J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 6591--6608

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Railway Track Image Analysis and Recognition Technology Based on Image Segmentation Algorithm

Yuefei Liu[1, ✉]

[1] *School of Geodesy and Geomatics, Wuhan University, Wuhan, Hubei, 430079, China*

ABSTRACT

The article takes the defect detection and recognition of railroad track as the main research point, and extracts, preprocesses and corrects the railroad track surface image by introducing image segmentation algorithm. Gabor function, K-means clustering method and conditional iterative pattern algorithm are embedded in the original Markov random field model to construct the improved two-layer graph model for railroad track defect segmentation. The recall, precision, mean average precision, and loss function of the improved Markov defect segmentation model are significantly better than those of the original model, and the mean average precision of the defect segmentation model is increased to 95.7% after the Gabor function, K-means clustering method, and conditional iterative pattern algorithm are applied. The improved Markov defect segmentation model fused with clustering features in this paper can better meet the classification and identification of railroad track defects.

Keywords: image segmentation, Markov random field, clustering features, defect segmentation mode, defect recognition

1. Introduction

As an important transportation infrastructure, railroad is an important part of a country's comprehensive transportation system, which plays an important role in supporting and promoting the rapid development of the national economy and society, guaranteeing the security of the national territory and energy, and providing comfortable and satisfactory travel services for the people [1-2]. With the increasing scale of China's railroad network and passenger transportation turnover, coupled with the complex and changing environmental factors such as operating weather conditions, the railroad department's requirements for the railroad safety operation monitoring system in terms of monitoring capability, early warning and prediction capability have been continuously improved [3-5]. At present, most of the railroad lines to implement video surveillance monitoring, the arrangement of video images obtained by the camera to manually realize the observation, the degree of automation

✉ Corresponding author.
E-mail address: yuefeiliu@whu.edu.cn (Y. Liu).

Received 05 March 2024; Revised 20 April 2024; Accepted 05 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-361

is low, often cost a lot of manpower and can not guarantee the reliability of the detection [6-8]. Only rely on human detection, fixed-point installation of monitoring points can not meet the needs of society to realize automation [9]. The application of machine vision combined with image processing technology for track obstacle detection can improve the automation degree of the detection system, but it is easily affected by the complex environment and environmental changes, and cannot achieve the expected processing effect, with lower reliability and accuracy [10-13]. And because these image data are unstructured data, the traditional analysis methods based on structured data are no longer applicable, while the shallow machine learning methods in the past are not sufficient to mine and analyze the data, resulting in a large amount of valuable information is not effectively used [14-17]. How to design a segmentation algorithm with good real-time performance and strong robustness is the key to quickly realize the identification and analysis of railroad track images based on machine vision.

A large number of scholars have conducted research on how to enhance the railroad safety monitoring system and improve the collection, mining, analysis, and identification of railroad operation safety big data. Wang, Y. et al. designed an adaptive railroad track region segmentation algorithm and used the Hough transform to synthesize the segmented image fragments into localized regions according to clustering rules to achieve accurate and fast classification and identification [18]. Shang, L. et al. proposed a two-phase railroad track defect monitoring method using image segmentation processing to generate cropped track images and input them into a convolutional neural network for classification, which significantly improved the accuracy of track defect monitoring [19]. Wei, X. et al. discussed a multi-objective defect recognition method for railroad track fasteners based on image processing and deep school technology, introducing variance projection and wavelet transform to process image data and combining with deep learning algorithms to extract and categorize image features, which effectively improves the accuracy and efficiency of the model for defect detection [20]. Gao, H. and He, T. introduced the region of interest delineation method into the detection of foreign object intrusion on railroad tracks, and used the improved Sobel image edge detection algorithm to delineate the region of interest of the track image to improve the accuracy of track image detection and recognition [21]. Wei, D. et al. constructed a railroad track line image segmentation model (RTLSeg) based on YOLACT and YOLACT++, and at the same time introduced the attention mechanism to further improve the detection and segmentation accuracy of the railroad track image, which meets the demand of the railroad track for rapid detection and accurate recognition [22]. The above literature shows that the combination of image segmentation processing algorithm and deep learning network plays an important role in realizing image recognition, segmentation and detection in railroad locomotive traveling process and shunting operation scene, and the information it mines provides support for intelligent analysis and decision-making of railroad operation safety.

In this paper, the image processing method is used on the railroad track to extract and preprocess the image of the railroad track area. In order to reduce the image error, the railroad track surface image is geometrically corrected and a projection model is established. Markov random field method is selected to segment the defects of the railroad track, Gabor preprocessing, K-means clustering and conditional iterative pattern algorithm are introduced to establish the improved two-layer graph model, and the improved Markov defects segmentation model fused with clustering features is constructed. Defect detection experiments are conducted on the improved defect segmentation model to understand the detection accuracy of the improved Markov defect segmentation model with fused clustering features. In addition, the defects of railroad tracks are classified and recognized to further explore the recognition effect of the model in this paper.

2. Railway track surface area extraction and image pre-processing

Image segmentation technology is a key step from image processing to image analysis, which plays an increasingly important role in remote sensing, military, meteorology and other fields. In this paper, it is applied to the railroad track to detect and recognize the defects of the railroad track. The original image collected on site contains noise and interference areas (track, gravel, weeds, etc.), if the defective target is detected directly from the original image, it will inevitably be interfered by the gravel on both sides of the track or weeds, etc., and these interference areas not only increase the difficulty of detecting defective targets, but also increase the detection time. For this reason, we should first separate the interference region and the rail area, extract the rail surface area (generally known as the rail surface area), and then pre-process the extracted rail surface area to attenuate the interference of noise and enhance the contrast of the image, to provide a good condition for the subsequent rail defect image segmentation, defective feature extraction and so on.

2.1. Track surface area extraction

Analysis of the collected original images shows that: the surface area of the rail is high and more uniform due to long-term friction with the train wheels, while the interference area usually includes the side part of the rail, the gravel or weeds on both sides of the track, which has a lower brightness. Based on this, this paper proposes a maximum brightness and rail surface area extraction method based on RGB color model, which is not only applicable to the rail surface extraction of color rail images, but also can be accurately extracted for grayscale images, has the advantages of wider applicability and fast speed, and can quickly and efficiently extract the rail surface area from the original image which contains the surface area of the rail and a large number of interference areas.

The rail surface extraction algorithm based on the maximum brightness sum is extracted according to the brightness value of the rail surface region in the captured original image is greater than the brightness value of the non-rail surface region. The specific process is as follows:

Step1: Calculate the maximum luminance sum of each column of the input image to generate the maximum luminance sum curve. Let the captured original image be f, the maximum luminance sum of image f in column x is $sum(x)$, and the formula for $sum(x)$ is:

$$sum(x) = \sum_{y=0}^{H} F(x,y) \tag{1}$$

Where, H is the height of image f, $F(x,y)$ is the maximum luminance value of image f at position (x,y), i.e. $F(x,y) = \max(f_R, f_G, f_B)$, (f_R, f_G, f_B) is the value of (R,G,B) at (x,y).

Step2: Coordinate transformation. The vertical coordinate of the maximum brightness and curve is $sum(x)$, and the horizontal coordinate is the width of the rail image W, when the rail image and the maximum brightness and curve are superimposed, it is not possible to display the maximum brightness and curve in the captured rail image, (the size of the captured image is 480pixel × 640pixel, then its height is 640pixel, and the maximum brightness and the maximum brightness of the column is the sum of the values of the 640 $\max(R,G,B)$, which is much larger than 640). This value is much larger than 640), so the following transformation of the vertical coordinates needs to be done, and the new vertical coordinates are $sum^1(x)$:

$$sum^1(x) = sum(x)/255 \tag{2}$$

Step3: Determine the start point BY and end point BY2 of the track surface boundary. Since the brightness of the surface area of the rail is greater than the mean value of brightness in Image f, the start and end values of the boundaries can be determined by automatically finding boundaries whose brightness is greater than the mean value:

$$Psum/255 = \sum_{x=1}^{W} sum^1(x)/(W \times H) \tag{3}$$

Where Psum is the average value of luminance of image f . Since $Psum$ is smaller than the luminance value of the rail surface, the first column greater than $Psum$ is set as the starting boundary value BY, and the last column greater than $Psum$ is set as the end value of the rail surface boundary BY2, and finally the rail surface region is extracted based on BY and BY2.

2.2. *Rail surface image pre-processing*

2.2.1. Rail Image Filtering. In the process of image acquisition and transmission, due to the performance of the CCD sensor, image acquisition card and the influence of the external environment, the captured pictures often contain noise, such as pretzel noise and Gaussian noise, etc. These noises will not only cause blurring of the details of the image and reduce the quality of the image, but also bring a lot of difficulties to the image processing, which will have a direct impact on the image segmentation, feature extraction and so on. Therefore, it is necessary to preprocess the extracted rail surface area. In this paper, different templates of mean filter, Gaussian filter, and median filter are selected to filter and analyze the extracted rail surface area, and finally the filtering algorithm suitable for this paper's data set is selected. Considering that the color image processing is more time-consuming, the complexity of the algorithm is higher, in order to consider real-time, the extracted rail surface image needs to be converted into gray-scale image for subsequent processing.

1) Mean value filtering. Mean value filtering is a linear filtering algorithm for local spatial domain processing [23], and its basic idea is neighborhood averaging, i.e., averaging the grayscale of the pixels in the neighborhood of the image, and taking the average value as the grayscale value of the target point, so that the impact of the noise point can be shared to the pixels in the neighborhood, thus reducing the overall impact of noise on the image. Although mean filtering can quickly filter out granular and Gaussian noise from an image, the blurring of image details increases accordingly as the radius of the filtering template increases.

The mean filter achieves the filtering process by weighted averaging the pixels in the neighborhood, let the input image be $f(x,y)$, take a neighborhood S for each pixel point of $f(x,y)$, calculate the mean value of the gray level of all pixels in S, and use it as the pixel value of the image $G(x,y)$ after the neighborhood averaging process, viz:

$$G(x,y)=\frac{1}{M}\sum_{(i,j)\subset S} f(i,j) \tag{4}$$

Where, S is the neighborhood window, M is the number of points of pixels in the neighborhood, and $G(x,y)$ is the gray value at (x,y) after mean filtering.

2) Gaussian filtering. Gaussian filtering is a simple and fast linear smoothing denoising algorithm [24]. The Gaussian filter gives different weights to different pixels in different locations when processing the image, with higher weights in the center and lower weights away from the center, the significance of such an arrangement of weights is that when filtering with this template, the overall gray scale distribution of the image can be retained more. Gaussian filter has a good filtering effect for white noise, but the adaptive function is poor due to the neglect of the image local gray scale similarity and local texture differences. The Gaussian filter is denoted as:

$$G(x,y)=\frac{1}{2\pi\sigma^2}\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right](-L\le x,y\le L) \tag{5}$$

where σ is the standard deviation and L is the template size.

3) Median Filtering. Median filtering has been widely used to filter out pepper noise and smooth image edges [25]. Linear filters (e.g., mean filter, Gaussian filter, etc.) will cause image blurring as the filter template increases during the smoothing and denoising process, while median filtering, as a typical nonlinear filtering technique, can avoid the problem under certain conditions. The principle is to sort

the pixel values in the sliding window and replace the center value of the window with the sorted median value, the median filter is defined as:

$$g(x,y) = Median\{f(k,l)\} \tag{6}$$

Where Median represents the operation of taking the median value, $f(k,l)$ is the gray value of pixel point (k,l) within the window. The size of the window is usually odd, if it is even, the median gray value of the center point of the window is equal to the average of the gray values of the middle two pixels.

2.2.2. Image contrast enhancement. Scientists have done a lot of research on the information processing mechanism of human visual system from biology, anatomy, neurophysiology, etc. Research shows that the human eye has a certain adaptive regulation of light intensity, but the visual system is not sensitive enough to the absolute brightness, and it needs to rely more on the luminance difference between the target and the background to make a judgment. Therefore, in order to better segment the track surface image, this paper introduces the gray scale contrast map algorithm to enhance the contrast between the background and the defects. The grayscale contrast map algorithm converts the grayscale map into a grayscale contrast map, and the conversion formula is:

$$C(x,y) = (R(x,y) - g(x)) / (R(x,y) + g(x)) \tag{7}$$

where $C(x,y)$ represents the generated grayscale contrast map, and $R(x,y)$ and $g(x)$ represent the gray value at (x,y) and the average gray value in column x of the grayscale image, respectively.

3. Orbital image correction

Since the camera cannot take pictures right above the track, we need to correct the position of the pictures taken. Use appropriate methods to minimize measurement errors.

3.1. Geometric Transformation Processing of Orbital Images

An image geometric transformation is a transformation of the spatial location of an image [26], which maps a coordinate position in the original image to a new coordinate position in another image. The key when we work with orbital images is to determine this mapping relationship, as well as to find the parameters of the transformation during the mapping process.

The geometric transformation of the track image does not affect the gray value of the image, which is good for identifying the defects of the image, the geometric transformation only changes the position of the elements in the image. The geometric transformation requires two parts of the operation: firstly, the mapping relationship in the image, for example, angular transformation or translation, determines the calculation method of the original image and the output image. In addition, attention needs to be paid to the distortion that is brought to the processed image when the interpolation algorithm is applied in the calculation process, because by performing the image transformation calculation, the pixel coordinates of the original image may be non-integer coordinates in the output image. This requires a difference calculation.

Let the original image $f(x_0, y_0)$ be geometrically transformed to produce a target image of $g(x_1, y_1)$. The transformation relation in this space can be expressed as:

$$x_1 = s(x_0, y_0) \tag{8}$$

$$y_1 = t(x_0, y_0) \tag{9}$$

where $s(x_0, y_0)$ and $t(x_0, y_0)$ are the coordinate transformation functions from $f(x_0, y_0)$ to $g(x_1, y_1)$.

The formula is as follows:

$$[x_1 \quad y_1 \quad 1] = [x_0 \quad y_0 \quad 1]\begin{pmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{pmatrix} = [x_0 * S_x \quad y_0 * S_y \quad 1] \tag{10}$$

The magnification factor of the image in the x-axis direction is S_x, and the magnification factor in the y-axis direction is S_y. The magnification of the image is performed by setting their values.

In the process of calculation, it is sometimes necessary to perform the inverse operation on the image, and its inverse operation formula is as follows:

$$[x_0 \quad y_0 \quad 1] = [x_1 \quad y_1]\begin{pmatrix} \frac{1}{S_x} & 0 & 0 \\ 0 & \frac{1}{S_y} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \left[\frac{x_1}{S_x} \quad \frac{y_1}{S_y} \quad 1\right] \tag{11}$$

It can be seen from the transformation of the image, the main work is to find out the relational equation of the transformation, and then the data processing in the process of transformation, this chapter through modeling to transform the image of the track, first of all, through the Matlab pairwise comparison of the image with the standard image, to obtain the coefficients in the transformation, in the data processing, for the non-integer coordinates of the point, the interpolation algorithm, the fitting instead. Finally, VC is applied to the program to correct the image.

3.2. *Interpolation Algorithm*

There are two methods for realizing geometric operations. The first is known as the forward mapping method, which is based on the principle of transferring the grayscale of the input image to the output image pixel by pixel, i.e., calculating the coordinates of the target image from the coordinates of the original image: $g(x_1, y_1) = f(a(x_0, y_0), b(x_0, y_0))$. Simple transformations of an image (transformations such as translation, mirroring, and so on) are generally performed in this way. The other is called the backward mapping method, which is the opposite of the forward mapping calculation method for image transformations, where a raster of pixels in the output image is first created and then corresponded one by one to the pixel coordinates in the original image. In this thesis, the camera is not mounted directly above the track and the image needs to be deskewed, the second mapping method is used.

In the backward mapping, interpolation calculation is needed, and the more commonly used methods are nearest difference, bilinear interpolation, and higher order interpolation. Nearest difference is simpler compared to other interpolation methods, where the value of the output pixel is determined by the pixel value of its nearest neighboring sampling point.

Nearest interpolation can be expressed as:

$$f(x, y) = g(round(x), round(y)) \tag{12}$$

Bilinear interpolation is a more accurate algorithm for processing the two-dimensional orientation of an image. It is obtained by interpolating the grayscale of a series of points around a pixel. The general form is the closest 2×2 neighborhood, and the weighted average of the other points is calculated to determine the value of that point.

Given that the point mapped back to the original image is known and is not at an integer coordinate position, and that the four points closest to that point are unit squares $f(0,0)$, $f(1,0)$, $f(0,1)$, and $f(1,1)$, use the above method to compute the value of point $f(x, y)$ inside the square.

The first two points at the upper end of the square are calculated to obtain:

$$f(x,0) = f(0,0) + x[f(1,0) - f(0,0)] \tag{13}$$

Calculate again for the two points at the lower end:

$$f(x,1) = f(0,1) + x[f(1,1) - f(0,1)] \tag{14}$$

Finally, the calculation is obtained in the vertical direction:

$$f(x,y) = f(1,y) + y[f(x,1) - f(x,0)] \tag{15}$$

Organize the equations above:

$$\begin{aligned} f(x,y) &= [f(1,0)+f(0,0)]x + [f(0,1)-f(0,0)]y \\ &+ [f(1,1)+f(0,0)-f(0,1)-f(1,0)]xy + f(0,0) \end{aligned} \tag{16}$$

Higher order interpolation is a more computationally cumbersome method. In image processing, bilinear interpolation utilizes linear transformations of gray levels that degrade the details of the image, and the slopes of the changes in the gray values may be discontinuous, which can produce effects we don't want, such as image distortion. These problems can be solved by higher order interpolation, which operates through convolution. The value of the output pixel is arrived at based on the value of other point pixels in the 4×4 field closest to it. In this way the image details are better preserved, but the computational effort increases considerably during image processing.

The results obtained by the recent interpolation method are worse for the boundary, and a comparison of the three interpolated images shows that the jaggedness of the boundary decreases in order as the order increases. The boundary of the recent interpolation is clearly not as good as the other two, while a comparison of the higher order interpolation with the bilinear interpolation shows that the difference between the two is not very significant. Higher-order interpolation is better because there are more sampling points, and the effect of more points is more accurate, but the number of points involved in the calculation increases the complexity of the calculation. In the process of running the program, the higher-order interpolation time is obviously longer than the other two methods, here we use the bilinear interpolation method. This method can satisfy the accuracy and save the calculation time at the same time.

3.3. *Modeling the projection*

In the previous section, we have briefly introduced the image transformation, in the process of image correction, the key is to find the transformation relationship between images, once the transformation relationship of the image is determined, the image can be programmed according to the algorithm mentioned above to transform the image. Let the original image be $f(x_0, y_0)$, through the mapping transformation, the obtained image is $g(x_1, y_1)$. In the actual operation, it is not easy to get the mapping relationship. In addition, it is also difficult to determine the parameters of the transformation.

For the orbit correction in this topic, we already have a suitable analytical is to describe it, i.e., modeling by bilinear equations:

$$x_1 = s(x_0, y_0) = c_1x_0 + c_2y_0 + c_3x_0y_0 + c_4 \tag{17}$$

$$y_1 = t(x_0, y_0) = c_5x_0 + c_6y_0 + c_7x_0y_0 + c_8 \tag{18}$$

4. Improved Markov defect segmentation model incorporating clustered features

4.1. *Application of Markov Random Fields to Machine Vision*

In rail surface defect segmentation applications, the rail surface image can be considered as a two-dimensional grid, where each grid point (either the result of the super-pixel segmentation can be considered as a grid point or each pixel point in the pixel matrix of the image can be considered as a grid point) corresponds to a pixel. Each grid point can have different states, such as normal, cracked, worn, etc. Accurate detection of rail surface defects is achieved by constructing a Markov model that utilizes the spatial correlation between grid points to infer the most probable state of each pixel point. The advantage of Markov random field in image segmentation task is that it can handle complex scenes and local details in images well, and more accurate segmentation results can be obtained by training and optimizing the model.

4.2. Segmentation modeling of rail defects

4.2.1. Gabor preprocessing. The recognition effect of rail defect images may be affected by a variety of external environmental factors, including the rail's own environment as well as lighting conditions. Referring to the role of convolutional layer in deep learning, the Gabor function convolution process is introduced into the model and this is used as the input preprocessing step for detecting defective datasets. Gabor function is a kind of linear filter [27], which can extract the features of the image in multiple directions and frequencies, by applying the Gabor function to the pre-processing of rail defect images, it can effectively reduce the influence of reflections and shadows of the rail track surface on the image quality, highlight the useful feature information, and make the features of the processed image more obvious, while the dimensionality of the data has also been reduced. Convolution kernels have different directions and frequencies, which can comprehensively cover various features of the image. By applying these convolution kernels to an image, the feature information of the image in different directions and frequencies can be extracted.

The impulse response of the Gabor function is defined by the product of a sinusoidal wave and a Gaussian function, and its frequency and directional properties are similar to those of the human visual system. Therefore, it is often applied to analyze the presence or absence of content of a particular frequency and direction at a specific point or region. Using the Fourier transform, the response of the Gabor function can be expressed as the result of the convolution of the Fourier transforms of each of the sinusoidal and Gaussian functions. In addition, the Gabor function contains real and imaginary components representing orthogonal orientations, which can be combined into a complex form or used individually.

The complex expression of the Gabor function is shown in equation (19):

$$g(x,y,\lambda,\theta,\varphi,\sigma,\gamma)=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\exp(i\left(2\pi\frac{x'}{\lambda}+\varphi\right)) \tag{19}$$

Where, $x'=x\cos\theta+y\sin\theta$, $y'=-x\sin\theta+y\cos\theta$. x, y denote the pixel coordinate positions, respectively. λ is the wavelength. θ is the direction. σ is the standard deviation of the Gaussian factor. γ is the spatial aspect ratio. φ is the phase offset. At $\varphi=0$, the Gabor function returns the real part, as shown in equation (20), and at $\varphi=\frac{\pi}{2}$, the Gabor function returns the imaginary part, as shown in equation (21):

$$g(x,y;\lambda,\theta,\varphi,\sigma,\gamma)=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \tag{20}$$

$$g(x,y;\lambda,\theta,\varphi,\sigma,\gamma)=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\varphi\right) \tag{21}$$

When the input image $X\in R^{h,w}$ is given, the image X is convolved with the Gabor filter as follows, which is calculated as shown in equation (22):

$$F_{\alpha,\beta}=X\otimes g(x,y,\lambda,\theta,\sigma,\gamma)_{\alpha,\beta}\in R^{\alpha,\beta} \tag{22}$$

where $F_{\alpha,\beta}$ is the convolution result to get the feature map.

4.2.2. Fusion clustering features. Given that it is difficult to accurately determine the parameters of the superpixel segmentation in the traditional Markov segmentation method, and that superpixel segmentation will have the problem of ignoring the information of the small-size defects in the rail surface defect image, these problems may lead to the results of the superpixel segmentation to be affected. In order to overcome the effects of the above problems, the improved Markov defect segmentation method in this paper decides to use the K-means algorithm to perform a preliminary cluster analysis of the pixel matrix. In contrast to hyperpixel segmentation which often only clusters pixel points in the surrounding domain, the K-means algorithm, a widely used clustering method in the field of data processing, determines the optimal category attribution for each data point by calculating the similarity between the data points.

The central goal of the K-means clustering algorithm is to find a set of clustering centers, the number k of which is predetermined by the need. By matching these centers with data samples $x_1, x_2 ... x_n$, the algorithm minimizes the objective function by partitioning the data set into k clusters. This process makes the distance between each data point and its cluster's clustering center $C_1, C_2 ... C_k$ as small as possible, resulting in efficient and accurate data classification.

For K-means clustering of the Gabor convolution processed feature maps, $F_{\alpha,\beta}$ in the previous step, the feature maps, $F_{\alpha,\beta}$ satisfy $F_{\alpha,\beta} = (x_{1,1}, x_{1,2} ... x_{\alpha,\beta})$, the number of cluster centers is specified to be positive k and the objective function is minimized by separating the samples into k clusters, and the objective function of the K-means clustering algorithm is shown in equation (23):

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} (C_i - x)^2 \tag{23}$$

The partial derivatives are obtained for the objective function as shown in Eq. (24) and the simplified result is shown in Eq. (25):

$$\frac{\partial}{\partial C_k} SSE = \frac{\partial}{\partial C_k} \sum_{i=1}^{k} \sum_{x \in C_i} (C_i - x)^2 = 0 \tag{24}$$

$$C_i = \frac{1}{m_i} \sum_{x \in C_i} x \tag{25}$$

where m_i is the number of points in the cluster and C_i denotes the mean vector of points in the group. The model is iterated by updating the mean vectors of the cluster classes, and the model is constructed when the mean vectors no longer change, making the distance between the data and its closest cluster center $C_1, C_2 ... C_k$ as small as possible.

4.2.3. Conditional Iterative Mode Algorithm. When the Markov two-layer graph model contains more variables in the labeled layer S, the computational cost of both the exact a posteriori inference method and the MAP inference is large, and in order to reduce the occupancy of computational power in the computational process of the model inference, the conditional iterative mode algorithm (ICM) can be introduced into the process of the model inference [28].

ICM is a deterministic algorithm based on local conditional probability, and its core idea is to optimize the energy function through iterative updating until it reaches a preset value of the minimum energy function value or the number of iterations. During each iterative update, the ICM algorithm selects the label that is adjacent to the current label and can lead to the greatest improvement in segmentation quality as the new label based on the surrounding pixel point similarity and neighborhood information.

First initialize all nodes based on conditional chaining law and Markov condition there:

$$\begin{aligned} p(x \mid C_i) &= \prod_{i=1}^{N} p(x_{\alpha,\beta} \mid x_{-\alpha,\beta}, y) \\ &= \prod_{i=1}^{N} p(x_i \mid N_{x_i}, C_i) \end{aligned} \tag{26}$$

The conditional probability of labeling layer S approximates the factorization of the conditional probability of $x_{\alpha,\beta}$ such that independent MAP estimation can be performed for all, $x_{\alpha,\beta}$ as shown in Equation (27):

$$x^*_{\alpha,\beta} = \arg\max p(x_{\alpha,\beta} \mid N_{x_{\alpha,\beta}}, C_i) \tag{27}$$

A MAP estimate for labeled layer S can be obtained as shown in equation (28):

$$x^* = \left\{x^*_{1,1}, x^*_{1,2}, \ldots, x^*_{\alpha,\beta}\right\} \tag{28}$$

Update each node value according to the current value of the other nodes, keep iterating until x^* converges, and the part of the label labeled as defective for the final iteration classification is used as the final classification result.

4.2.4. Improved two-layer graph modeling. In the first three sections of this chapter, the improved part of the improved Markov defect segmentation model is introduced respectively, and the improved Markov defect segmentation model can complete the task of rail defect segmentation well on the results of rail surface area localization in Chapter 2, and the algorithmic architecture of the improved Markov defect segmentation model is shown in Fig. 1.

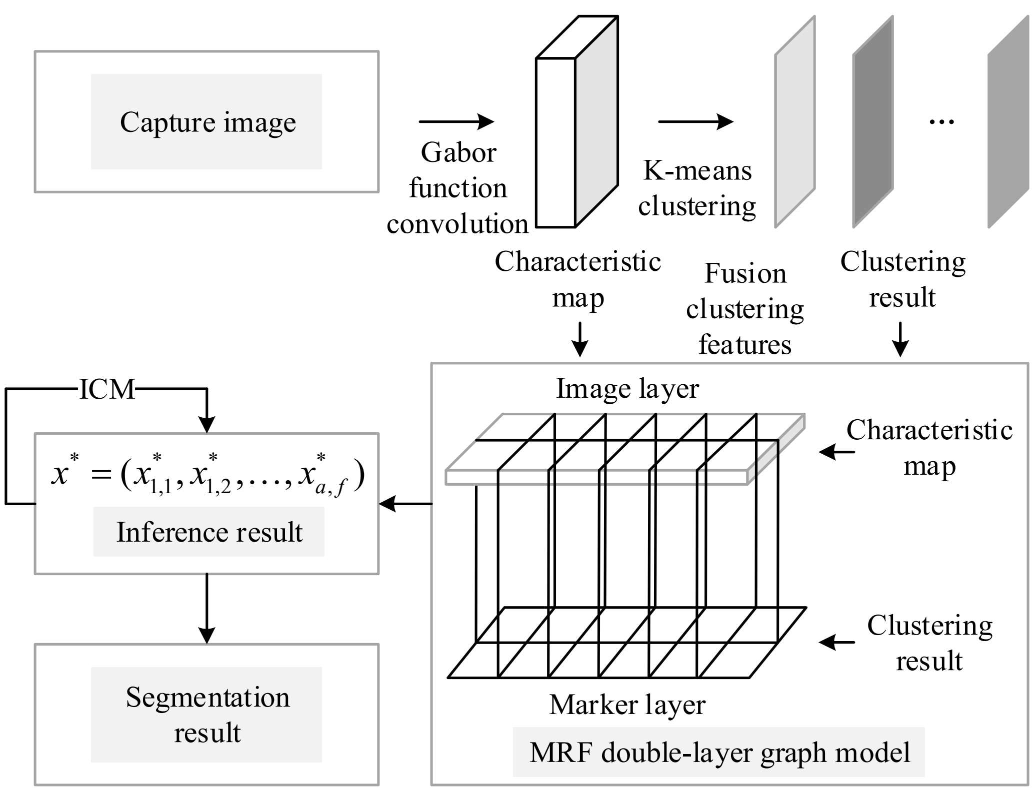


Fig. 1. Improved Markov defect segmentation model algorithm architecture

The process of segmentation of rail surface defects by the improved model is as follows: the rail test dataset after rail surface area localization is processed by Gabor convolution and then converted to grayscale image, and the result is taken as the image observation layer of the bilayer model, and at the same time, the initial clustering is completed by using the K-means clustering algorithm, and the result of the clustering is taken as the image marking layer of the bilayer model, which together form a Markov bilayer model, and then the ICM algorithm is used to perform model inference. Then the ICM algorithm is used for model inference, in the process of model inference, the ICM algorithm updates and iterates the label value in the image labeling layer corresponding to the current pixel in the image observation layer according to the similarity between the current pixel and the surrounding pixels and the neighborhood information until convergence or the maximum number of iterations is reached, and when inference is finished, we get a biclassified labeling matrix and transform the inference result into a binary image, which is the improved Markov defect model. is the result obtained from the improved Markov defect model.

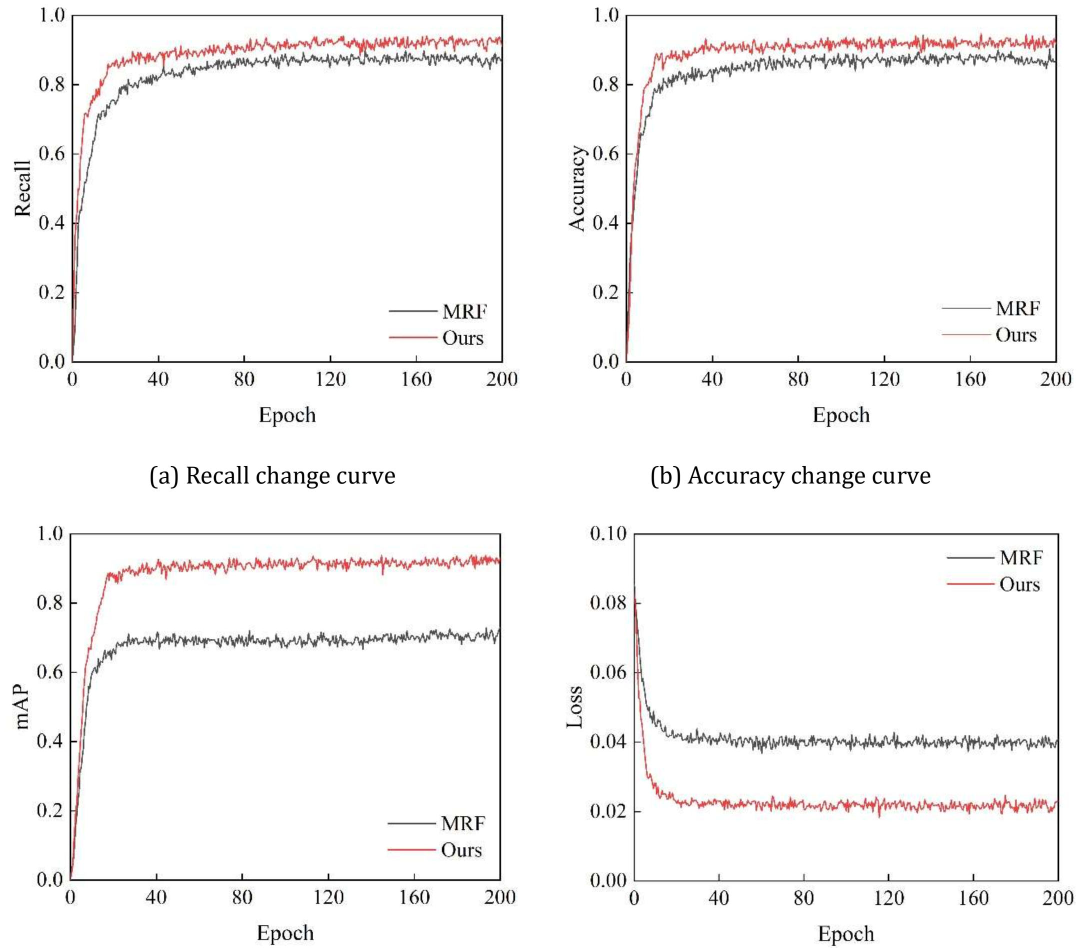
5. Identification and analysis of railroad track defects

5.1. Railway track defect detection

In order to verify the effectiveness of this paper's algorithm, the proposed algorithm is compared with the original Markov defect segmentation algorithm (MRF) in the same railroad defect sample database, and the comparison results are shown in Fig. 2. Figures (a), (b) and (c) show the comparison of model recall and precision and average precision mean before and after the improvement of the target detection algorithm, and Figure (d) shows the loss function comparison.

From Fig. 2(a) it can be seen that MRF reaches about 0.8 recall after 30 rounds of training and 0.85 after 80 rounds, while after algorithmic improvement recall also goes to about 0.89 after 20 rounds of training and 0.91 after 40 rounds. Finally after improvement recall is increased by 6%. MRF in Fig. 2(b) reaches an accuracy of about 0.8 after 20 rounds and stabilizes to 0.86 after 60 rounds. After improvement, the accuracy stabilizes to about 0.9 in 40 rounds, and also stabilizes to about 0.95 after 50 rounds of training, which is an improvement of 8%. After analysis, it can be seen that the improved Markov defect detection model focuses on more neglected semantic information.

In Fig. 2(c), the average precision mean reaches a stable value after training and reaches about 0.95 after 40 rounds, which shows that the mAP value before and after the improvement is obviously much higher, proving that the algorithm can detect the target more accurately after the improvement. After analysis, it can be seen that the MRF downsampling multiplier is larger, and the deeper the feature map is, the less likely it is to learn the feature information of the target, and after the improvement, according to the depth of each feature map, it can capture more useful information, resist unnecessary regional responses, and obtain more complex semantic representations. As can be seen in Fig. 2(d), the MRF reaches a loss value of about 0.04 only after 40 rounds, and eventually stabilizes to about 0.041. With the improved loss function, the initial loss value is around 0.08 and eventually stabilizes to 0.023, and at 8 rounds, the loss value has been reduced to under 0.04, which proves that the improved prediction frame can quickly approximate the real frame, reduce the regression loss of the prediction frame, and improve the regression accuracy.



(a) Recall change curve (b) Accuracy change curve

(c) mAP change curve (d) The loss function iteration process

Fig. 2. The comparison of the primary algorithm and the improved algorithm

In order to verify the effect of each improvement separately, ablation experiments were conducted on the same railroad dataset. Gabor image processing, K-means clustering, and conditional iterative mode are added to the original model in turn. The results of the ablation experiments are shown in Table 1, and the "√" in the table is to add the improvement.

As can be seen from Table 1, when the Gabor function is added to the network, it reduces the impact due to environmental issues and improves the image contrast, which leads to an increase in the mean average accuracy by 1.6%. When K-means is introduced the mean value of average accuracy is increased by 2.8%, and the experimental results show that the K-means clustering method is able to improve the recognition of defective targets in complex environments and enhance the focus on defects. When the ICM method is added the average accuracy mean is improved by 1.8%. Finally after all the improvements are applied, the average accuracy mean can reach 95.7%, but the detection rate is not much different and has good real-time performance.

Table 1. Improved ablation experiment results

Gabor	K-means	ICM	Precision	mAP	Recall	FPS
			0.915	0.893	0.864	46
√			0.929	0.909	0.849	41
	√		0.932	0.921	0.867	30
		√	0.905	0.911	0.865	36

√	√	√	0.925	0.957	0.967	44

5.2. *Identification of railroad track defects*

The defective images of railroad tracks are classified and recognized using the improved Markov defect segmentation model of this paper. Before classification, the defects are first characterized. Rail defects are divided into two categories, crack defects and scar defects, to find out the eigenvalues that can distinguish between the two, the selection of eigenvalues should be differentiated, for example, crack defects present elongated shape, while scar defects present circular or elliptical shape, after analyzing the roundness rate, rectangularity rate, and slenderness ratio as the classification features. To calculate the above three features it is necessary to get the minimum outer rectangle of the defects, and the calculation of roundness rate, rectangularity rate, and slenderness ratio needs to get the geometrical features of the minimum outer rectangle of the cracks and scars.

Before classification first determine the training samples, 30 railroad track crack scar images as training samples, in this paper defect segmentation model input training samples, 3 types of features, 2 types of defects, the number of samples 30, to 1 that the type of defects for the crack, 0 that the type of defects for the scar, the training error is set to 0.01, the training samples feature value information as shown in Table 2.

Table 2. Training samples

Number	Round rate	Rectangular rate	Spindle ratio	Expected output
1	0.235	0.4184	4.2024	0
2	0.7294	0.447	17.1428	1
3	0.682	0.5716	10.5438	1
4	0.1364	0.4428	9.0689	1
5	0.0406	0.3657	6.4377	0
6	0.0414	0.3632	10.0369	1
7	0.2088	0.7358	6.7886	0
8	0.1825	0.4439	17.4309	1
9	0.1749	0.8082	4.6928	0
10	0.3666	0.3306	2.5418	0
11	0.0312	0.4536	13.6438	1
12	0.1259	0.6608	9.1282	1
13	0.6539	0.5243	4.2008	0
14	0.3556	0.5385	9.7597	1
15	0.7534	0.3539	9.7746	1
16	0.3608	0.6086	5.7573	0
17	0.0286	0.5768	3.1455	0
18	0.6487	0.5123	16.4931	1
19	0.5074	0.5941	5.7327	0
20	0.3082	0.7713	11.9807	1
21	0.0066	0.6836	18.3055	1
22	0.3532	0.5162	4.5884	0
23	0.3351	0.3636	14.5547	1
24	0.6138	0.4029	5.1927	0
25	0.4032	0.3797	4.9575	0
26	0.0421	0.7313	8.1271	1
27	0.4312	0.7291	5.8067	0
28	0.3758	0.6358	6.9437	0
29	0.6437	0.3707	11.3411	1
30	0.1804	0.3434	3.8375	0

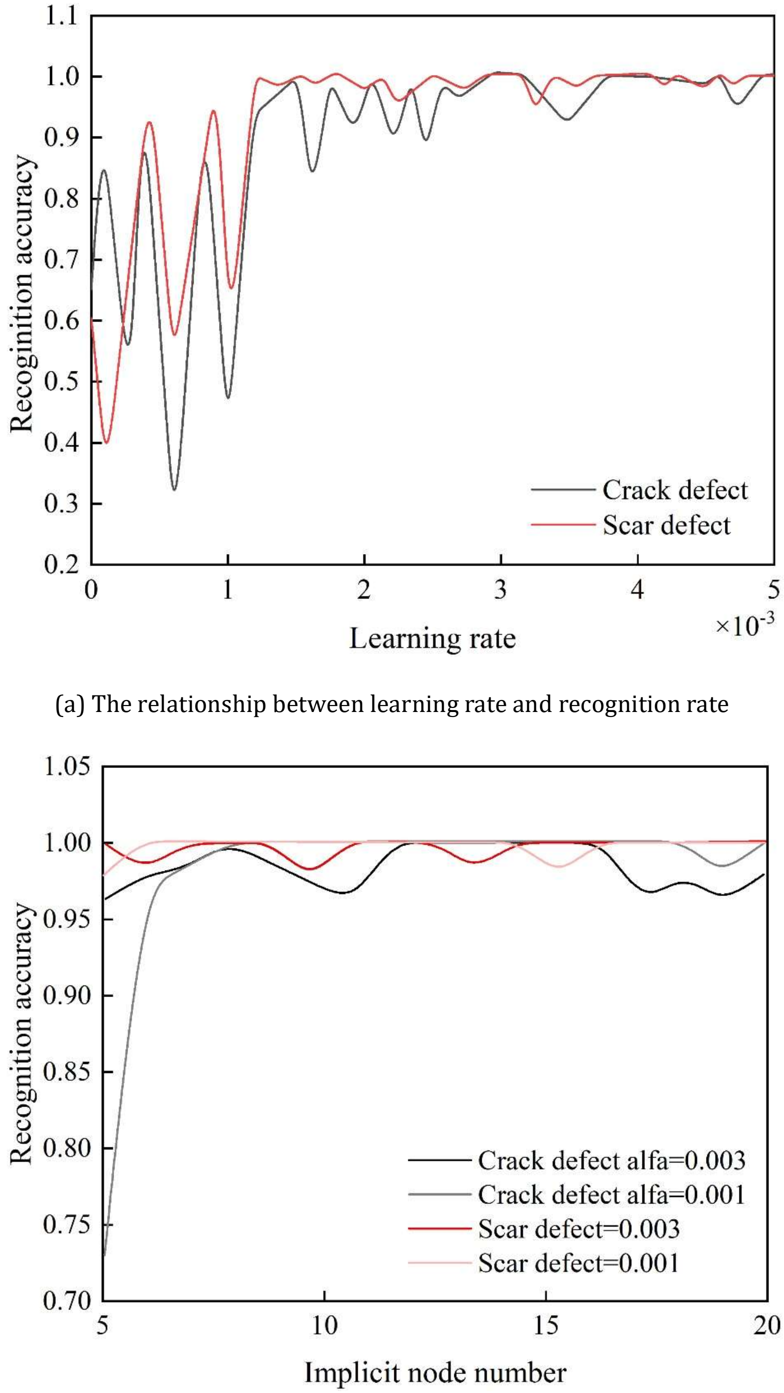
The values in the table are calculated based on the above feature values, by inputting training samples to the defect segmentation model of this paper for network training, at this time the network has reached the training purpose and meets the error of 0.01. Inputting classification samples to the trained network, the classification samples are classified, and the results are shown in Table 3. By fusing the clustering features of the improved Markov defect segmentation model classification can be observed that the actual results are the same as the defect results, defect classification is reliable. This model satisfies the classification requirements.

Table 3. Classification of classified samples

Number	Round rate	Rectangular rate	Spindle ratio	Classification result	Real result
1	0.4146	0.5437	6.6357	0	0
2	0.2502	0.6679	3.1058	0	0
3	0.5629	0.3542	15.728	1	1
4	0.5427	0.6413	13.4146	1	1
5	0.2797	0.2648	14.8631	1	1
6	0.4709	0.3794	2.1663	0	0
7	0.2139	0.6479	12.3613	1	1
8	0.8116	0.6058	3.8476	0	0
9	0.4452	0.5232	4.8542	0	0
10	0.7062	0.5933	9.8543	1	1
11	0.7726	0.6557	4.9348	0	0
12	0.2595	0.5123	9.2174	1	1
13	0.5026	0.3049	9.8566	1	1
14	0.4513	0.6705	6.0049	0	0
15	0.4615	0.6379	13.0772	1	1
16	0.4995	0.4814	2.7345	0	0
17	0.4848	0.4394	7.0152	1	1
18	0.7706	0.4362	5.8459	0	0
19	0.5282	0.5756	15.5923	1	1
20	0.2065	0.6907	9.3813	1	1
21	0.2269	0.7044	2.3001	0	0
22	0.6625	0.6687	11.1322	1	1
23	0.3406	0.5889	3.5199	0	0
24	0.6479	0.3456	4.3995	0	0
25	0.6982	0.7415	10.2429	1	1
26	0.5198	0.3274	4.1266	0	0
27	0.7129	0.7193	7.1538	1	1
28	0.4715	0.3056	7.3942	1	1
29	0.3796	0.3546	4.0378	0	0
30	0.6954	0.6702	4.9723	0	0

In this paper defect segmentation model for railroad track defect classification and recognition, the conditional iterative mode algorithm is used. For this reason, when discussing the relationship between the learning rate, the number of hidden layer nodes and the recognition rate, an initial value is selected for its learning rate, and then the number of its hidden layer nodes is changed, and the relationship between its learning rate, the number of hidden layer nodes and the classification and recognition rate is obtained by training and learning of segmentation for many times, as shown in Figure 3.

From Fig. 3(a), it can be seen that when the number of hidden layer nodes of the model network in this paper is 10, the recognition rate of the network for crack defects and scar defects gradually converges to 1 with the increase of the learning rate, i.e., the recognition rate increases with the increase. From Fig. 3(b), it can be seen that when the initial values of the learning rate are taken as 0.003 and 0.001, respectively, the network has the best classification and recognition ability when the number of implied layer nodes is between 8 and 16, and when the number of implied layer nodes is further increased, its classification and recognition rate instead shows a decreasing trend.



(a) The relationship between learning rate and recognition rate

(b) The relationship between the number of implicit layer nodes and recognition rate

Fig. 3. Relationship between the number of nodes and the recognition rate of the learning rate

6. Conclusion

The author performs region extraction, preprocessing and correction on the railroad track images and improves them by adding Gabor function, K-means clustering and conditional iterative pattern algorithms on the basis of the original Markov random field model in order to improve the image recognition.

Comparing the original Markov random field method with the improved Markov segmentation model, it is found that the recall and precision of the improved model are increased by 6% and 8%

respectively compared to the original model. The improved loss function is stabilized at 0.023, which is lower than the pre-improvement loss function of 0.041. The mean average precision of the original model is improved by 1.6%, 2.8%, and 1.8% after adding the Gabor function, K-means, and ICM alone, respectively. After all the improvements were applied, the mean average accuracy was increased to 95.7%. In the classification of crack defects and scar defects of railroad tracks, the improved Markov defect segmentation model incorporating clustering features can meet the needs of identification and classification.

References

[1] Lu, C., & Cai, C. (2020). Overview on safety management and maintenance of high-speed railway in China. Transportation Geotechnics, 25, 100397.

[2] Wang, W., Cai, K., Du, W., Wu, X., Tong, L. C., Zhu, X., & Cao, X. (2020). Analysis of the Chinese railway system as a complex network. Chaos, Solitons & Fractals, 130, 109408.

[3] Song, H., Song, X., & Dong, H. (2022). Application and evaluation of image-based information acquisition in railway transportation. Journal of Intelligent & Robotic Systems, 106(1), 9.

[4] Ji, A., Woo, W. L., Wong, E. W. L., & Quek, Y. T. (2021). Rail track condition monitoring: A review on deep learning approaches. Intelligence & Robotics, 1(2), 151-175.

[5] Aela, P., Cai, J., Jing, G., & Chi, H. L. (2024). Vision-based monitoring of railway superstructure: A review. Construction and Building Materials, 442, 137385.

[6] Di Summa, M., Griseta, M. E., Mosca, N., Patruno, C., Nitti, M., Renò, V., & Stella, E. (2023). A review on deep learning techniques for railway infrastructure monitoring. IEEE Access, 11, 114638-114661.

[7] Fan, X., Jiao, X., Shuai, M., Qin, Y., & Chen, J. (2023). Application research of image recognition technology based on improved SVM in abnormal monitoring of rail fasteners. Journal of Computational Methods in Sciences and Engineering, 23(3), 1307-1319.

[8] Kapoor, R., Goel, R., & Sharma, A. (2022). An intelligent railway surveillance framework based on recognition of object and railway track using deep learning. Multimedia tools and applications, 81(15), 21083-21109.

[9] Karakose, M., Yaman, O., Murat, K., & Akin, E. (2018). A new approach for condition monitoring and detection of rail components and rail track in railway. International Journal of Computational Intelligence Systems, 11(1), 830-845.

[10] Tastimur, C., Karakose, M., & Akin, E. (2017). Image processing based level crossing detection and foreign objects recognition approach in railways. International Journal of Applied Mathematics Electronics and Computers, (Special Issue-1), 19-23.

[11] Rampriya, R. S., Sabarinathan, & Suganya, R. (2021). RSNet: Rail semantic segmentation network for extracting aerial railroad images. Journal of Intelligent & Fuzzy Systems, 41(2), 4051-4068.

[12] Chandran, P., Asber, J., Thiery, F., Odelius, J., & Rantatalo, M. (2021). An investigation of railway fastener detection using image processing and augmented deep learning. Sustainability, 13(21), 12051.

[13] Liu, S., Wang, Q., & Luo, Y. (2019). A review of applications of visual inspection technology based on image processing in the railway industry. Transportation Safety and Environment, 1(3), 185-204.

[14] Bojarczak, P., & Nowakowski, W. (2021). Application of deep learning networks to segmentation of surface of railway tracks. Sensors, 21(12), 4065.

[15] Dharshan, S., Divya, B., DV, S. R., & Jayakumar, M. (2024, February). Detection of fault in Railway track using Image processing and Fuzzy logic. In 2024 Second International Conference on Emerging Trends in

Information Technology and Engineering (ICETITE) (pp. 1-7). IEEE.

[16] Rampriya, R. S., Jenefa, A., Prathiba, S. B., Julus, L. J., Selvaraj, A. K., & Rodrigues, J. J. (2024). Fault Detection and Semantic Segmentation on Railway Track Using Deep Fusion Model. IEEE Access.

[17] Fuzhi, W., & Changlin, S. (2024). Semantic Image Segmentation Based on Superpixel-pooling DCNN with Random Walks on Graph and Application to Intrusion Detection of Railway. IEEE Access.

[18] Wang, Y., Zhu, L., Yu, Z., & Guo, B. (2019). An adaptive track segmentation algorithm for a railway intrusion detection system. Sensors, 19(11), 2594.

[19] Shang, L., Yang, Q., Wang, J., Li, S., & Lei, W. (2018, February). Detection of rail surface defects based on CNN image recognition and classification. In 2018 20th International Conference on Advanced Communication Technology (ICACT) (pp. 45-51). IEEE.

[20] Wei, X., Wei, D., Suo, D., Jia, L., & Li, Y. (2020). Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model. IEEE Access, 8, 61973-61988.

[21] Gao, H., & He, T. (2023, May). Segmentation of railway region of interest based on image recognition. In Second International Conference on Algorithms, Microchips, and Network Applications (AMNA 2023) (Vol. 12635, pp. 237-241). SPIE.

[22] Wei, D., Wei, X., Tang, Q., Jia, L., Yin, X., & Ji, Y. (2023). RTLSeg: A novel multi-component inspection network for railway track line based on instance segmentation. Engineering Applications of Artificial Intelligence, 119, 105822.

[23] Huiyi Zhu,Zhen Sui,Jianliang Xu & Yeshen Lan. (2024). Fault Diagnosis of Mechanical Rolling Bearings Using a Convolutional Neural Network–Gated Recurrent Unit Method with Envelope Analysis and Adaptive Mean Filtering. Processes,12(12),2845-2845.

[24] Yu Wei,Yongbo Li & Xianzhi Wang. (2025). Incipient bearing fault detection using adaptive fast iterative filtering decomposition and modified Laplace of Gaussian filter. Structural Health Monitoring,24(2),909-924.

[25] Vishwas Rathi,Abhilasha Sharma & Amit Kumar Singh. (2025). Multispectral images reconstruction using median filtering based spectral correlation. Image and Vision Computing,156,105462-105462.

[26] Fang Lei,Shi Zelin,Liu Yunpeng,Li Chenxi,Pang Mingqi & Zhao Enbo. (2023). A general geometric transformation model for line-scan image registration. EURASIP Journal on Advances in Signal Processing,2023(1).

[27] GomezCardona Daniel,Favazza Christopher P,Leng Shuai,Schueler Beth A & Fetterly Kenneth A. (2021). Task-specific efficient channel selection and bias management for Gabor function channelized Hotelling observer model for the assessment of x-ray angiography system performance. Medical physics,48(7),3638-3653.

[28] J. Gimenez,A. Amicarelli,J. M. Toibero,F. di Sciascio & R. Carelli. (2018). Iterated Conditional Modes to Solve Simultaneous Localization and Mapping in Markov Random Fields Context. International Journal of Automation and Computing,15(3),310-324.