct a SFC rail fastener dataset by the method described in Section 4.1. A new evaluation metric is then proposed to quantitatively characterize the effect of rail fastener fastclip region positioning based on a significance detection model. In addition, by conducting experiments on the rail fastener image dataset, the results of the experiments performed on the present algorithm are qualitatively and quantitatively validated.

**4.1. Dataset.** Our experimental data were derived from original rail images taken and captured in real railway project scenarios, and we segmented the fastener regions on both sides of the steel rail in the original images and constructed the SFC-type rail fastener dataset using the method described in Section 3.1. Our dataset contains images of rail fasteners on the left and right sides of the tracks, pseudo-labels corresponding to the images, the ground truth, and the true state of the fasteners in the images. In addition, we classified the type of fasteners according to the state of the rail fasteners into positive and negative data samples, where the positive samples include the image data of the rail fasteners in the normal state, and the negative samples include the image data of the rail fasteners in the loose, detached, missing, and dislocated states.

The SFC rail fastener dataset contains 2000 positive sample data in the normal state and 2000 negative sample datasets with defects (including 500 negative sample data in each of the three different states). In addition, we divided the equal amount of rail fastener image data for different state conditions in the dataset into a training set and a test set by random selection, with 2400 samples for training and 1600 samples for testing.

**4.2. Analysis of Experimental Results.** We first quantitatively evaluated the positioning effect of the rail fastener fastclip region and then conducted extensive experiments on the rail fastener image dataset to further qualitatively and quantitatively validate the experimental results of the detection system designed in this work.

**4.2.1. Experimental Analysis of Fastener Railroad Fastclip Positioning.** In existing studies on the positioning of the rail fastener area by computer vision and image processing techniques, existing methods can usually only qualitatively describe and evaluate the results of fastener positioning experiments. Given the lack of uniform evaluation criteria for the positioning results of the rail fastener images, the experimental results of the fastener region positioning are difficult to quantitatively evaluate and compare. Thus, we introduced the evaluation parameters in the visual attention mechanism [30–34].

The Precision-Recall curve, the F-measure [35], and the mean absolute error (MAE) are used as indicators for evaluating the accuracy of fastener positioning for the effective evaluation and analysis of the accuracy of positioning results from a quantitative perspective.

In the fastener fastclip image dataset, a corresponding ground truth was first constructed for each rail fastener image in the dataset by manual labeling at the pixel level. Then, the resultant diagram of the fastener fastclip region localization obtained by this method was threshold selected as a binary segmentation from 0 to 255. Finally, for each threshold condition, the fastclip positioning results plot was compared pixel-by-pixel with the true value plot, and the evaluation parameters of the positioning results relative to the ground truth were calculated. The experimental results on the positioning of the fastener fastclips were evaluated from a quantitative point of view. Precision represents the ratio of the correctly positioned area of the fastener fastclip in the positioning result map to the actual overall area of the fastclip in the positioning result, and Recall represents the ratio of the correctly positioned area of the fastener fastclip in the positioning result map to the theoretical area of the fastclip in the true value map. The F-measure is a comprehensive indicator for evaluating the final positioning result. The MAE is used to measure the positioning error of the positioning result relative to the true value image, which is calculated as follows:

$$\text{Precision} = \frac{T_p}{T_p + F_p}, \quad (13)$$

$$\text{Recall} = \frac{T_p}{T_p + F_N}, \quad (14)$$

$$F_\lambda = \frac{(1 + \lambda^2) \text{Precision} \times \text{Recall}}{\lambda^2 \text{Precision} + \text{Recall}}, \quad (15)$$

$$\text{MAE} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H |\bar{S}(x, y) - \bar{G}(x, y)|. \quad (16)$$

In formulas (13) and (14),  $T_p$  represents the pixel sample data of the real fastener area located, and  $F_p$  and  $F_N$ , respectively, indicate the false and missed pixel data samples of the positioning experiment results relative to the theoretical railroad fastclip area in the truth map. In formula (15), the value of  $\lambda^2$  is 0.3 [35].  $W$  and  $H$  in formula (16) are used to represent the length and width of the input rail fastener image to be processed, and parameters  $x$  and  $y$  represent the horizontal and vertical coordinates of the pixel node in the image, respectively.

We quantitatively evaluated the results of the localization of the fastclip region on a dataset of 200 images of rail fasteners in different states. In addition, the fastener positioning results obtained by this method were compared with the positioning effects of several other classical target saliency detection algorithms on rail fasteners, including the SER [36], SWD [37], SIM [38], SS [39], MR [40], RCRR [41], and ACM [42] algorithms. The representative serial numbers in Figure 7 are 2–8, respectively, and the graphs of the Precision-Recall curve, F-measure, and MAE of various algorithms for locating the fastclip area in the rail fastclip image are presented.

As shown in Figure 7(a), the localization experiments performed for the fast clip region significantly outperform several other saliency detection algorithms. The results show that the proposed localization method achieves the best accuracy of more than 0.9. Although the background region in the localization results was clearly separated from the foreground fastclip region, which can effectively meet the localization and segmentation needs of the postexperiment, the recall values below 0.4 do not correspond to the ideal situation (high accuracy value of 0.9) and show an increasing trend because they were not treated by the model as completely irrelevant black backgrounds. In addition, all methods converged toward an accuracy value of 0.12 when a threshold value near 255 was selected for the maximum recall value. During the convergence of the curves, the accuracy of the methods was significantly better than that of the other algorithms at any Recall value. The experimental results show that the fastener fastclip location effect of the proposed method is more accurate than those of the other methods and has a lower false alarm rate and robustness. Figures 7(b) and 7(c) show an F-measure and MAE of 0.5632 and 0.1302, respectively, for the experimental results obtained by this method. The positioning area of the rail fasteners obtained by this method is more accurate than the positioning effect of several other algorithms, and the average absolute errors of the positioning results are smaller than those of other methods, indicating that this method can position rail fasteners more precisely.

#### 4.2.2. Functional Region Semantic Segmentation Results.

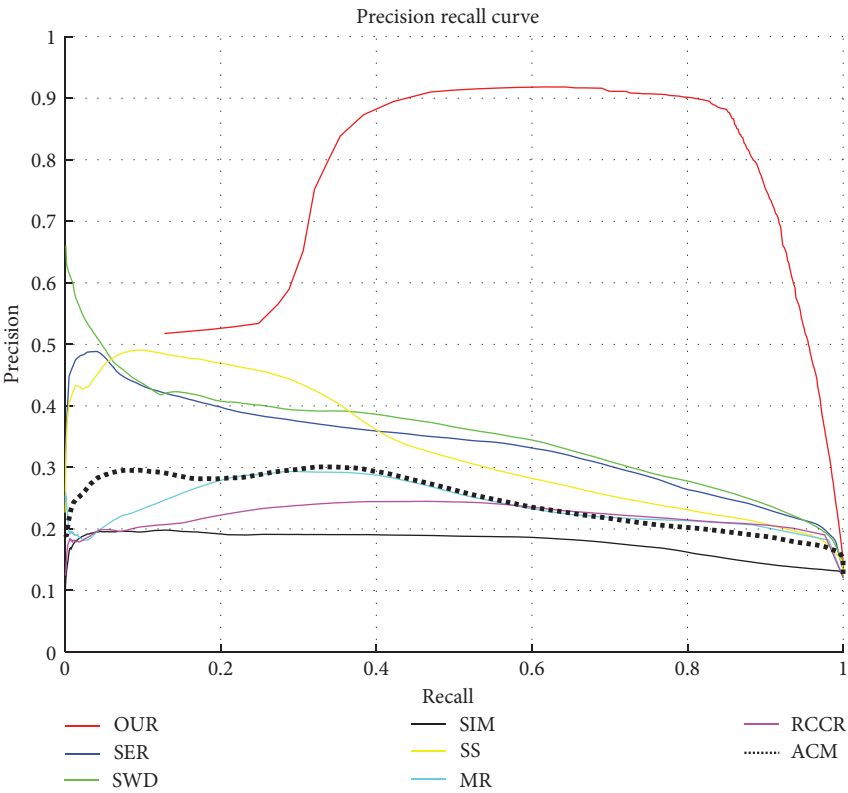
We validated the detection effect of the proposed method on the constructed SFC-type rail fastener dataset. The segmentation network was modeled and trained on the training set based on the pseudolabeling of each image sample, namely, the semantics and segmentation of the “rail” and “fastclip” regions of the image with different properties.

Next, we fed 1600 rail fastener image samples from the test set into a trained semantic segmentation model for automatic region identification and semantic segmentation, enabling accurate targeting and segmentation tasks for specific functional regions in the image samples.

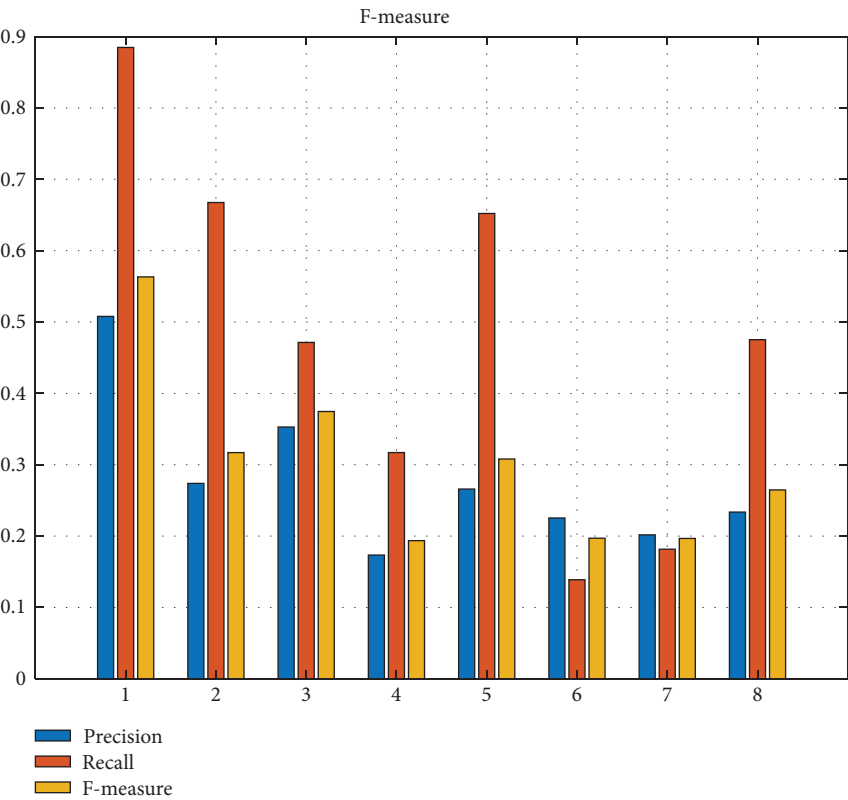
The images of rail fasteners in four different states for the dataset were constructed. Accurate regional localization and semantic segmentation can be performed through the proposed method. The experimental results are shown in Figure 8. Figure 8(a) shows the rail fastener images of the five different fastener state types obtained by this method, where the rail fastener image in the normal, loose, separate, missing, and dislocated states are shown from the top to bottom; Figure 8(b) shows the rail fastener remarkable images for the completed fastclip remarkable area positioning; Figure 8(c) shows the functional area ground truths for the manually labeled multiple fastener images; Figure 8(d) shows the target-level pseudolabels for the rail fastener image constructed by this method, where the red and green areas are the labels for the fastclip and rail areas in the image, respectively; Figure 8(e) shows the predicted functional area results for the rail fastener image for different state types, where the gray area represents the predicted rail area, the white area represents the predicted fastclip area, and the other background and unrelated functional areas are black. In addition, Figures 8(f)–8(j) show the corresponding schematic of the fastener image data on the other side of the rail.

The experimental results in Figure 8 indicate that the functional region semantic segmentation results obtained by this method and shown in Figures 8(e) and 8(j) have an excellent experimental effect compared to the corresponding ground truth shown in Figures 8(c) and 8(h). The experimental results show that our segmentation results are accurate and automatic for the functional areas of interest in all five types of rail fastener images and can effectively filter out other areas and complex backgrounds. Therefore, the region segmentation method established in this work has significant validity and accuracy for rail fastener images.

**4.2.3. Rail Fastener State Detection Results.** In the field of rail fastener detection, the performance of rail fastener detection and classification methods is an important factor in confirming system reliability. For the 800 data samples in the test set used for algorithm validation in different states, we detected and classified the states of the rail fasteners using the proposed vector-based geometry measurement method. In the test results obtained by this method, we assumed that parameter  $T_p$  indicates the number of samples of true positives detected by this method,  $T_N$  indicates the number of samples of true negatives detected,  $F_p$  indicates the number of samples of false positive data detected, and  $F_N$  indicates the number of samples of false negative data detected. The Precision, Recall, and F-measure of the state detection and classification results of the rail fasteners were, respectively, calculated by equations (13)–(15) (the value of parameter  $\lambda$  in equation (15) is set to 1 in this experiment), and the accuracy of the experimental results was evaluated in



(a)



(b)

FIGURE 7: Continued.

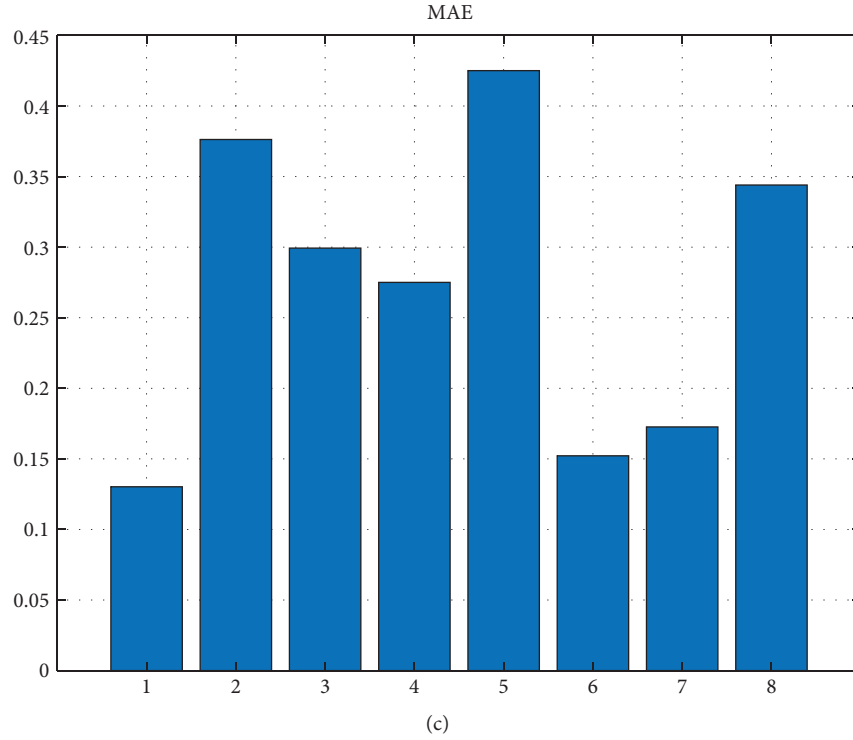


FIGURE 7: Evaluation curve of the results of the positioning of the fastener fastclip area.

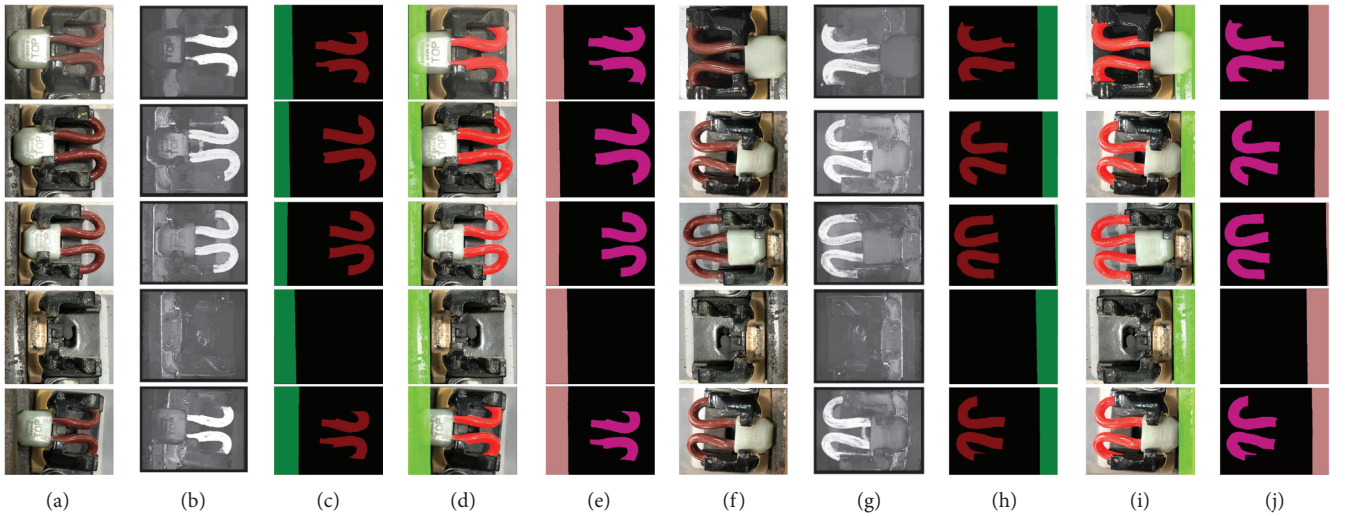


FIGURE 8: Schematic of the prediction results of semantic segmentation of rail fastener images for five state types.

conjunction with the Accuracy index. The mathematical expression of the Accuracy indicator is

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}. \quad (17)$$

The experimental results are shown in Table 1, and the proposed rail fastener detection method can achieve 93.13% Accuracy on the validation dataset (contains 800 positive samples and 800 negative samples of various data types).

Average Accuracy and Recall rates of 92.17% and 90.50% were achieved for the five different states of the rail fastener image data in the validation set, with an average F-measure of 0.9127. In addition, the experimental results indicate that the proposed method can obtain more accurate detection results for samples of rail fasteners in the normal, loosening, separate, missing, and dislocated states.

The experimental results indicate that the rail fastener detection system and proposed method obtain good experimental results on the SFC-type rail fastener dataset.



TABLE 1: Detection and classification results of rail fasteners in five state types.

Fastener status	Type	Experimental results					Evaluation index			
		Normal	Loosing	Separate	Missing	Dislocated	Precision (%)	Recall (%)	<i>F</i> -measure	Accuracy (%)
Real sample	Normal	780	11	4	0	5	94.43	97.50	0.9594	93.13
	Loosing	18	158	15	0	9	85.41	79.00	0.8208	
	Separate	4	9	184	1	2	90.20	92.00	0.9109	
	Missing	0	0	0	200	0	99.50	100.00	0.9975	
	Dislocated	24	7	1	0	168	91.30	84.00	0.8750	

TABLE 2: Accuracy results of state detection of multiple detection methods in SFC-type rail fasteners.

Method	Ours (%)	DCNN [19] (%)	PHOG + SVR [1]	HOG + SVM [10]
Accuracy	93.31	88.79	85.67%	79.29%

Specimens of rail fasteners in normal, loosing, missing, and dislocated states can be obtained with high accuracy in the experiment, but the accuracy against the loose state of the fasteners is only 85.41%. The difference between the characteristics of the image samples of rail fasteners in the loosened state and the normal and disengaged state is small that even experienced professional inspectors cannot accurately determine whether the rail fasteners are in the loosened state in the inspection work. Consequently, detecting rail fasteners in a loose state is difficult and challenging, so the accuracy of the results of testing the state of rail fasteners in the loose state is affected compared to several other samples of state rail fasteners. However, considering the overall experimental results, the proposed method can be used to conduct an accurate state inspection of rail fasteners in different states with obvious accuracy and reliability and can achieve better inspection results in practical engineering applications.

To demonstrate the validity of our method, the detection accuracy metrics of different detection methods were calculated and compared on the validation dataset constructed above. Table 2 shows the total accuracy of different methods for SFC-type fastener images. The table shows that our method significantly outperforms several other detection methods mainly because the other schemes do not consider the state detection of the offset fastener during design and implementation, producing unsatisfactory experimental results and accuracy for this part of the negative sample detection. In addition, the difficulty of detecting the negative sample data in the loose state also affects the detection performance of other methods and can thus affect the overall performance of the detection method to some extent.

To test the performance of the detection method, the proposed algorithm was tested on a dataset of SFC-type rail fastener images taken and constructed in a real railway scenario. The proposed semantic-segmentation-based rail fastener detection method was executed on a computer equipped with an Intel Xeon W-2150B processor (10 cores and 10 threads). Experimental calculations indicate that the time to detect the state and classification of each rail fastener using the proposed method is 1.135 s, which can meet the requirements of practical engineering inspection tasks. In addition, we evaluated the efficiency of our semiautomatic labeling method by randomly selecting 50

images of rail fasteners from several types of data. The average time for the semiautomatic annotation of a single image in the dataset was 5.256 s, while the average time for manual annotation was 132 s. This method is 25 times more efficient than the manual method in annotating all the images. In summary, this method can effectively reduce the cost of manual annotation and has very reliable detection performance and a good generalization ability, which can meet the accuracy and efficiency requirements of engineering inspection scenarios.

## 5. Conclusions and Future Work

In this study, we aim to address the limitations of traditional rail fastener detection methods and deep learning theory in engineering applications. A semantic segmentation-based rail fastener state recognition algorithm is proposed. First, we propose a functional area positioning and labeling method based on the salient detection model in the system, and a novel SFC-type rail fastener dataset was constructed by labeling and constructing pseudolabels for fastclip and rail regions through an interactive semiautomatic labeling method. Then, the fastener state in the original rail image was classified and produced by the detection system designed in this study. Second, a rail fastener state detection method based on the semantic segmentation model was designed, and the fastener state was detected by the semantic segmentation network and vector distance calculation method. The experimental results show that the proposed method has good accuracy and robustness while effectively saving cost and can achieve good experimental results in practical application scenarios.

Although the proposed method was validated by numerous effective experiments on our dataset and achieved promising results, it can still be improved. Although the results based on the saliency detection model satisfy the requirements of the algorithm, the model can be further optimized and improved. Moreover, detecting the state of the rail fastener image of the loose fastclip is still challenging and difficult. In future work, we plan to address some of the above limitations. Nevertheless, we believe that the proposed inspection method is a major step forward in the automation of rail inspection and has significant implications for practical engineering applications.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 51975347 and 51907117) and Key Science and Technology Support Project of the Shanghai Science and Technology Commission (no.18030501300).

## References

- [1] T. Prasongpongchai, T. H. Chalidabhongse, and S. Leelhapantu, "A vision-based method for the detection of missing rail fasteners," in *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 419–424, Kuching, Malaysia, September 2017.
- [2] P. De Ruvo, A. Distanto, E. Stella, and F. Marino, "A GPU-based vision system for real time detection of fastening elements in railway inspection," in *Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2333–2336, Cairo, Egypt, November 2009.
- [3] X. Yuan, B. Liu, and H. Chen, "Algorithm and program design for fastener locating and detection using wavelet transformation and template matching," in *Proceedings of the 2017 IEEE 17th International Conference on Communication Technology (ICCT)*, pp. 1116–1121, Chengdu, China, October 2017.
- [4] J. Yang, W. Tao, M. Liu, Y. Zhang, H. Zhang, and H. Zhao, "An efficient direction field-based method for the detection of fasteners on high-speed railways," *Sensors*, vol. 11, no. 8, pp. 7364–7381, 2011.
- [5] E. Stella, P. Mazzeo, M. Nitti et al., "Visual recognition of missing fastening elements for railroad maintenance," in *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems*, pp. 94–99, Singapore, September 2002.
- [6] H. Fan, P. C. Cosman, Y. Hou, and B. Li, "High-speed railway fastener detection based on a line local binary pattern," *IEEE Signal Processing Letters*, vol. 25, no. 6, pp. 788–792, 2018.
- [7] Y. Xia, F. Xie, and Z. Jiang, "Broken railway fastener detection based on Adaboost algorithm," in *Proceedings of the 2010 International Conference on Optoelectronics and Image Processing*, pp. 313–316, Haikou, China, November 2010.
- [8] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 4, pp. 877–888, 2014.
- [9] R. A. Khan, S. Islam, and R. Biswas, "Automatic detection of defective rail anchors," in *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1583–1588, Qingdao, China, October 2014.
- [10] X. Gibert, V. M. Patel, and R. Chellappa, "Robust fastener detection for autonomous visual railway rail inspection," in *Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 694–701, Waikoloa, HI, USA, January 2015.
- [11] W. F. Zhu, X. J. Chen, Z. W. Li et al., "A SAFT method for the detection of void defect in side a ballastless rail structure using ultrasonic array sensors," *Sensors*, vol. 19, no. 21, p. 15, 2019.
- [12] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 153–164, 2017.
- [13] H. Cui, J. Li, Q. Hu, and Q. Mao, "Real-time inspection system for ballast railway fasteners based on point cloud deep learning," *IEEE Access*, vol. 8, pp. 61604–61614, 2020.
- [14] X. Gibert, V. M. Patel, and R. Chellappa, "Material classification and semantic segmentation of railway rail images with deep convolutional neural networks," in *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP)*, pp. 621–625, Quebec, Canada, September 2015.
- [15] C. Aytekin, Y. Rezaeitabar, S. Dogru, and I. Ulusoy, "Railway fastener inspection by real-time machine vision," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 7, pp. 1101–1107, 2015.
- [16] Q. Mao, H. Cui, Q. Hu, and X. Ren, "A rigorous fastener inspection approach for high-speed railway from structured light sensors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 143, pp. 249–267, 2018.
- [17] H. Zhang, J. Yang, W. Tao, and H. Zhao, "Vision method of inspecting missing fastening components in high-speed railway," *Applied Optics*, vol. 50, no. 20, pp. 3658–3665, 2011.
- [18] Z. Peng, C. Wang, Z. Ma et al., "A multifeature hierarchical locating algorithm for hexagon nut of railway fasteners," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, 2019.
- [19] J. Liu, Y. Huang, Q. Zou et al., "Learning visual similarity for inspecting defective railway fasteners," *IEEE Sensors Journal*, vol. 52, 2019.
- [20] Y. Lin, C. Hsieh, W. Huang, S. Hsieh, and W. Hung, "Railway rail fasteners fault detection using deep learning," in *Proceedings of the 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, pp. 187–190, Yunlin, Taiwan, October 2019.
- [21] B. Dong, Q. Li, J. Wang, W. Huang, P. Dai, and S. Wang, "An end-to-end abnormal fastener detection method based on data synthesis," in *Proceedings of the 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 149–156, Portland, OR, USA, November 2019.
- [22] E. Rahtu and J. Heikkilä, "A simple and efficient saliency detector for background subtraction," in *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops*, pp. 1137–1144, Kyoto, Japan, October 2009.
- [23] A. Criminisi, G. Cross, A. Blake et al., "Bilayer segmentation of live video," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pp. 53–60, New York, NY, USA, June 2006.
- [24] M. Szummer, P. Kohli, and D. Hoiem, "Learning CRFs using graph cuts," in *Proceedings of the European Conference on Computer Vision. railroad fastcliper*, pp. 582–595, Berlin, Heidelberg, April 2008.
- [25] N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa, "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5001–5009, Salt Lake City, UT, USA, June 2018.

- [26] H. Zhao, J. Shi, X. Qi et al., "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890, Honolulu, HI, USA., July 2017.
- [27] 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, Las Vegas, ND, USA, July 2016.
- [28] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, <https://arxiv.org/abs/1511.07122>.
- [29] L. C. Chen, G. Papandreou, I. Kokkinos et al., "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [30] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [31] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 545–552, Columbia, UK, May 2007.
- [32] G. Li, Y. Xie, T. Wei et al., "Flow guided recurrent neural encoder for video salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3243–3252, Salt Lake City, UT, USA, June 2018.
- [33] X. Zhang, T. Wang, J. Qi et al., "Progressive attention guided recurrent network for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 714–722, Salt Lake City, UT, USA, June 2018.
- [34] W. Wang, J. Shen, X. Dong et al., "Salient object detection driven by fixation prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1711–1720, Salt Lake City, UT, USA, June 2018.
- [35] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequencytuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, Miami, FL, USA, June 2009.
- [36] H. J. Seo and P. Milanfar, "Nonparametric bottom-up saliency detection by self-resemblance," in *2009 IEEE Computer Society Conference On Computer Vision and Pattern Recognition Workshops*, pp. 45–52, Miami, FL, USA, June 2009.
- [37] L. Duan, C. Wu, J. Miao et al., "Visual saliency detection by spatially weighted dissimilarity," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR 2011*, pp. 473–480, Colorado Springs, CO, USA, June 2011.
- [38] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.
- [39] X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2012.
- [40] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166–3173, Portland, OR, USA, June 2013.
- [41] Y. Yuan, C. Li, J. Kim, W. Cai, and D. D. Feng, "Reversion correction and regularized random walk ranking for saliency detection," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1311–1322, 2018.
- [42] Q. Zhang, L. Zhang, W. Shi, and Y. Liu, "Airport extraction via complementary saliency analysis and saliency-oriented active contour model," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 7, pp. 1085–1089, 2018.