SCueU-Net: Efficient Damage Detection Method for Railway Rail

JUN LU[1], BO LIANG[1,2], QUJIANG LEI[2], XIUHAO LI[2], JUNHAO LIU[2], JI LIU[2], JIE XU[2], AND WEIJUN WANG[2]

[1]Shaanxi University of Science and Technology, Xi'an 710021, China
[2]Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou 511548, China

Corresponding author: Qujiang Lei (qj.lei@aaia-ai.org)

ABSTRACT Automatic detection of industrial product damage using machine learning is a promising research area. At the same time, various machine learning methods based on convolutional neural networks have a very important role in the application of visual automatic detection. Therefore, the machine vision-based automatic detection of high-speed railway rail damage has received widespread attention. This paper proposes an efficient detection method for the damage of high-speed railway rails called SCueU-Net. For the first time, the combination of U-Net graph segmentation network and the saliency cues method of damage location is applied to the task of high-speed railway rail damage detection. The experimental results show that our method has a detection accuracy rate of 99.76%, which is 6.74% higher than the recent method in damage identification accuracy, which improves the detection efficiency of rail damage significantly.

INDEX TERMS High-speed railway, machine learning, data augmentation, rail damage detection.

I. INTRODUCTION

As the development of high-speed railways around the world enters a booming period, high-speed railways play a pivotal role in the field of passenger transportation. Therefore, railway malfunction is directly related to social and economic development and safety etc. [1]. The surface defects of high-speed railway tracks are one of the most common and important forms of malfunction, include track surface cracks, rolling contact fatigue due to high stress, track head spalling, and track ripples [2], [3]. They are potentially serious risks when high-speed tracks are running, such as high-speed train running noise and train derailment accidents. Therefore, during the daily maintenance of the railway, it is necessary to detect the railway track to ensure reliability, safety and service period.

Historically, detection tasks have been performed primarily by professional engineers. Manual detections not only cause serious waste of human resources, but also bring time-consuming, low accuracy, and subjective assessments. Therefore, automatic and non-destructive testing systems are strongly driven by situation [4]. Recently, many non- destructive testing methods have been developed, such as fiber Bragg grating sensors [5]–[7], ultrasonic testing [8], [9], eddy currents [10]–[12], MEMS gyroscope sensors [13], and visual inspection systems (VIS) [14], [15]. Many of above detection methods will be restricted by the detection environment, but the impact of VIS detection methods on environmental factors shows extraordinary robustness. Therefore, VIS has been widely used in various industry inspections and has received widespread attention [16]. Due to the development of machine vision technology, which applications include not only track detection (such as track profile measurement [17], gauge measurement [18], fastening element inspection [19]–[23], orbit surface inspection [24], [25]), but also extended to a variety of industrial products (such as electronic component housings [26], LCD panels [27], large-caliber optical components [28], magnetic tiles [29], etc.).

With the development of machine vision technology in different detection scenarios, classic machine vision methods can indeed solve many visual detection problems. However, with the upcoming Industry 4.0 area. The automatic manipulation is becoming a requirement, which requires highly adaption to new products [30]. Traditional machine vision is not adapt to diverse scenarios. Generally, the particular domain requires highly matches feature. Decisions are then made

using a hand-made rule-based approach or using a learning-based classifier such as SVM, decision tree, or KNN. Since such classifiers are not as powerful as deep learning methods, hand-made features play a very important role. A wide variety of filter banks, histograms, and wavelet transforms, morphological manipulations, and other techniques are used to manually craft appropriate features. Therefore, the manual engineering of features plays an important role in the tradition method. However such features are not suitable for diverse tasks, and lead to a long development cycle, because the machine vision method must make different manual features to adapt to different product. Due to the many drawback of traditional machine learning methods, we can improve flexibility in data-driven machine learning methods. Using an appropriate amount of training data, the proposed method can quickly adapt to new detection scenarios.

This paper uses the most advanced deep learning methods to solve the problem of surface defects detection. In recent years, image recognition rebased on deep learning methods has become the most common method in the field of machine vision. When applied to surface quality control problems [31]–[34], deep learning methods can achieve excellent results and can be adapted to different situation. Compared with classic machine vision methods, deep learning can directly learn features from the underlying data, has a higher ability to express complex structures, and can completely replace the manual production of features with an automated learning process. With rapid adaptation to new products, this method is well suited for the flexible automation required by Industry 4.0.

In order to solve the problem of visual inspection of discrete defects on the surface of high-speed railway tracks (see Figure 1). We focus on deep learning methods for non-destructive testing based on deep learning methods. Although the conventional visual detection method of track surface damage has achieved remarkable success. However, there are still two major challenges:

a) The background of the detection image is complex, and image noise interference cannot be avoided.
b) The traditional damage image detection methods have weak generalization and low accuracy, lead to lower overall detection efficiency.

Compared traditional visualized surface damage detection methods. This paper proposes a method called SCueU-Net, to detect the damage of the surface image of high-speed railway rails by combining the saliency cues of the damaged area [29] with a U-Net network. It overcomes the interference of image noise and solves the current problem of low detection efficiency. In addition, in order to curtail the research period, this paper selects Type-I RSDDs high-speed railway track dataset [42] for experiment. Experiment results prove that the method is the most advanced way to solve such problems.

Section II introduces the work has been done by others. Section III introduces the details of the proposed method. Section IV discusses the setup of experiment and analysis the result, Section V brings the conclusions and future work.

II. RELATED WORK

The visual inspection system (VIS) includes two main aspects: the image acquisition subsystem (IAS) and the image processing subsystem (IPS) [15], [35]. IAS provides hardware support, while IPS provides the capability to accurately and efficiently analyze large 3D and 2D image databases. IPS is the core part of VIS. Therefore, in recent years, a large number of conventional image processing techniques have been proposed in the related paper in the field of visual inspection. For example, various texture analysis techniques [36] have been widely used for the classification, inspection, and segmentation of orbital surface images. The paper [37] proposed a system for real-time detection of railway track surface using a 3D laser camera. The system marks the health and fault of the track contour images collected by the laser camera, and uses the random forest (RF) method to train to create a decision forest, and then real-time status monitoring of the track. The paper [38] proposed an optimization method based on Chouqet fuzzy integral to improve the overall accuracy of the method described in the paper [37]. The paper [39] proposed a method of using the IMU original to deblur the acquired image and then applying the CNN method to classify the railway track image and then complete the track status monitoring method, and achieved good results. The paper [40] proposed a method for railway line condition monitoring and surface defect detection. In the detection method, the classic image processing method is used, and the relevant features are extracted by applying Canny edge extraction and Hough line transformation to the images captured from the left and right cameras of the track. Use the obtained features to create a decision tree, classify the straight lines in the image through the decision tree, and detect the track status. At the same time, based on the obtained rail surface characteristics, the damage of the rail surface was detected. The paper [41] proposed a method for fuzzy classification of track surface fault types using image characteristic curves of different track defects. The paper [42] proposed a method to analyze the thermal imaging data of railway track and pantograph based on complex fuzzy system, and then to detect the railway status. The paper [43] proposed a method for condition monitoring and fault detection of the acquired orbital image using image processing and particle swarm optimization (PSO) methods. In paper [24], three filtering methods based on texture analysis were tested, and Gabor texture features were selected for railway surface images. Then k-nearest neighbor method (KNN) was used to identify possible defects. The paper [44] replaced the global Gabor texture feature of the orbit image with the local frequency feature to ensure the detection performance of local or small ripples. In [24], the geometric features (area, centroid, filled area, and perimeter) of the defect area were extracted based on the maximum stable extreme value region technology for further processing and analysis. In addition,

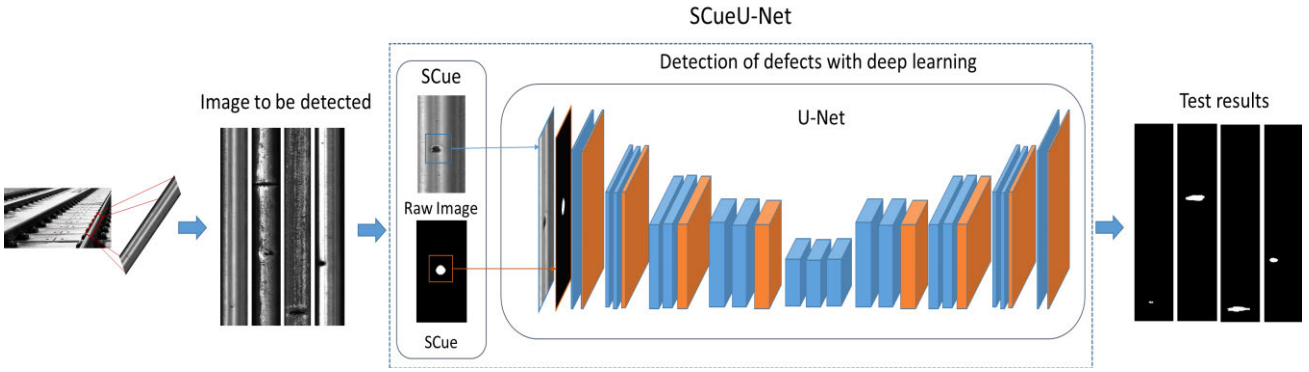


FIGURE 1. Surface damage detection scheme for high-speed railway track.

wavelet transforms using spectral features have also been applied in orbit detection [45], [46]. However, these visual detection methods rely heavily on the extracted global or local feature intensity information to perform well on a uniform texture orbit surface image. However, they cannot detect defects with extreme appearance and dynamic background due to the large amount of random noise caused by complex orbital conditions and image acquisition equipment in actual working environments [47].

In view of the shortcomings of the above visual inspection methods. Another way to model orbital surface images is to automatically locate defects after image enhancement. A method of LN + DLBP [35] is proposed for real-time discrete defect detection. First, the proposed system uses nonlinear and illumination-independent local normalization (LN) methods to enhance the contrast between defects and the background. Subsequently, a defect location based on the projected contour algorithm was proposed to identify possible defects. Instead of LN + DLBP, MLC + PEME was proposed in [15] to detect possible discrete defects in images. MLC (similar to Michelson's local contrast) first enhances image contrast, and then PEME (proportionally emphasizes maximum entropy) automatically locates possible defects by maximizing object entropy, while keeping the defect ratio at a low level. In [48], a new Anti-Perona-Malik (APM) inverse diffusion model was proposed for image enhancement. Then, adaptive threshold binarization can easily find the required defects. The coarse extractor based on the mean shift algorithm in the paper [49] searches the background pattern in the longitudinal direction from the observed sequence, and then finds outliers containing real defects and other noise points. A coarse-fine extractor (CFE) method integrating vertical context information and horizontal prior information is proposed to filter these noise points. Although the above methods have made many attempts to improve the performance of visual inspection, they are still affected by irregular factors (noise points) to varying degrees, so there are still false detections that cannot be tolerated in automated detection scenarios.

Nowadays, deep learning methods based on convolutional neural networks have been rapidly applied and widely concerned in the field of visual detection. Shortly after emerge of AlexNet, deep learning methods began introduced to more surface defect detecting problems [50]. The research in the paper [51] shows that for surface defect classification, deep learning methods can perform better than traditional machine vision methods. Traditional machine vision methods combine hand-designed features with support vector machines. The deep learning method used five convolution layer structure to detect the steel defect of the input image. They achieved excellent results. However, their work was limited to shallow networks because they did not use ReLU and batch normalization.

The paper [52] proposed a more efficient network for to perform defect segmentation. They implemented 10 fully convolutional layer structure network with ReLU and batch normalization techniques to segment defects. In addition, they proposed an additional decision network on top of the segmentation network function to classify the existence of defects sort by image classification. This enables them to improve the accuracy.

The paper [53] proposed a deep segmentation convolutional neural network called U-Net for processing medical images. The network structure is based on full convolutional network. In order to achieve high performance in segmentation with less number of dataset, the paper introduce an upgrade plan for this CNN structure. One of the key upgrading is, during the up sampling, there are many channels which allows CNN sends the contextual information to the large with high resolution ratio. Moreover, the network does not have any fully connected layers and uses only the effective part of each convolution to get the full context in the input image. This method allows seamless segmentation of arbitrarily large images through an overlapping tile strategy. In order to predict pixels in the boundary area of the image, the missing context can be inferred by mirroring the input image.

Compared with related methods, this paper proposes SCueU-Net that a detecting approach of the high-speed railway rails surface damage image of by combining the saliency cues of the damaged area [29] and a U-Net segmentation

network. It overcomes image noise, and raise the detection efficiency of traditional image processing methods. The Type-I RSDDs dataset [49] was enhanced with data and trained on the proposed method. The experimental results were compared with the methods of LN + DLBP, MLC + PEME, paper [45], and CFE. The results prove that our method significantly improves the accuracy of rail damage detection, and significantly improves the efficiency of visual detection of rail damage.

III. PROPOSED APPROACH

A. OVERVIEW

This section introduces the method of high-speed railway track damage detection based on SCueU-Net. Figure 2 shows the overall structure of the proposed method, which includes the three main phases.

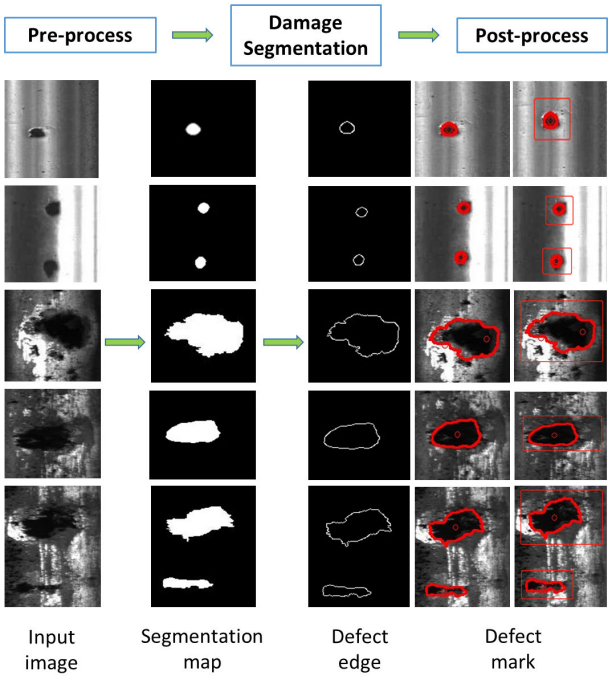


FIGURE 2. Pipeline of Type-I RSDDs inspection.

1) PRE-PROCESS

We obtain the channel and pixel information of the Type-I RSDDs dataset, which is convenient for calculating suitable deep learning network parameters. In addition, because of the sample number of the Type-I RSDDs dataset is small, under fitting is prone to occur during network training, which will cause large surface damage segmentation errors. For this problem, data augmentation processing is performed on the selected dataset.

2) DAMAGE FEATURE SEGMENTATION

In the paper, we will the saliency cues of the orbital damage image [29] and the U-Net deep CNN as a feature extraction method for surface damage detection. This method is guided by the salient features of the damage image, and feeds the detection samples to the U-Net network. The learning network has a large number of functional channels in the up sampling section, which allow the network to propagate contextual information to higher resolution layers. The expansion path is approximately symmetrical with respect to the contraction path in the network, so that it forms a "U" structure [53]. In addition, the network does not have any fully connected layers and uses only the effective part of each convolutional layer, that is, the segmentation map contains only pixels, and a complete context can be obtained in the input image. This strategy allows seamless segmentation of arbitrarily large images through an overlapping tile strategy. In order to predict pixels in the boundary area of the image, the missing context can be inferred by mirroring the input image. This slicing strategy is very important for applying the network to large pictures. Otherwise, the resolution will be limited by the GPU memory.

3) POST-PROCESS

The post-processing stage is based on the segmentation map, calculated by the U-Net network and marked with a rectangular frame on the original input image (see Figure 3). As the conclusion, the detection results of the rail surface damage can be better recognized by the engineer and the engineer can better make an appropriate evaluation of the detection objects.

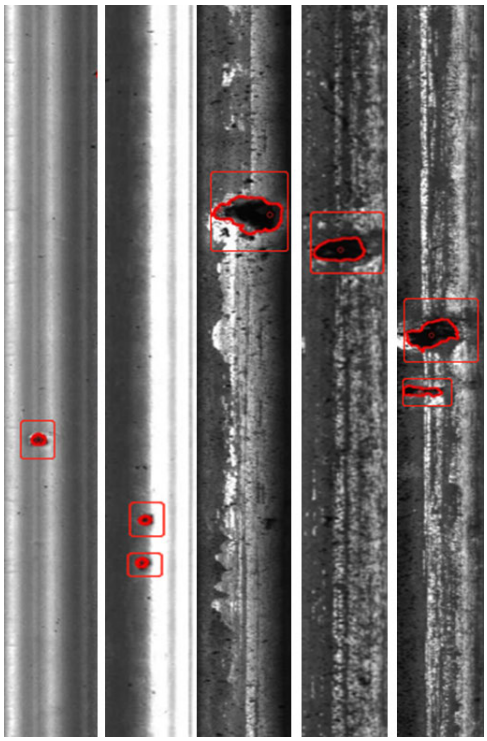


FIGURE 3. Post-processing results of orbital damage images from Type-I RSDDs dataset.

B. PRE-PROCESS

In case of only a few training samples are available, data augmentation is essential to impart the required invariance

and robustness to the network. For orbital damage images, we need to extend the Type-I RSDDs dataset by applying translation, rotation, and scaling invariance. The image samples after data augmentation are shown in Figure 4. In addition, the Dropout layer at the end of the U-Net network contraction path further performs implicit data expansion [46].

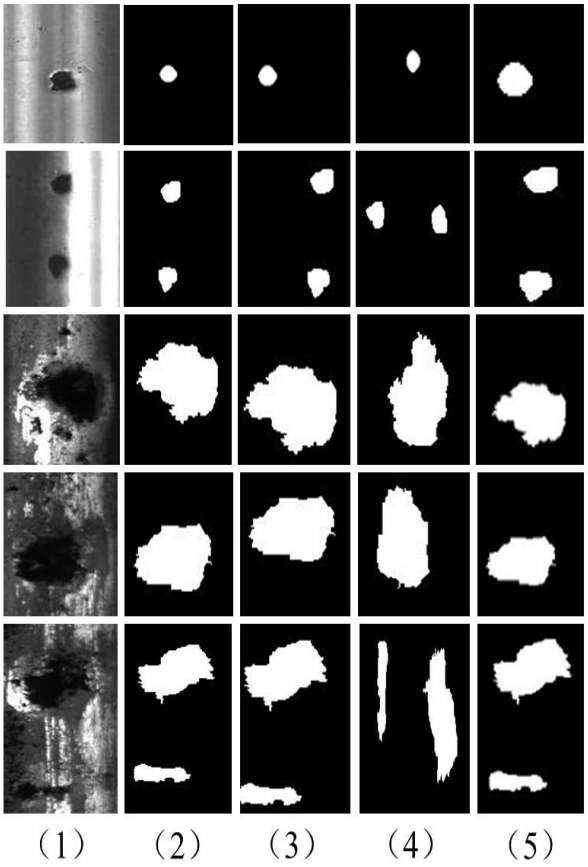


FIGURE 4. Data augmentation from Type-I RSDDs dataset: (1) Defect part, (2) Ground truth, (3) Translation transform, (4) Rotate transform and (5) Scale transform.

C. DAMAGE FEATURE SEGMENTATION

1) DAMAGE FEATURE SALIENCY CUE

The two main types of surface defects, holes and cracks in rails are mainly studied, and their characteristics in grayscale original images are analyzed in-depth. Five important clues that have been found useful for multi-cue saliency construction:

a) The reflection intensity of holes and cracks is weaker than their surroundings;
b) Holes and cracks often have strong corner and edge responses;
c) The wear value of local defects is relatively high;
d) Image background texture rules;
e) Both types of defects can be easily detected by human visual attention mechanisms.

a: DARKER CUE

Once damage occurs, the microscopic geometry of the surface of the rail will change, which will cause the diffuse reflection of the microscopic surface to change. As a result, damage such as holes and cracks appears darker than the surrounding environment. However, conventional binarization still cannot accurately divide those defects, because the illumination has not been standardized in the image. Therefore, adaptive binarization is used to calculate darker clues, which is defined as:

$$D = \begin{cases} 1, & if\ I_R - I > t; \\ 0, & otherwise; \end{cases} \tag{1}$$

where I_R is the mean filter blurred image of the original image I in the R×R blur window, and t is a constant threshold greater than 0. When the gray value of a pixel is less than the average value of its R×R local neighborhood, the pixel is considered darker than another pixel.

b: STRUCTURE TENSOR CUE

Different diffuse reflections between injured and non-damaged areas can cause sharp edges or corners around the damage. The corner or edge response of the defect will be stronger than the background. Harris [54] found that structural tensor can be used to detect angles and edges. The structure tensor is the Hessian matrix of the image. The structure tensor pixels (x, y) of the anchor point are defined as:

$$M = \begin{bmatrix} G & F \\ F & H \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2 I}{\partial x^2} & \dfrac{\partial I}{\partial x} \cdot \dfrac{\partial I}{\partial y} \\ \dfrac{\partial I}{\partial x} \cdot \dfrac{\partial I}{\partial y} & \dfrac{\partial^2 I}{\partial y^2} \end{bmatrix} \tag{2}$$

let λ_1 andλ_2 be the eigenvalues of M, then we get:

$$\begin{cases} \lambda_1 = \dfrac{G + H + \sqrt{(G-H)^2 + 4F^2}}{2} \\ \lambda_2 = \dfrac{G + H - \sqrt{(G-H)^2 + 4F^2}}{2} \end{cases} \tag{3}$$

Harris proved that when one eigenvalue is large and the other eigenvalue is small, an edge response will occur, and a corner response will occur if and only if both eigenvalues are large. Therefore, let $A = (\lambda_1 - \lambda_2)^2$ represent the edge response and the anchor pixel, and $B = |\lambda_1 + \lambda_2|$ represent the angular response of the anchor pixel. Then, through equation (3), we get:

$$\begin{cases} A = (G-H)^2 + F^2 \\ B = G + H \end{cases} \tag{4}$$

Finally, the definition of the structural tensor of the image is:

$$S_{ST} = \frac{N\,(Sal_A) + N\,(Sal_B)}{2} \tag{5}$$

whereSal_A and Sal_B are calculated by equation (4), and represent the saliency map of edges and corners of the input image, and $N\,(\cdot)$ represents the normalization of the saliency map.

c: CUE FUSION

In the third tip, we point out that the gray levels of holes and cracks are distinguishable from the surrounding environment, where the damage color is easily noticeable due to its rarity. Locally rare gray levels of holes and cracks may not be uncommon throughout the image. Capturing color rarity helps pinpoint salient features, which reduces computational complexity. Based on the above considerations, the AC [55] model was chosen to perform this specific task. Specifically, it calculates the target pixel quality and measures the Euclidean distance between the value of the selected center pixel and the average of multiple neighborhoods with different sizes. The color rarity of the local area is thus obtained.

According to the fourth tip we listed, the distribution of the background texture follows a fixed pattern. These textures are approximately parallel and heavy, producing a lot of noise on the predictions. PHOT [56] can be used to remove these interferences from background textures. BMS [57] detects salient objects by simulating human visual attention systems.

Based on the above considerations, a significant clue Scue was proposed. Their construction is:

$$\text{Scue} = \frac{S_{BMS} \cdot (D\omega_D + 1.0) \cdot (S_{AC} + S_{ST} + S_{PHOT}\omega_{PHOT})}{m} \tag{6}$$

whereS_{BMS}, S_{AC}, S_{ST}, S_{PHOT}, and D represent the saliency maps of BMS, AC, structure tensor, and dark cue, respectively. ω_D, ω_{PHOT} are the weights of darker cue, PHOT saliency map, and m is the normalization constant.

The images are added pixel-by-pixel interactively to complement the salient areas of each operating image; these areas are therefore enhanced. At the same time, image doubling will only enhance the areas with higher saliency values in the two images. Otherwise, the areas with lower saliency values in the two images will be weakened.

We implement pixel-by-pixel addition and multiplication on image fusion, where $\omega_D = \omega_{PHOT} = 3$ and$m = 5$. Testing on ST, PHOT and AC, most high probability density regions can be correctly distinguished as surface damage, but still cannot map the entire defect shape. Therefore, we use image addition to complement these saliency maps. PHOT saliency can accurately identify defects based on the high probability density of these regions. So we assign a higher weight toω_{PHOT}. At the same time, due to the strict restrictions on the darker hint D, direct multiplication may cause false detection. We use weights $D\omega_D+1.0$ for image multiplication to relax the constraints on darker cues. Image multiplication reduces the probability of being incorrectly detected as damage.

2) U-NET NETWORK ARCHITECTURE

The network structure is shown in Figure 5. It consists of a contraction path (left) and an expansion path (right) [53]. The contraction path follows the typical architecture of a convolutional network. It consists of the reuse of two 3×3 convolutional layers. Each convolutional layer is followed by a ReLU function activation layer and a maximum pooling layer with a step size of 2 and a pooling kernel of 2×2 down sampling. In each down sampling step, the number of feature channels is doubled. Each step in the expansion path includes up sampling the feature map, followed by a 2×2 convolution to halve the number of feature channels, and cascading the feature maps correspondingly cropped from the shrinking path. There are plus two 3×3 convolution layers, each convolution unit is followed by a ReLU function activation layer. Due to the loss of boundary pixels, each convolution unit must be cropped. In the last layer, a 1×1 convolution is used to map the feature vectors of each 64 channels to the required classes. The network has a total of 23 convolutional layers. In order to seamlessly stitch the output segmentation map, it is important to choose the input tile size. In order to apply all 2×2 maximum pooling operations to layers of equal x and y size.

3) TRAINING

Input image data and implement training of detection network through Pytorch deep learning algorithm platform. Because the convolution is not filled, the output image is a constant border width smaller than the input image. In order to minimize the overhead and maximize the use of GPU memory, we tend to use larger input tiles in large batches, reducing the batch to a single image. Therefore, we use high momentum. In this way, a large number of previously seen training samples can determine the model update in the current optimization step.

The energy function is a combination of a pixel-level soft-max function and a cross-entropy loss function on the final feature map. The soft-max function can be defined as:

$$p_{k(\mathrm{X})} = \frac{\exp\left(a_k\left(\mathrm{X}\right)\right)}{\sum_{k=1}^{K} \exp\left(a_k\left(\mathrm{X}\right)\right)} \tag{7}$$

where a_k (X) represents the feature channel k at the pixel position X $\in\Omega$ with$\Omega \subset \mathbb{Z}^2$. K is the number of classes, and $p_{k(\mathrm{X})}$ is the maximum likelihood function. That is, $p_{k(\mathrm{X})} \approx 1$ for k with the largest activationa_k (X), and $p_{k(\mathrm{X})} \approx 0$ for all other k. The cross-entropy loss function is the deviation of the penalty term $p_{l(\mathrm{X})}$ (X) from 1, and is defined as:

$$E = \sum\nolimits_{\mathrm{X}\in\Omega} \omega\left(\mathrm{X}\right) \log\left(p_{l(\mathrm{X})}\left(\mathrm{X}\right)\right) \tag{8}$$

where l:$\Omega \rightarrow \{1, \ldots, K\}$ is the true label of each pixel; ω:$\Omega \rightarrow \mathbb{R}$ is the weight map we introduced to make some pixels more important in training.

Pixel-level discontinuities between the damaged surface of the rail and the surrounding normal tissue can be calculated using morphological operations to calculate the segmentation boundary. The segmentation weight map is then calculated as:

$$\omega\left(\mathrm{X}\right) = \omega_c\left(\mathrm{X}\right) + \omega_0 \cdot \exp\left(-\frac{\left(d_1\left(\mathrm{X}\right) + d_2\left(\mathrm{X}\right)\right)^2}{2\sigma^2}\right) \tag{9}$$

where ω_c: $\Omega \rightarrow \mathbb{R}$ is a weight map used to balance the class frequency. d_1: $\Omega \rightarrow \mathbb{R}$ represents the distance from the damaged area to the nearest boundary, and d_2: $\Omega \rightarrow \mathbb{R}$ represents

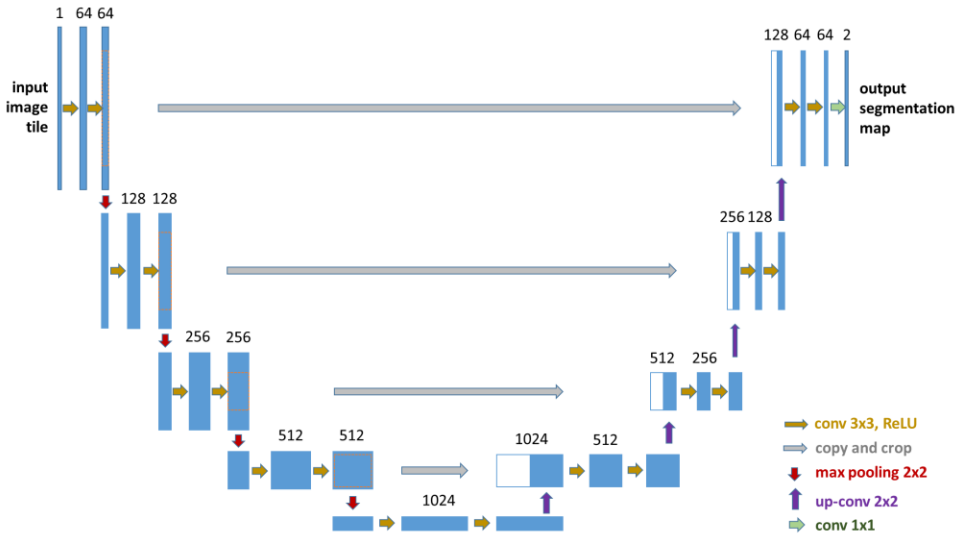


FIGURE 5. Segmentation network architecture of railway track surface damage image based on U-Net.

the distance from the damaged area to the second closest boundary. In our experiments, we set $\omega_0 = 10$and $\sigma \approx 5$ pixels.

In deep learning networks with multiple convolutional layers and multiple paths, the initialization of network weight parameters is very important. Otherwise, some parts of the network may experience repeated activations, while others are ineffective because they are under-activated. Ideally, the initial weights should be adjusted so that each feature map in the network has approximately unit variance. For U-Net network architecture (alternative convolution and ReLU layers). We can implement parameter initialization by extracting initial weights from a Gaussian distribution with standard deviation$\sqrt{2}/N$, where N represents the incoming node of a neuron number [53]. For example, for the 3×3 convolution of the previous layer and 64 feature channels, $N=9 \times 64=576$.

IV. EXPERIMENT AND RESULT ANALYSIS

A. EXPERIMENTAL SETUP

In order to evaluate the accuracy of the orbital damage extractor based on the U-Net network. We implemented the proposed method using Pytorch on the Ubuntu 16.04 operating system, and selected the Type-I RSDDs dataset in an NVIDIA GeForce GTX1080 Ti (11 GiB). This dataset contains multiple types of orbital damage defects (see Figure 6), including 67 original images and their annotated images. After the data enhance, the entire dataset has 201 samples. The augmented dataset is divided into a training set, a validation set, and a test set. The number of samples in each sub-dataset is 141, 21, and 39.

B. EVALUATION METRICS

The evaluation is performed through 3x cross-validation, which ensuring that the same sample will never appear in the training and test sets at the same time. In order to verify

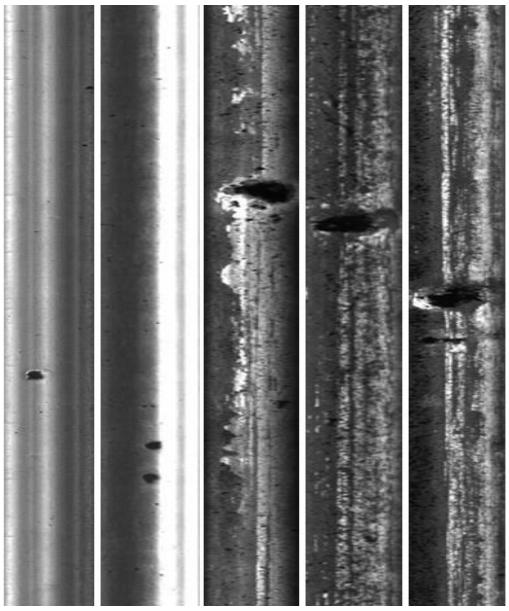


FIGURE 6. Example rail surface images from Type-I RSDDs dataset.

the orbital damage detection method proposed in the paper, we consider four different evaluation indicators: Accuracy (Acc), recall (RE), Loss curve (loss), ROC curve (Receiver Operating Characteristic). The accuracy rate is our most common evaluation index. Generally speaking, the higher the accuracy rate, the better the classifier. The recall rate is the extent to which the detection process covers the entire sample. It measures how many positive examples are classified into positive examples and measures the classifier's ability to recognize positive examples. Because a positive example usually means the existence of a dangerous source in track defect detection, it is necessary to choose the recall rate to evaluate the proposed detection method. Loss curve can prove that the

proposed loss function converges on the learning process of track damage defects. In addition, in the actual detection of orbital damage detection, the positive and negative samples distribution varies widely, and a class imbalance phenomenon often occurs, that is, negative samples are much more than positive samples (or vice versa). In this case the reliability can be assessed by the method of the ROC curve, the large of area under the ROC curve (AUC) the more reliable model.

In addition, we also compare the detection results of the proposed detection method with the methods of LN + DLBP, MLC + PEME, paper [45], and CFE on the same data, and the comparison indicators are accuracy Acc and recall RE. Where Acc and RE are defined by equations (4) and (5):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$RE = \frac{TP}{TP + FN} \tag{11}$$

where TP indicates the number of correctly classified as positive examples; FP indicates the number of incorrectly classified as positive examples; TN indicates the number of correctly classified as negative examples; FN indicates the number of incorrect examples as negative examples.

C. IMPLEMENTATION DETAILS

The network architecture is implemented by Pytorch and trained with the Adam optimizer, where $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate for softmax-cross entropy loss is 0.0002. Considering the image size and GPU memory limitations, we set only a single image for each iteration, that is batchsize = 1.

During the learning process, training samples are selected irregularly. However, the selection mechanism was modified to ensure that the number of positive and negative samples send to the network were consistent. This is achieved by feeding a defective image for each even number of iterations, and a non-defective image for each odd number of iterations. This mechanism ensures that the system learns defective images at a constant rate. On the other hand, because there are more non-defective samples, the learning process becomes unbalanced because more non-defective samples are used, which leads to a significant reduction in the learning rate. The network has been trained for 100 epochs equivalent to 14,100 steps, and alternates between defective and non-defective images at each training step. An epoch is considered to be over if and only if all samples of a sub-dataset are facilitated at one time.

D. RESULT ANALYSIS

We evaluate the performance of method on the Type-I RSDDs training dataset and validation dataset, and test it on the corresponding test dataset. The results obtained are compared with other advanced methods, and the results prove our performance of the method on the corresponding evaluation indicators has indeed improved significantly.

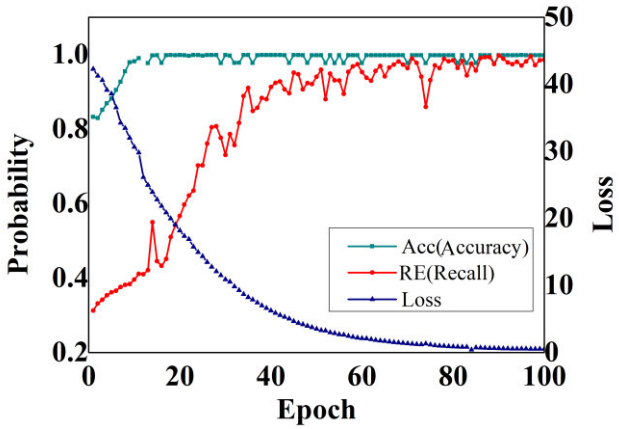


FIGURE 7. Performance curve of the method proposed.

1) PERFORMANCE OF PROPOSED METHOD

We draw the data output during the learning process into a graph as shown in Figure 7. From the figure, we can see that the method proposed in the paper has an accuracy rate (Acc) of about 99% and a recall rate (RE) is maintained of about 98%, and the loss of the entire learning model converges steadily. These evaluation indicators prove that our proposed method can improve accuracy the current problems of rail damage detection, and the detection method is more stable.

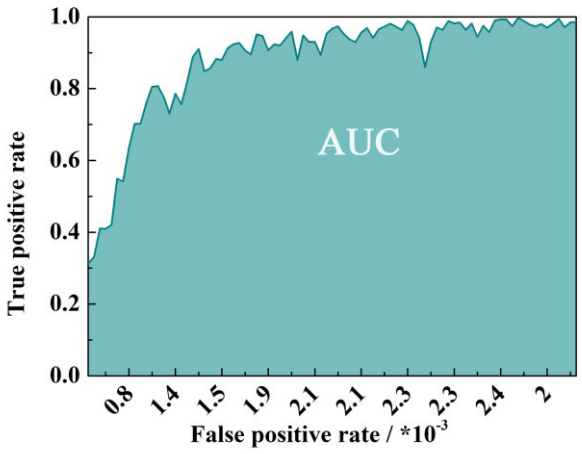


FIGURE 8. ROC curve of the method proposed during training.

In addition, we also draw the ROC curve of the learning model (see Figure 8). From the area under the ROC curve (AUC), we can see that our proposed learning model is more reliable.

2) COMPARISON

We compared the performance indicators of the method proposed in the paper on the test set with other advanced methods (see Figure 9). It can be seen from the figure that our method reached 99.76% in terms of accuracy (Acc). The

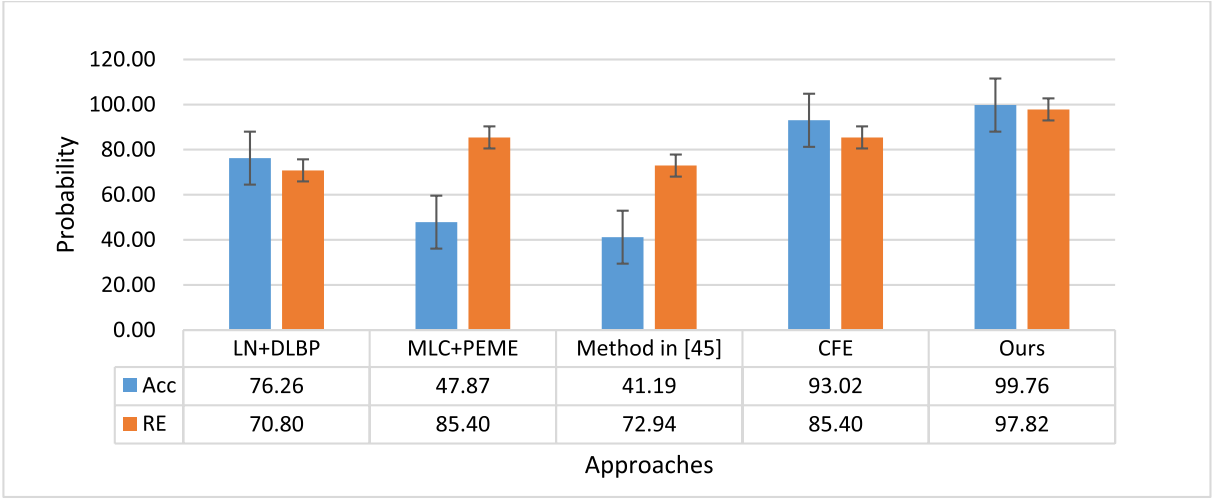


FIGURE 9. Comparison between our method and the other method for Type-I RSDDs dataset.

recall the rate (RE) reached 97.82%, which was 6.74% and 12.42% higher than the highest level in other methods. The comparison results show that our method has good accuracy and recall rate in railway track damage detection tasks, which also means that our detection method has higher efficiency and reliability than other detection methods.

In addition, we compared the visual inspection results of the proposed method with other methods (see Figure 10). The results prove that the final detection result of our method is more intuitive and accurate, which allows engineers to evaluate the railway track damage more quickly and accurately.

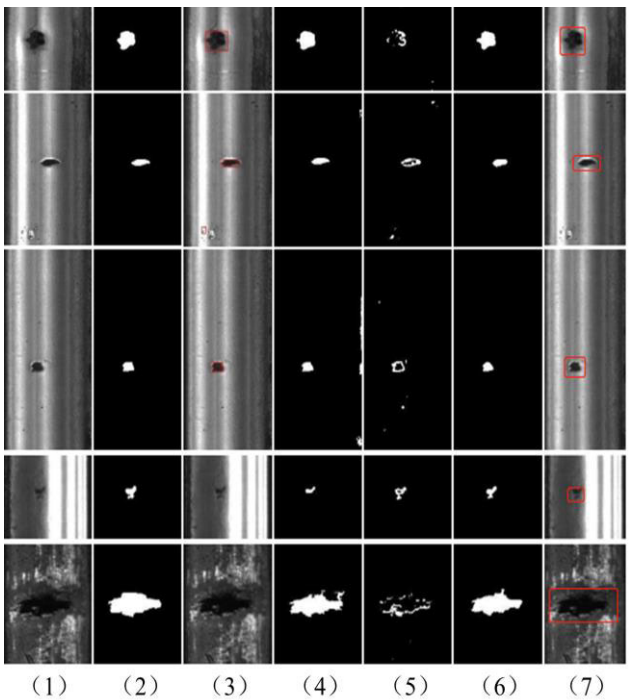


FIGURE 10. Sample defect images and inspection results of different methods from Type-I RSDDs dataset: (1) Defect part, (2) Ground truth, (3) LN+DLBP, (4) MLC+PEME, (5) Method in [45], (6) CFE and (7) Ours method.

V. CONCLUSION AND FUTURE WORK

A. CONCLUSION

In this paper, we propose a method called SCueU-Net, which for the first time completes high-speed railway surface damage detection task by combining the saliency cues of damaged areas [29] and U-Net graph segmentation network [53]. In addition, we choose the Type-I RSDDs to extent dataset as a training sample so that our method can learn the orbital damage features and during the actual detection. By evaluating the accuracy and other indicators of the proposed method, the results prove that our detection accuracy can reach 99.76%, which is an improvement of 6.74% over the best performance of other existing methods. The method produced satisfactory performance of high-speed railway damage visual detection. This result shows a significant improvement for the visual detection of actual track damage. At the same time, our method effectively overcomes the shortcomings of other detection methods. Frankly, our method also has a potential concern. The number and type of data sets used for the experiment are somewhat thin. In actual application, there may not be a high success rate in the experiment. Therefore, this is also our future work direction.

B. FUTURE WORK

In the future, we plan to expand the existing damage image dataset and make appropriate modifications to the U-Net network to make the existing model more robust in order to better apply it to the new visual inspection field and to promote global industry 4.0 Field to make more contributions. Although the developed method is specifically designed for railway inspection tasks, it can be easily extended to other field of industrial products with relatively consistent backgrounds, such as fabrics and paper, by modifying relevant structures and parameters. In addition, we will also use classic image processing methods to reduce the noise of the collected

images, and further reduce the interference of environmental noise during detection.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers and the associate editor for their prolific comments which significantly improved the quality of their work.

REFERENCES

[1] E. Resendiz, J. Hart, and N. Ahuja, "Automated visual inspection of railroad tracks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 751–760, May 2013.

[2] D. F. Cannon, K.-O. Edel, S. L. Grassie, and K. Sawley, "Rail defects: An overview," *Fatigue Fract. Eng. Mater. Struct.*, vol. 26, no. 10, pp. 865–886, Oct. 2003.

[3] R. Clark, "Rail flaw detection: Overview and needs for future developments," *NDT E Int.*, vol. 37, no. 2, pp. 111–118, Mar. 2004.

[4] M. Ph Papaelias, C. Roberts, and C. L. Davis, "A review on non-destructive evaluation of rails: State-of-the-art and future development," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 222, no. 4, pp. 367–384, Jul. 2008.

[5] M. L. Filograno, P. Corredera Guillen, A. Rodriguez-Barrios, S. Martin-Lopez, M. Rodriguez-Plaza, Á. Andres-Alguacil, and M. Gonzalez-Herraez, "Real-time monitoring of railway traffic using fiber Bragg grating sensors," *IEEE Sensors J.*, vol. 12, no. 1, pp. 85–92, Jan. 2012.

[6] M. L. Filograno, P. Corredera, M. Rodriguez-Plaza, A. Andres-Alguacil, and M. Gonzalez-Herraez, "Wheel flat detection in high-speed railway systems using fiber Bragg gratings," *IEEE Sensors J.*, vol. 13, no. 12, pp. 4808–4816, Dec. 2013.

[7] C.-L. Wei, C.-C. Lai, S.-Y. Liu, W. H. Chung, T. K. Ho, H.-Y. Tam, S. L. Ho, A. McCusker, J. Kam, and K. Y. Lee, "A fiber Bragg grating sensor system for train axle counting," *IEEE Sensors J.*, vol. 10, no. 12, pp. 1905–1912, Dec. 2010.

[8] R. S. Edwards, S. Dixon, and X. Jian, "Characterisation of defects in the railhead using ultrasonic surface waves," *NDT E Int.*, vol. 39, no. 6, pp. 468–475, Sep. 2006.

[9] Y. Li, H. Trinh, N. Haas, C. Otto, and S. Pankanti, "Rail component detection, optimization, and assessment for automatic rail track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 760–770, Apr. 2014.

[10] B. Gao, L. Bai, W. L. Woo, G. Y. Tian, and Y. Cheng, "Automatic defect identification of eddy current pulsed thermography using single channel blind source separation," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 913–922, Apr. 2014.

[11] L. Cheng and G. Y. Tian, "Surface crack detection for carbon fiber reinforced plastic (CFRP) materials using pulsed eddy current thermography," *IEEE Sensors J.*, vol. 11, no. 12, pp. 3261–3268, Dec. 2011.

[12] X. Li, B. Gao, W. L. Woo, G. Y. Tian, X. Qiu, and L. Gu, "Quantitative surface crack evaluation based on eddy current pulsed thermography," *IEEE Sensors J.*, vol. 17, no. 2, pp. 412–421, Jan. 2017.

[13] A. Broquetas, A. Comerón, A. Gelonch, J. Fuertes, J. Castro, D. Felip, M. López, and J. Pulido, "Track detection in railway sidings based on MEMS gyroscope sensors," *Sensors*, vol. 12, no. 12, pp. 16228–16249, Nov. 2012.

[14] M. Singh, S. Singh, J. Jaiswal, and J. Hempshall, "Autonomous rail track inspection using vision based system," in *Proc. IEEE Int. Conf. Comput. Intell. for Homeland Secur. Pers. Saf.*, Oct. 2006, pp. 56–59.

[15] Q. Li and S. Ren, "A visual detection system for rail surface defects," *IEEE Trans. Syst., Man, Cybern., C (Appl. Rev.)*, vol. 42, no. 6, pp. 1531–1542, Nov. 2012.

[16] D.-M. Tsai, S.-C. Wu, and W.-Y. Chiu, "Defect detection in solar modules using ICA basis images," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 122–131, Feb. 2013.

[17] C. Alippi, E. Casagrande, F. Scotti, and V. Piuri, "Composite real-time image processing for railways track profile measurement," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 3, pp. 559–564, Jun. 2000.

[18] Z. Zhang, Q. Feng, Z. Gao, C. Kuang, C. Fei, Z. Li, and J. Ding, "A new laser displacement sensor based on triangulation for gauge real-time measurement," *Opt. Laser Technol.*, vol. 40, no. 2, pp. 252–255, Mar. 2008.

[19] P. L. Mazzeo, M. Nitti, E. Stella, and A. Distante, "Visual recognition of fastening bolts for railroad maintenance," *Pattern Recognit. Lett.*, vol. 25, no. 6, pp. 669–677, Apr. 2004.

[20] F. Marino, A. Distante, P. L. Mazzeo, and E. Stella, "A real-time visual inspection system for railway maintenance: Automatic hexagonal-headed bolts detection," *IEEE Trans. Syst., Man Cybern., C (Appl. Rev.)*, vol. 37, no. 3, pp. 418–428, May 2007.

[21] C. Aytekin, Y. Rezaeitabar, S. Dogru, and I. Ulusoy, "Railway fastener inspection by real-time machine vision," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45, no. 7, pp. 1101–1107, Jul. 2015.

[22] 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 Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.

[23] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.

[24] C. Mandriota, M. Nitti, N. Ancona, E. Stella, and A. Distante, "Filter-based feature selection for rail defect detection," *Mach. Vis. Appl.*, vol. 15, no. 4, pp. 179–185, Oct. 2004.

[25] A. K. Dubey and Z. A. Jaffery, "Maximally stable extremal region marking-based railway track surface defect sensing," *IEEE Sensors J.*, vol. 16, no. 24, pp. 9047–9052, Dec. 2016.

[26] D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, Mar. 2020.

[27] Y. Gan and Q. Zhao, "An effective defect inspection method for LCD using active contour model," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 9, pp. 2438–2445, Sep. 2013.

[28] X. Tao, Z. Zhang, F. Zhang, and D. Xu, "A novel and effective surface flaw inspection instrument for large-aperture optical elements," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 9, pp. 2530–2540, Sep. 2015.

[29] Y. Huang, C. Qiu, Y. Guo, X. Wang, and K. Yuan, "Surface defect saliency of magnetic tile," in *Proc. IEEE 14th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2018, pp. 1–6.

[30] E. Oztemel and S. Gursev, "Literature review of industry 4.0 and related technologies," *J. Intell. Manuf.*, vol. 31, no. 1, pp. 127–182, Jan. 2020.

[31] P.-H. Chen and S.-S. Ho, "Is overfeat useful for image-based surface defect classification tasks?" in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 749–753.

[32] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter, "Deep convolutional neural networks for detection of rail surface defects," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Oct. 2016, pp. 2584–2589.

[33] D. Weimer, H. Thamer, and B. Scholz-Reiter, "Learning defect classifiers for textured surfaces using neural networks and statistical feature representations," *Procedia CIRP*, vol. 7, pp. 347–352, 2013.

[34] C.-F.-J. Kuo, C.-T.-M. Hsu, Z.-X. Liu, and H.-C. Wu, "Automatic inspection system of LED chip using two-stages back-propagation neural network," *J. Intell. Manuf.*, vol. 25, no. 6, pp. 1235–1243, Dec. 2014.

[35] Q. Li and S. Ren, "A real-time visual inspection system for discrete surface defects of rail heads," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2189–2199, Aug. 2012.

[36] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 7, no. 3, pp. 1–22, 2008.

[37] Y. Santur, M. Karaköse, and E. Akin, "Learning based experimental approach for condition monitoring using laser cameras in railway tracks," *Int. J. Appl. Math., Electron. Comput.*, vol. 4, nos. 1, pp. 1–5, Dec. 2016.

[38] Y. Santur, M. Karakose, and E. Akin, "Chouqet fuzzy integral based condition monitoring and analysis approach using simulation framework for rail faults," in *Proc. IEEE 14th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2016, pp. 345–350.

[39] Y. Santur, M. Karaköse, and E. Akin, "An adaptive fault diagnosis approach using pipeline implementation for railway inspection," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 26, no. 2, pp. 987–998, Mar. 2018.

[40] M. Karakose, O. Yaman, K. Murat, and E. Akin, "A new approach for condition monitoring and detection of rail components and rail track in railway," *Int. J. Comput. Intell. Syst.*, vol. 11, no. 1, pp. 830–845, Jan. 2018.

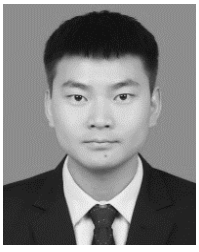
[41] O. Yaman, M. Karakose, and E. Akin, "A vision based diagnosis approach for multi rail surface faults using fuzzy classificiation in railways," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, Oct. 2017, pp. 713–718.

[42] M. Karakose and O. Yaman, "Complex fuzzy system based predictive maintenance approach in railways," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6023–6032, Sep. 2020.

[43] O. Yaman, M. Karakose, and E. Akin, "PSO based diagnosis approach for surface and components faults in railways," *Int. J. Comput. Sci. Softw. Eng.*, vol. 5, no. 5, p. 89, 2016.

[44] Q. Li, Y. Tan, Z. Huayan, S. Ren, P. Dai, and W. Li, "A visual inspection system for rail corrugation based on local frequency features," in *Proc. IEEE 14th Int. Conf. Dependable, Autonomic Secure Comput., 14th Int. Conf. Pervas. Intell. Comput., 2nd Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2016, pp. 18–23.
[45] A. Caprioli, A. Cigada, and D. Raveglia, "Rail inspection in track maintenance: A benchmark between the wavelet approach and the more conventional Fourier analysis," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 631–652, Feb. 2007.
[46] M. Molodova, Z. Li, A. Nunez, and R. Dollevoet, "Monitoring the railway infrastructure: Detection of surface defects using wavelets," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 1316–1321.
[47] H. Trinh, N. Haas, Y. Li, C. Otto, and S. Pankanti, "Enhanced rail component detection and consolidation for rail track inspection," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2012, pp. 289–295.
[48] Z. He, Y. Wang, F. Yin, and J. Liu, "Surface defect detection for high-speed rails using an inverse P-M diffusion model," *Sensor Rev.*, vol. 36, no. 1, pp. 86–97, Jan. 2016.
[49] J. Gan, Q. Li, J. Wang, and H. Yu, "A hierarchical extractor-based visual rail surface inspection system," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7935–7944, Dec. 2017.
[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.
[51] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout, "Steel defect classification with max-pooling convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–6.
[52] D. Racki, D. Tomazevic, and D. Skocaj, "A compact convolutional neural network for textured surface anomaly detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1331–1339.
[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[54] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 147–151.
[55] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. Int. Conf. Comput. Vis. Syst.*, 2008, pp. 66–75.
[56] D. Aiger and H. Talbot, "The phase only transform for unsupervised surface defect detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 295–302.
[57] J. Zhang and S. Sclaroff, "Saliency detection: A Boolean map approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 153–160.

JUN LU was born in Shandong, China, in 1961. He received the B.E. and M.E. degrees from Northwestern Polytechnical University, Xi'an, China. He is currently a Professor with the College of Mechanical and Electrical Engineering, Shaanxi University of Science and Technology, Xi'an. His current research interests include machine vision and intelligent robotic system applications.

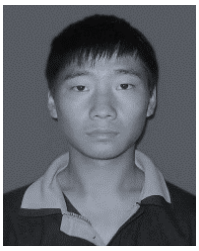


BO LIANG received the B.E. degree in mechanical design manufacture and automation from the Shaanxi University of Science and Technology, Xi'an, China, in 2018, where he is currently pursuing the master's degree. He is currently an Intern with the Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou, China. His current research interests include machine vision and automated inspection.

QUJIANG LEI received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2018. He currently works as an Associate Professor with the Intelligent Robot and Equipment Center, Guangzhou Institute of Advanced Technology (GIAT), Chinese Academy of Sciences, Guangzhou, China. His research interests include intelligent robots, collaborative robots, robotic systems integration, robot vision, artificial intelligence, machine learning, and human–computer interaction.

XIUHAO LI received the master's degree from South China Agricultural University, in 2019. She currently works as an Assistant Engineer with the Intelligent Robot and Equipment Center, Guangzhou Institute of Advanced Technology (GIAT), Chinese Academy of Sciences, Guangzhou, China. Her research interests include computer vision and machine learning.

JUNHAO LIU is currently pursuing the degree with Arizona State University, Tempe, AZ, USA. He is also an Intern with the Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou, China. His research interests include computer vision, computer science, and software engineering.



JI LIU is currently pursuing the degree with the College of Mechanical Engineering, Xi'an Engineering University, Xi'an, China. He currently works with the Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou, China. His main research interests include machine vision and intelligent robots.

JIE XU received the master's degree in mechanical engineering from the Guangdong University of Technology, Guangzhou, China, in 2017. He currently works as an Assistant Engineer with the Intelligent Robot and Equipment Center, Guangzhou Institute of Advanced Technology (GIAT), Chinese Academy of Sciences, Guangzhou. His research interests include system integration of industrial robots, multi-robot cooperated control, and deep learning.

WEIJUN WANG received the Ph.D. degree from Hanyang University, Seoul, South Korea, in 2012. He currently works as an Associate Professor with the Intelligent Robot and Equipment Center, Guangzhou Institute of Advanced Technology (GIAT), Chinese Academy of Sciences, Guangzhou, China. His research interests include intelligent robots, human–computer interaction, and mechanical design.