

IEICE **TRANSACTIONS**

on Information and Systems

DOI:10.1587/transinf.2024EDP7049

Publicized:2024/09/12

This advance publication article will be replaced by
the finalized version after proofreading.



A PUBLICATION OF THE INFORMATION AND SYSTEMS SOCIETY

The Institute of Electronics, Information and Communication Engineers

Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Multi-Scale Rail Surface Anomaly Detection Based on Weighted Multivariate Gaussian Distribution

Yuyao LIU[†], Qingyong LI[†], *Nonmembers*, Shi BAO^{††}, *Member*, and Wen WANG^{†a)}, *Nonmember*

SUMMARY Rail surface anomaly detection, referring to the process of identifying and localizing abnormal patterns in rail surface images, faces the limitation of robustness because of the large diversity of scale, quantity, and morphology of surface anomalies. To address this challenge, we propose a multi-scale rail surface anomaly detection method (MRS-AD) based on a distribution model, which cooperates neighborhood information to precisely locate rail surface anomalies. Specifically, MRS-AD integrates multi-scale structures to enhance the perception of different scale information of anomalies. Furthermore, the neighborhood information is utilized to capture the correlations between adjacent regions, and thereby a weighted multivariate Gaussian distribution model is estimated to improve the recognition capability of anomalous morphologies. To validate the effectiveness of MRS-AD, we collected and built a Rail Surface Anomaly Detection dataset (RSAD), considering the scale and quantity of rail surface anomalies. Extensive experiments on RSAD, RSDD and NEU-RSDD-2 demonstrate the superiority of MRS-AD. The code and dataset are publicly available at <https://github.com/lyy70/MRS-AD>

key words: anomaly detection, multi-scale structure, distribution model, neighborhood information

1. Introduction

Ensuring the safety and smooth operation of the railway transportation system is imperative, demanding timely inspection and swift rectification of potential faults. Automatic detection of rail surface defects serves as an effective means to achieve this goal. Traditional strategies for rail surface defect detection heavily rely on hand-crafted feature extraction [1–3]. However, constrained by various factors, these approaches are no longer fully adaptable to the demands of modern railway inspection. With the development of artificial intelligence, supervised deep learning has made significant strides in railway detection [4–6]. Nevertheless, these methods often require substantial human effort for defect data screening and labeling, leading to high implementation costs. Recently, many efforts are devoted to unsupervised anomaly detection, which aims to identify abnormal images and locate anomalous regions using a model trained solely on anomaly-free images. Unsupervised anomaly detection has played a crucial role in various domains, including industry [7], network [8], and video surveillance [9]. Therefore,

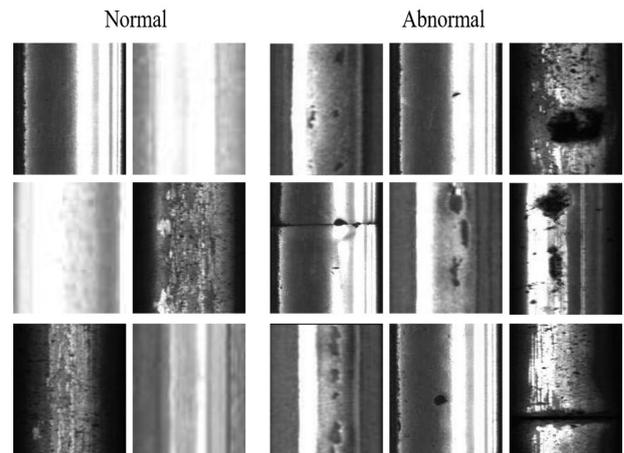


Fig. 1: Sample images from RSAD. The left side represents normal samples, while the right side represents anomalies with significant differences in terms of scale, quantity, and morphology.

in this paper, we also resort to unsupervised anomaly detection to tackle rail surface anomaly detection for seeking practical solutions.

Typical rail surface anomalies include scratches, missing pieces, cracks, etc. These anomalies often exhibit various quantities at different scales, with some regions having dense or sparse distributions of anomalies. Moreover, the diverse morphology of anomalies make them challenging to distinguish anomalies from normal regions, as illustrated in Figure 1. Distribution models [10, 11] are commonly employed to understand the spatial distribution characteristics of normal samples, thereby distinguishing abnormal samples. However, these methods lack the ability to capture the visual morphology of abnormal samples, leading to poor performance under complex scenarios. To tackle this problem, neighborhood information [12, 13] is utilized to typically exhibit a more detailed consideration of the surrounding environment, allowing them to capture subtle variations in anomalies and identify potential anomaly regions more accurately. However, it is crucial to note that these models may face unique challenges in perceiving anomaly scales. Niu et al. [14] proposed a reconstruction network to identify and locate anomalies with significant variations. Nevertheless, this approach still struggles from tackling samples with subtle variations in anomaly appearance.

To comprehensively address the challenges of rail sur-

[†]Key Laboratory of Big Data & Artificial Intelligence in Transportation (Beijing Jiaotong University), Ministry of Education, Beijing, 100044, China.

^{††}The author is with the School of Information Engineering, Inner Mongolia University of Technology, Hohhot, 010080 China.

a) E-mail: Wen Wang@bjtu.edu.cn

face anomalies, an unsupervised Multi-scale Rail Surface Anomaly Detection method (MRS-AD) based on distribution modeling is proposed by exploring coordinated neighborhood information. Specially, MRS-AD integrates a multi-scale structure into the distribution model with the aim of understanding the spatial distribution of anomalies and capturing information differences across different scales. Simultaneously, coordinated neighborhood information is employed to analyze the morphological edges of anomalies, enhancing the localization capability for anomalies. The contributions of this paper are summarized as follows:

- To address the diversity of scale and quantity for surface anomalies on railway tracks, a multi-scale distribution model is designed to comprehensively encompass anomalous features.
- In response to the diverse morphology of surface anomalies on railway tracks, neighborhood information is integrated into the distribution model, thereby further enhancing the accuracy of detection.
- A Rail Surface Anomaly Detection Dataset (RSAD) is released including anomalies with different scale, quantity and morphology.

This paper is organized as follows. Section 2 presents the current research status of anomaly detection in the field of railways. Section 3 gives the reasons and implementation details of the MRS-AD. Section 4 discusses the experimental results and evaluates the performance. Section 5 reports conclusions.

2. Related work

2.1 Anomaly Detection

Unsupervised methods for industrial anomaly detection and localization can be broadly categorized into two types: reconstruction-based methods and embedding-based methods. In the former, the training of neural network is exclusively dedicated to the reconstruction of normal images, making anomalous images easily identified due to poor reconstructions. Thus, anomaly scores are usually expressed using reconstruction errors. Reconstruction-based methods primarily leverage models such as Autoencoders (AEs) [15–21] and Generative Adversarial Networks (GANs) [22–24]. Wang et al. [17] employed VQ-VAE to create a discrete latent space, resampling discrete latent codes deviating from the normal distribution and utilizing the resampled latent codes for image reconstruction. DRÆM [18] introduces a synthetic strategy for simulating anomalies, conducting distinctive training of autoencoder models by augmenting anomaly data. GANomaly [23] trains an adversarial autoencoder to enhance image reconstruction quality, defining the anomaly score as the difference between the latent space representations of the original and reconstructed images.

For the latter, deep neural networks are employed to extract reference vectors that meaningfully describe the entire images. Anomaly scores are typically represented by

the distance between the embedding vector of a test image and the reference vectors in the training dataset. Typical methods utilize networks pre-trained on ImageNet for feature extraction [10–12, 25–28]. PaDiM [10] computes and stores statistical information on each normal feature as preparation for distinguishing anomalies. PatchCore [25] highlights that biased features extracted by pre-trained CNNs can affect precise anomaly localization, hence advocating for the use of intermediate features with smaller biases. CFA [26] addresses the mismatch problem caused by biased features by adopting coupled hyperspheres, thereby enhancing anomaly recognition accuracy.

While existing anomaly detection methods have demonstrated excellent performance, they have not fully exploited the inherent relationships between multi-scale semantic features in the context of rail surface environments. Additionally, their characterization of anomaly morphology is not accurate.

2.2 Rail Surface Defect Detection

Based on the output form of the final defect detection results, methods for rail surface defect detection can be classified into three categories: defect image classification, defect object detection, and defect segmentation. Defect image classification aims to identify defect images in normal samples and output the corresponding defect categories. Niu et al. [14] proposed a robust anomaly detection model (APGVR-GAN) to enhance the capability of classifying anomalies and normal samples. Yang et al. [29] detected abnormal regions of rail objects by using visual symmetry or occurrence frequency. In contrast to defect classification tasks, defect object detection locates defects through bounding boxes. Yu et al. [2] introduced a coarse-to-fine rail defect detection method (CTFM) to identify defects at the sub-image, region, and pixel levels. Defect segmentation, due to its ability to achieve pixel-level defect localization, is more widely applied. Ma et al. [30], based on the sparsity of rail surface defect images, proposed SC-OCDA to address pixel-level distribution alignment between the source and target domains. However, in practical rail surface anomaly detection, anomalies are relatively rare, but their distribution differences are significant, and there is a wide variety of scale and morphological changes.

Considering the complex spatial characteristics exhibited by rail surface anomalies, this work adopts an embedding-based approach for rail surface anomaly detection. To alleviate the mismatch problem caused by Pre-trained networks, MRS-AD utilizes features from specific layers of the Pre-trained network to adapt to rail surface images.

3. Method

The MRS-AD aims to address the complexity of anomaly scale, quantity, and morphology, facilitating accurate classification and localization of anomalies. As illustrated in

Figure 2, in the training phase, a multi-scale distribution model is firstly established based on the characteristics of anomaly scale and quantity. To better capture anomaly morphology, neighborhood information is utilized to form a weighted multivariate Gaussian distribution. Specifically, the target pixel is weighted based on the similarity of the distribution between the target pixel and its neighboring pixels, with similarity measured by the Bhattacharyya distance [12]. Finally, in the inference phase, Mahalanobis distance [10] is employed to express the anomaly score of the test image.

3.1 Multi-Scale Distribution model

Different scales of anomalies refer to microscopic small damages and welds, or macroscopic long-distance cracks and large-area depressions. The uneven distribution of anomaly quantity means that some areas on the rail surface exhibit numerous and dense anomalies, while other areas are relatively clear and sparse. Therefore, adopting a distribution model and integrating a multi-scale structure can not only understand the spatial distribution of normal samples but also effectively capture anomaly context information.

For this purpose, the WideResNet-50 Pre-trained with ImageNet(Frozen CNN) is employed as the backbone network to construct the multi-scale distribution model. Herein, the training set $X_{\text{train}} = \{x \mid y = 0\}$ consists of N_{train} nominal images, and the test set $X_{\text{test}} = \{x \mid y = 0 \text{ or } 1\}$ consists of N_{test} images that are either nominal or anomalous, where $y \in \{0, 1\}$ denotes image x as nominal with 0 and anomalous with 1. The intermediate two layers of feature maps are extracted from the specific hierarchical structure of WideResNet-50, forming the feature set $\phi_j(x)$.

Due to the presence of redundant information, features extracted by the Pre-trained ResNet may weaken the representation of anomaly features [10]. Consequently, we resort to reducing feature dimensions for highlighting the abnormal clues. (see Section 4.4 for details). Simultaneously, to further amplify the differences between anomalous and normal regions and better preserve essential image characteristics, max-pooling is adopted to process the feature set $\phi_j(x)$, and the processed features are denoted as $\{e_i^{(h,w)} \mid i \in \llbracket 1, N \rrbracket\}$.

For all data during the preprocessing stage, which is normalized using given mean and standard deviation values. Consequently, it can be assumed that the hidden layer feature values of the training data approximately follow a Gaussian distribution. Therefore, the mean and covariance are computed at each position (h, w) on the feature map $e_i^{(h,w)}$ and repeated across different feature scales to form a multi-scale distribution model. The definitions of mean and covariance are defined as follows:

$$\mu^{(h,w)} = \frac{1}{N} \sum_{i=1}^N e_i^{(h,w)}, \quad (1)$$

$$\Sigma^{(h,w)} = \frac{1}{N-1} \sum_{i=1}^N \left(e_i^{(h,w)} - \mu^{(h,w)} \right) \left(e_i^{(h,w)} - \mu^{(h,w)} \right)^T + \epsilon_1 I. \quad (2)$$

When computing the covariance matrix, a regularization term $\epsilon_1 I$ is introduced to ensure that the covariance matrix remains full rank and invertible, thereby enhancing the stability and reliability of the estimation [10].

3.2 Weighting of Neighborhood Similarity

Rail surface anomalies exhibit various morphologies, including cracks, depressions, detachments, and pressure blocks, as shown in the anomaly part of Figure 1. Cracks are typically long and thin, depressions and detachments appear as local pits or peeling, while pressure blocks indicate local protrusions or compressions on the rail surface. These diverse anomaly morphologies, combined with the blurred boundaries with normal regions, emphasize the need to effectively capture local features and anomaly patterns by considering coordinated neighborhood information based on a multi-scale distribution model.

To achieve this, the content of the target pixel is enhanced by utilizing information from neighboring pixels. In other words, the distribution of the target pixel is assisted in definition by the similarity in distribution between neighboring pixels and the target pixel. Figure 3 illustrates the process of weighted similarity distribution estimation, involving a specific method for similarity calculation.

Firstly, the neighborhood is defined as the set of p pixels adjacent to the target pixel (h, w) :

$$N_p^{(h,w)} = \{(h', w') \mid h' \in [h - \lfloor p/2 \rfloor, h + \lfloor p/2 \rfloor], w' \in [w - \lfloor p/2 \rfloor, w + \lfloor p/2 \rfloor]\}. \quad (3)$$

Secondly, Bhattacharyya distance [12] is a method used to measure the similarity between two probability distributions. We use it to calculate the similarity between the target pixel (h, w) and all pixels in the neighborhood $N_p^{(h,w)}$. The Bhattacharyya distance is based on the Bhattacharyya coefficient (BC), which measures the overlap between two distributions, and its negative exponential value is considered to represent the similarity of these two distributions. Thus, the similarity m_a defined by the Bhattacharyya distance is defined as follows:

$$m_a = D_{\text{Batt}} \left(\mathcal{N}(e_i^{(h,w)}), \mathcal{N}(N_p^{(h,w)}) \right) = e^{-\frac{BC}{\gamma}}, \quad (4)$$

$$BC \left(\mathcal{N}(\mu_1, \Sigma_1), \mathcal{N}(\mu_2, \Sigma_2) \right) \approx \frac{1}{8} (\mu_1 - \mu_2)^T \Sigma'^{-1} (\mu_1 - \mu_2), \quad (5)$$

where the μ_1 and μ_2 represent the means of the target pixel (h, w) and its neighborhood pixels $N_p^{(h,w)}$, Σ' denotes the average of the covariance Σ_1 of the target pixel and the covariance Σ_2 of the neighborhood pixels. Here, the balancing parameter γ is used to measure the similarity or dissimilarity between the two probability distributions.

Lastly, the distribution of the target pixels is weighted using the similarity m_a . To ensure the stability of the computation results, normalization of the similarity m_a is necessary, as indicated in Eq. 6. Subsequently, the weighted

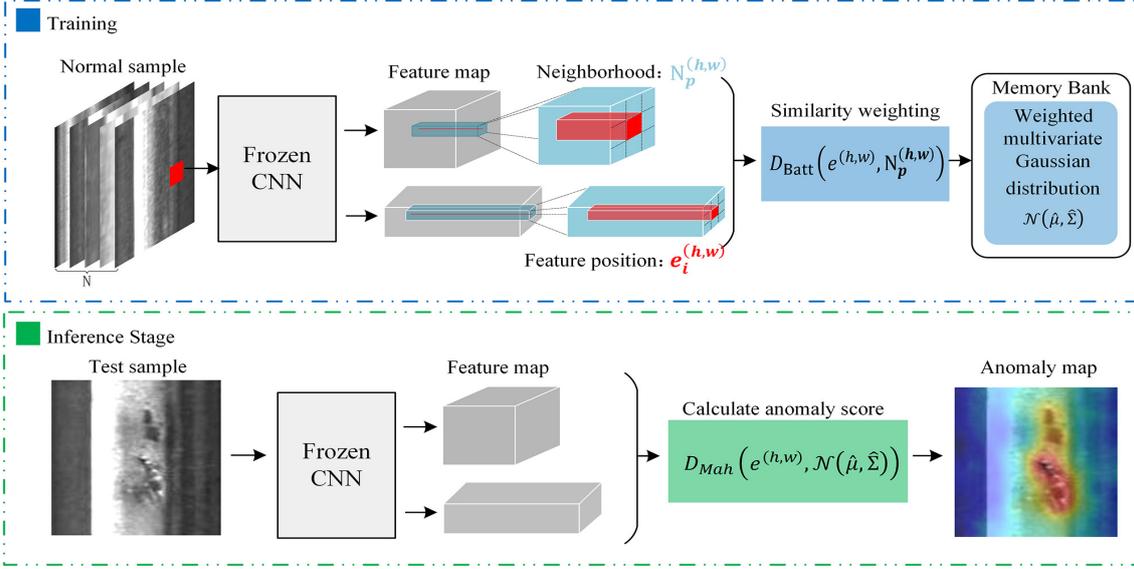


Fig. 2: Overview of the MRS-AD. In training, two-scale distribution models are estimated by applying the similarity between the distribution of the target pixel and its neighboring pixels as weights. In inference, the anomaly score of a pixel (h, w) is computed using the Mahalanobis distance between the target pixel in the test image and nominal distributions.

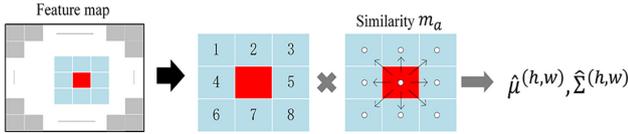


Fig. 3: Estimation of weighted similarity distribution.

mean $\hat{\mu}^{(h,w)}$ and covariance $\hat{\Sigma}^{(h,w)}$ are estimated for each feature position (h, w) to precisely describe the distribution of each pixel in the training images.

$$m'_a = \frac{m_a}{\sum_{a \in N_p^{(h,w)}} m_a}, \quad (6)$$

$$\hat{\mu}^{(h,w)} = \frac{1}{N} \sum_{i=1}^N \sum_{a \in N_p^{(h,w)}} m'_a e_i^a, \quad (7)$$

$$\hat{\Sigma}^{(h,w)} = \frac{1}{N - \sum_{a \in N_p^{(h,w)}} (m'_a)^2} \times \sum_{i=1}^N \sum_{a \in N_p^{(h,w)}} m'_a (e_i^a - \hat{\mu}^{(h,w)}) (e_i^a - \hat{\mu}^{(h,w)})^T + \epsilon_2 I. \quad (8)$$

The formation of the weighted multivariate Gaussian distribution is illustrated in Algorithm 1. This algorithm is designed not only to capture anomalies across multiple scales but also to effectively identify the edges of anomalous shapes.

3.3 Computation of anomaly score

During the inference process, the anomaly score $D_{Mah}^{(h,w)}$ for each position feature $e^{(h,w)}$ on the test image is determined

Algorithm 1: Weighted multivariate Gaussian distribution

Data: Set of training data x , N

Result: weighted: $\hat{\mu}^{(h,w)}$ and $\hat{\Sigma}^{(h,w)}$

```

1 for  $x$  in  $N_{train}$  do
2   Extract feature:  $\phi_j(x) \leftarrow \text{Pre-train CNN}(x)$ ;
3   Pre-processing the extracted features:
4      $e_i^{(h,w)} \leftarrow \text{preprocessing}(\phi_j(x))$ ;
5     Calculating similarity as in Eq.4 and Eq.5:
6      $m_a \leftarrow D_{Batt}\left(\mathcal{N}(e_i^{(h,w)}), \mathcal{N}(N_p^{(h,w)})\right)$ ;
7     Weighted results as in Eq.6, Eq.7 and Eq.8:
8      $\hat{\mu}^{(h,w)}, \hat{\Sigma}^{(h,w)} \leftarrow \mathcal{F}_{Gaussian}(e_i^{(h,w)}, m_a)$ ;
9 end
    
```

by its Mahalanobis distance [10] from the weighted multivariate Gaussian distribution. The feature extraction for each position on the test image aligns with the training process. The definition of the anomaly score $D_{Mah}^{(h,w)}$ is as follows:

$$D_{Mah}^{(h,w)} = \sqrt{(e^{(h,w)} - \hat{\mu}^{(h,w)})^T (\hat{\Sigma}^{(h,w)})^{-1} (e^{(h,w)} - \hat{\mu}^{(h,w)})}, \quad (9)$$

where $\hat{\mu}^{(h,w)}$ and $\hat{\Sigma}^{(h,w)}$ represent the mean and covariance of the weighted multivariate Gaussian distribution. To obtain an anomaly map with the same resolution as the input sample, linear interpolation is performed on Mahalanobis distances at different scales, and the average is taken as the final anomaly score map. Subsequently, Gaussian smoothing with $\sigma = 4$ is applied for further refinement, resulting in pixel-level anomaly scores, and the maximum value is extracted to obtain the image-level anomaly score.

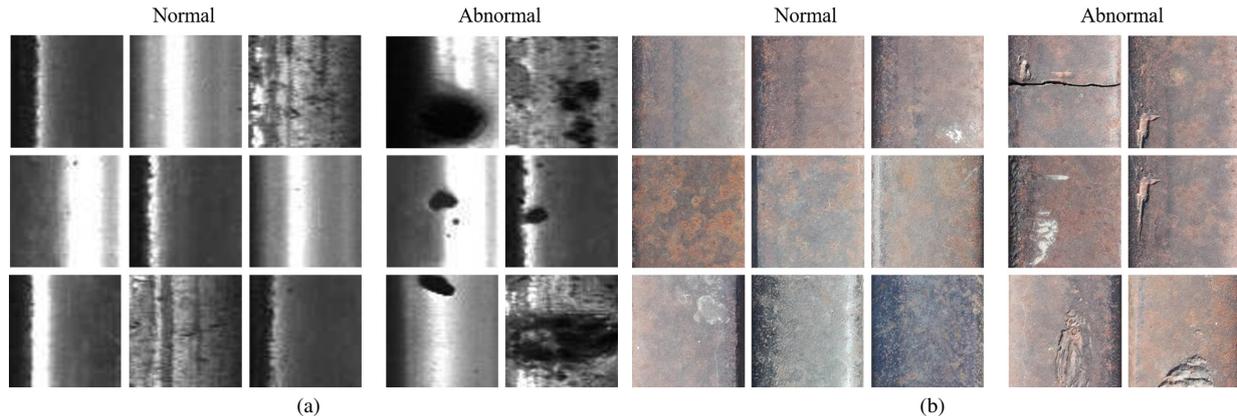


Fig. 4: Partial samples of industrial dataset. (a) RSDD dataset. (b) NEU-RSDD-2 dataset. The left column of each part is the normal images, and the right side is the abnormal images.

4. Experiments

4.1 Dataset

To assess the anomaly detection and localization performance of the MRS-AD, experiments were conducted on three rail surface anomaly datasets. Additionally, some classical anomaly detection algorithms were employed for experimental comparisons.

RSAD dataset. As depicted in Figure 1, the RSAD dataset is derived from the official RSDD dataset and other rail surface images, providing images of rail surface anomalies in different environments. Here, based on the characteristics of rail surface anomaly scales and quantities, anomaly images are categorized into three types of diseases: mild, moderate, severe. To enrich the content of the RSAD dataset, we have mixed these three types of diseases to create a mix disease, with each category containing different normal, abnormal and pixel-level Ground-Truth images provided by professional railway inspectors. The RSAD dataset consists of 3,502 normal images and 1,288 anomalous images for research and evaluation. During the experiments, rail surface images from different environments were uniformly resized to 224 x 224 for consistency with other anomaly detection methods.

RSDD Dataset. As shown in Figure 4(a), this dataset is constructed from high-speed railways and operational heavy-duty railways. In the experiments, it aligns with the dataset provided by Niu et al. [14] and includes 400 normal images and 295 anomaly images without pixel-level annotations. The images were also uniformly resized to 224 x 224 during the experiments.

NEU-RSDD-2 Dataset. As depicted in Figure 4(b), the NEU-RSDD-2 dataset was collected on sections of narrow-gauge railways. In the experiments, it aligns with the dataset provided by Niu et al. [14]. Due to prolonged exposure to outdoor conditions, the steel rail surfaces are influenced by factors such as rain and air, resulting in severe interference

such as rust and mottling. This dataset comprises a total of 699 images, including 415 normal images and 284 anomaly images, with no pixel-level annotations. The image size is 224 x 224.

4.2 Implementation details and Metrics

All experiments were implemented in PyCharm using PyTorch. To maintain the same experimental setup as the comparative anomaly detection methods [10] [25] [26], we employed the WideResNet-50 pre-trained on ImageNet as the backbone network, and executed it on an NVIDIA RTX 3080ti GPU and an Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz. To achieve optimal detection performance, the following settings were used: neighborhood size $p = 3$, balance parameter $\gamma = 0.75$, regularization term $\epsilon_1 = 0.5$, $\epsilon_2 = 0.0015$. The feature map dimension D_2 for $\phi_2(x)$ is 250, while the dimension D_3 for feature map $\phi_3(x)$ is set to 450. The parameters for the ablation study were kept consistent.

In anomaly detection, the Area Under the Receiver Operator Curve (AUROC) [10] is commonly used as an evaluation metric. This study assesses the performance on the RSAD datasets in terms of anomaly classification (Image-wise AUROC) and localization (Pixel-wise AUROC). Additionally, the Pixel-wise F1max score is employed to evaluate the model's pixel-level localization performance.

Consistent with prior work [14], the following metrics were used to evaluate performance on the RSDD dataset: Accuracy (ACC), Precision (PRE), True Positive Rate (TPR), F1 score, and Matthews Correlation Coefficient (MCC, ranging from -1 to 1).

4.3 The experimental results and analysis

Experimental Evaluation on RSAD. The MRS-AD is compared with three previous anomaly detection algorithms: PaDiM [10], PatchCore [25], and CFA [26], assessing their anomaly detection capabilities in rail surface scenarios. Ta-

Table 1: Image / Pixel-wise AUROC(%) of anomaly detection methods on RSAD.

Method	Mild_disease	Moderate_disease	Severe_disease	Mix_disease	Average
PaDiM [10]	98.3 / 99.0	95.7 / 95.6	93.4 / 88.8	96.2 / 94.3	95.9 / 94.4
PatchCore [25]	90.1 / 97.1	96.2 / 95.1	99.0 / 92.3	95.6 / 94.6	95.6 / 94.6
CFA [26]	95.8 / 95.4	97.1 / 93.9	98.9 / 88.8	96.8 / 92.2	96.8 / 92.9
MRS-AD	97.3 / 98.2	97.1 / 95.9	97.6 / 93.1	97.1 / 95.8	97.3 / 95.8

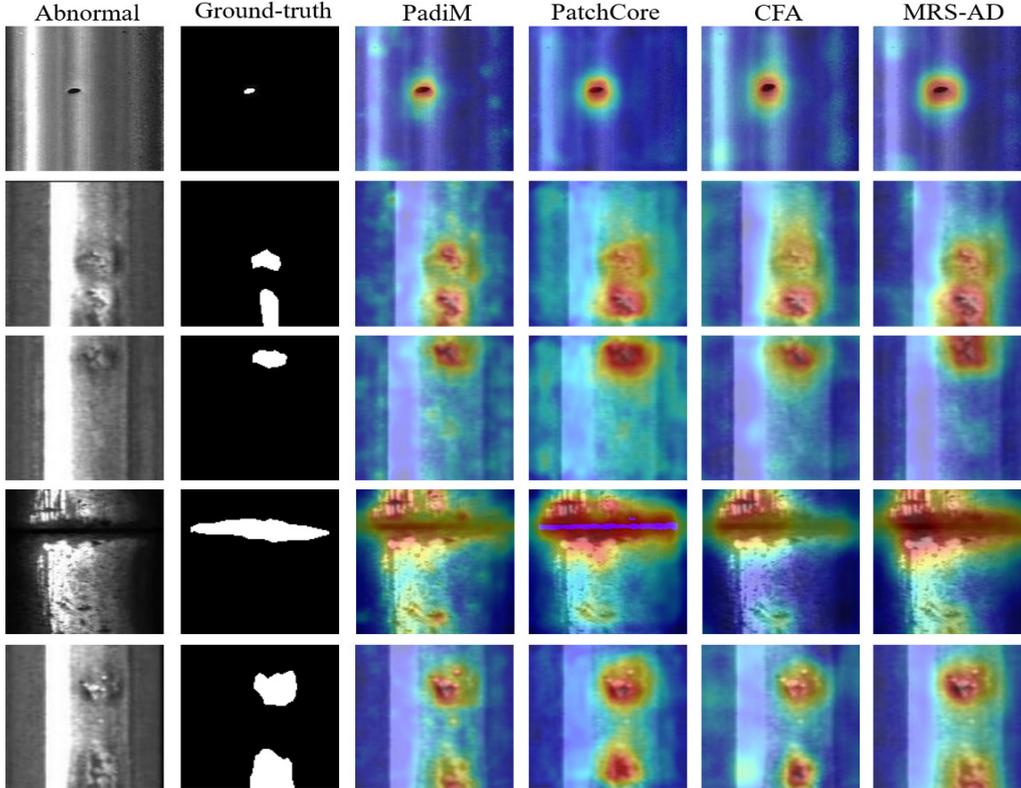


Fig. 5: Visualization results of anomaly detection methods on the RSAD dataset. The first column represents abnormal images, the second column is the ground truth, then displays PaDiM’s heatmap, PatchCore’s heatmap, and CFA’s heatmap, the last column presents the heatmap generated by the MRS-AD.

Table 1 quantitatively compares the detection performance of the MRS-AD and other methods in terms of anomaly classification and localization (Image/Pixel-wise AUROC).

MRS-AD demonstrates excellent average detection results. In the RSAD dataset, mild disease refers to defects of small-scale; moderate disease encompasses fewer instances of Medium-scale defects as well as numerous smaller defects; severe disease involves large-scale defects; and mixed disease includes all three types of defects. MRS-AD utilizes a multi-scale feature structure and integrates neighborhood information to perceive defects across different scales and enhance edge detection accuracy across various defect morphologies, thereby achieving optimal average detection results. Besides, in scenarios involving mild defects, MRS-AD demonstrates comparable performance to PaDiM [10], which uses a smaller receptive field to accurately detect small-scale defects, as shown in Figure 5 of the manuscript.

In cases of severe defects, PatchCore [25] shows a slight advantage by computing the maximum anomaly value of neighbors.

Figure 6 illustrates the performance of the MRS-AD in terms of Pixel-wise F1max. The F1 score provides a comprehensive evaluation of the model’s performance under imbalanced class distributions, balancing precision and recall, thereby offering effective guidance for a holistic assessment of binary classification models. To more intuitively compare the differences among various methods in anomaly localization, the maximum value in the F1 score is selected for comparison. It is worth noting that when it comes to severe and mixed diseases, the stability of PatchCore is slightly higher than that of MRS-AD, and it has better generalization ability.

Experimental Evaluations on RSDD and NEU-RSDD-2. Table 2 reports the detection performance of the

Table 2: The results of image-level evaluation metrics of MRS-AD on the RSDD and NEU-RSDD-2 datasets.

Method	RSDD						NEU-RSDD-2					
	AUC	ACC	PRE	TPR	F1	MCC	AUC	ACC	PRE	TPR	F1	MCC
AnoGAN [22]	0.72	0.60	0.85	0.26	0.41	0.29	0.48	0.29	0.77	0.06	0.11	0.02
GANomaly [23]	0.83	0.74	0.92	0.52	0.69	0.53	0.65	0.35	0.88	0.13	0.24	0.12
VAE [15]	0.76	0.64	0.88	0.33	0.49	0.36	0.61	0.32	0.85	0.10	0.18	0.08
OCGAN [24]	0.66	0.54	0.71	0.12	0.21	0.13	0.64	0.29	0.76	0.06	0.11	0.01
MemAE [20]	0.83	0.73	0.91	0.5	0.67	0.51	0.62	0.37	0.90	0.16	0.28	0.14
DAPE [21]	0.76	0.60	0.83	0.24	0.38	0.27	0.58	0.37	0.91	0.17	0.29	0.15
APGVR-GAN [14]	0.87	0.79	0.92	0.63	0.77	0.60	0.81	0.60	0.96	0.48	0.64	0.39
MRS-AD	0.92	0.86	0.88	0.90	0.86	0.71	0.87	0.83	0.93	0.93	0.89	0.52

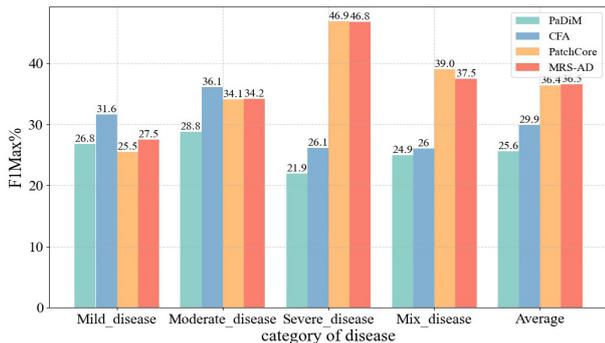


Fig. 6: Performance of Pixel-wise F1max% for anomaly detection algorithms on RSAD.

MRS-AD and the method mentioned by Niu et al. [14] in the same environment. The results indicate that, despite the MRS-AD exhibiting some limitations in predictive accuracy based on specific auxiliary evaluation metrics, these shortcomings are attributed to the reconstruction network’s emphasis on providing a more precise representation of the anomaly region. Nevertheless, it continues to demonstrate commendable classification capabilities, rendering it well-suited for scenarios characterized by imbalances among different categories.

4.4 Ablation study

Evaluation of the effectiveness of key design components.

The effectiveness of each module in the MRS-AD was examined, and ablation experiments were conducted to assess and compare their impact on detection performance.

As depicted in Table 3, individual modules contribute differently to anomaly detection. The multi-scale architecture enables the detection of anomaly at various scales, Pre-processing enhances feature extraction, and the neighborhood information module improves target pixel recognition. This demonstrates the effectiveness of combining weighted similarity inference with both the multi-scale architecture and neighborhood information.

Table 4 investigates the impact of using different scales on detection results in the RSAD dataset, where layer1, layer2, and layer3 represent three specific feature layers extracted from a network Pre-trained on ImageNet.

Table 3: The influence of each key component on mean AUROC in the RSAD dataset.

Baseline	Multi-Scale	Neighbor Information	Image-wise (Average)	Pixel-wise (Average)
✓			94.8	94.3
✓	✓		95.7	95.0
✓		✓	94.9	95.6
✓	✓	✓	97.3	95.8

Table 4: The influence of applying different layers on mean AUROC in the RSAD dataset, Layer1*2*3 denotes the concatenation of three layers of features, Layer1+2+3 indicates the independent processing of three layers of features.

Layer used	Image-wise (Average)	Pixel-wise (Average)
Layer3	94.8	94.3
Layer1*2*3	96.6	95.4
Layer2*3	97.1	95.8
Layer1+2+3	96.6	95.5
Layer2+3	97.3	95.8

Layer1*2*3 denotes the concatenation of three layers of features, where the concatenation process aligns with that employed in PaDiM [10]. Subsequently, the integrated features are used for neighborhood information fusion. Layer1+2+3 indicates the fusion of the three layers independently with neighborhood information. The experimental results indicate that employing a single-layer feature yields less informative content and results in inferior detection performance. Utilizing three layers of features enhances the detection performance for small anomalies, albeit with increased redundancy. However, leveraging the features from the last two layers not only alleviates biases linked to the categories of ImageNet but also adjusts to the dataset, retaining a richer set of semantic information. This adaptation contributes to more effective detection results.

Investigating hyperparameters. Firstly, the impact of different neighborhood sizes p on detection performance was investigated. As shown in Figure 7, the performance reached the optimal level when the neighborhood size was set to 3. With an increase in the neighborhood range, image-level AUROC gradually decreased, while pixel-level AUROC showed

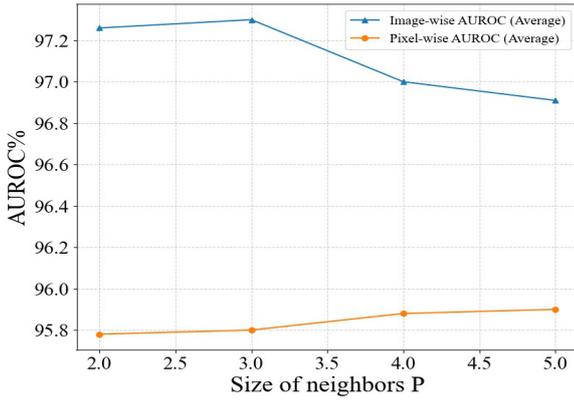


Fig. 7: The influence of applying different neighborhood size on mean AUROC in the RSAD dataset.

Table 5: The influence of applying different dimensionality reduction methods on Image/Pixel-wise AUROC in the RSAD dataset.

Dimension Reduction	Randomly Select [10]	Top-d Select [12]
$D_2 = 350$ $D_3 = 550$	96.7 / 95.6	95.6 / 94.8
$D_2 = 250$ $D_3 = 450$	97.3 / 95.8	95.6 / 94.7
$D_2 = 150$ $D_3 = 350$	96.4 / 95.4	95.4 / 94.3

ROC gradually decreased, while pixel-level AUROC showed a gradual improvement. This indicates that the model might focus more on the overall features of the image, neglecting local details, resulting in the local details of anomalies being subdued. However, at the same time, it captured richer anomaly cues, leading to better localization performance.

Next, the impact of different dimensionality reduction methods and the number of reduced dimensions was investigated. For the distribution model, this study mainly explored two dimensionality reduction approaches: 1) random channel reduction [10], and 2) selecting the top-d dimensions with fewer non-zero values [12]. Table 5 demonstrates the influence of these two dimensionality reduction methods on the detection results. The strategy of random channel reduction is more reasonable. This is because in channels with fewer non-zero values, there might still be some anomalous information that plays a crucial role in the weighted process of neighboring pixels. Therefore, random reduction was applied independently to features $\phi_j(x)$ at different scales. The optimal detection performance is achieved when the feature map dimensions of $\phi_2(x)$ and $\phi_3(x)$ are set to 250, 450.

4.5 Discussion

Despite demonstrating high efficacy in railway surface anomaly detection, MRS-AD still exhibits certain limitations.

Table 6: Inference times of various anomaly detection models on the RSAD dataset.

	PaDiM [10]	PatchCore [25]	CFA [26]	MRS-AD
Inference time(sec.)	0.1274	0.0069	0.0409	0.1325

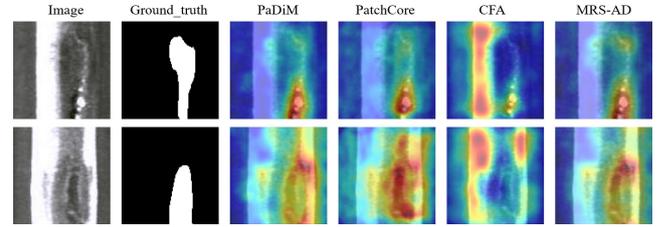


Fig. 8: Instances of detection failure on the RSAD dataset.

Table 6 compares model inference times reported in PaDiM [10] to assess the computational costs across different anomaly detection methods. Due to the computation of neighborhood similarity, the MRS-AD method requires more time for inference but exhibits superior detection performance compared to PaDiM [10].

Figure 8 illustrates instances where MRS-AD fails to detect anomalies in the RSAD dataset. It can be observed that other comparison methods also exhibit poor detection results. Although PatchCore [25] can detect defect areas extensively, it fails to accurately detect defect edges. MRS-AD utilizes a distribution model that integrates neighborhood information. When railway defects take on elongated shapes, they extend beyond the neighborhood's sensitivity range, leading to altered anomaly decisions in adjacent regions and resulting in misclassification. Therefore, dynamically adjusting the neighborhood range is a challenge we aim to address in the future.

5. Conclusion

This paper has proposed a novel model for rail surface anomaly classification and localization, which takes into consideration the scale, morphology, and quantity characteristics of rail anomalies. The model leverages a multi-scale distribution model and collaborates with neighborhood information to better identify and locate anomalies. The experimental results demonstrate that MRS-AD exhibits the best performance in unsupervised railway surface anomaly detection. The model shows robustness in handling mixed disease, validating the effectiveness of this approach. Additionally, a rail surface anomaly dataset is created based on the scale and quantity characteristics of anomalies, providing a more comprehensive validation for detection methods.

Future work will focus on developing anomaly detection solutions tailored to real-world scenarios, with a particular emphasis on incorporating zero/few-shot anomaly detection approaches to address rail surface anomalies.

Acknowledgements

This work was supported in part by the Fundamental Research Funds for the Central Universities (Science and technology leading talent team project, Grant 2022JBQY007), the National Natural Science Foundation of China (Grant 62276019, U2034211, 62006017), the China Postdoctoral Science Foundation (Grant 2022M721826).

References

- [1] J. Gan, Q. Li, J. Wang, and H. Yu, "A hierarchical extractor-based visual rail surface inspection system," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7935–7944, 2017.
- [2] H. Yu, Q. Li, Y. Tan, J. Gan, J. Wang, Y.-a. Geng, and L. Jia, "A coarse-to-fine model for rail surface defect detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 3, pp. 656–666, 2019.
- [3] Q. Luo, X. Fang, J. Su, J. Zhou, B. Zhou, C. Yang, L. Liu, W. Gui, and L. Tian, "Automated visual defect classification for flat steel surface: A survey," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9329–9349, 2020.
- [4] W. Cui, J. Wang, H. Yu, W. F. Peng, L. Wang, S. Wang, P. Dai, and Q. Li, "From digital model to reality application: A domain adaptation method for rail defect detection," in *Chinese Conference on Pattern Recognition and Computer Vision*, 2021.
- [5] X. Ni, Z. Ma, J. Liu, B. Shi, and H. Liu, "Attention network for rail surface defect detection via consistency of intersection-over-union(iou)-guided center-point estimation," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1694–1705, 2022.
- [6] Y. Wu, Y. Qin, Y. Qian, F. Guo, Z. Wang, and L. Jia, "Hybrid deep learning architecture for rail surface segmentation and surface defect detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, pp. 227–244, 2021.
- [7] R. Saiku, J. Sato, T. Yamada, and K. Ito, "Enhancing anomaly detection performance and acceleration," *IEEJ Journal of Industry Applications*, vol. 11, no. 4, pp. 616–622, 2022.
- [8] H. TIAN, K. GUO, X. GUAN, and Z. WU, "Anomaly detection of network traffic based on intuitionistic fuzzy set ensemble," *IEICE Transactions on Communications*, vol. E106.B, no. 7, pp. 538–546, 2023.
- [9] W. SHAO, R. KAWAKAMI, and T. NAEMURA, "Anomaly detection using spatio-temporal context learned by video clip sorting," *IEICE Transactions on Information and Systems*, vol. E105.D, no. 5, pp. 1094–1102, 2022.
- [10] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: a patch distribution modeling framework for anomaly detection and localization," in *International Conference on Pattern Recognition*. Springer, 2021, pp. 475–489.
- [11] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "Cutpaste: Self-supervised learning for anomaly detection and localization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9664–9674.
- [12] J. Jang, E. Hwang, and S.-H. Park, "N-pad: Neighboring pixel-based industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4364–4373.
- [13] J. Bae, J. Lee, and S. Kim, "Pni: Industrial anomaly detection using position and neighborhood information," 2022.
- [14] M. Niu, Y. Wang, K. Song, Q. Wang, Y. Zhao, and Y. Yan, "An adaptive pyramid graph and variation residual-based anomaly detection network for rail surface defects," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [15] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations*, 2013.
- [16] V. Zavrtnik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognition*, vol. 112, p. 107706, 2021.
- [17] L. Wang, D. Zhang, J. yi Guo, and Y. Han, "Image anomaly detection using normal data only by latent space resampling," *Applied Sciences*, 2020.
- [18] V. Zavrtnik, M. Kristan, and D. Skočaj, "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8330–8339.
- [19] K. SATO, S. NAKATA, T. MATSUBARA, and K. UEHARA, "Few-shot anomaly detection using deep generative models for grouped data," *IEICE Transactions on Information and Systems*, vol. E105.D, no. 2, pp. 436–440, 2022.
- [20] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. Van Den Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1705–1714.
- [21] J. Liu, K. Song, M. Feng, Y. Yan, Z. Tu, and L. Zhu, "Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection," *Optics and Lasers in Engineering*, vol. 136, p. 106324, 2021.
- [22] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 146–157.
- [23] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*. Springer, 2019, pp. 622–637.
- [24] P. Perera, R. Nallapati, and B. Xiang, "Ocgan: One-class novelty detection using gans with constrained latent representations," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2893–2901.
- [25] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14 318–14 328.
- [26] S. Lee, S. Lee, and B. C. Song, "Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization," *IEEE Access*, vol. 10, pp. 78 446–78 454, 2022.
- [27] J. Yi and S. Yoon, "Patch svdd: Patch-level svdd for anomaly detection and segmentation," in *Proceedings of the Asian conference on computer vision*, 2020.
- [28] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "Simplenet: A simple network for image anomaly detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20 402–20 411.
- [29] T. Yang, Y. Liu, Y. Huang, J. Liu, and S. Wang, "Symmetry-driven unsupervised abnormal object detection for railway inspection," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 12, pp. 11 487–11 498, 2023.
- [30] S. Ma, K. Song, M. Niu, H. Tian, Y. Wang, and Y. Yan, "Shape-consistent one-shot unsupervised domain adaptation for rail surface defect segmentation," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 9, pp. 9667–9679, 2023.



Yuyao Liu Yuyao Liu received the M.E. degree in science from Inner Mongolia University of Technology, China. He is currently pursuing his Ph.D. degree with the School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China. His research interests include computer vision and image processing.



Wen Wang Wen Wang received Ph.D. degree in computer science and technology from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in June 2017. Currently, she is a lecturer at Beijing Jiaotong University, Beijing. Her research interests mainly include computer vision, pattern recognition and machine learning.



Qingyong Li Qingyong Li received the Ph.D. degree in computer science and technology from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2006. He is currently a professor with Beijing Jiaotong University, Beijing. His research interests include computer vision and artificial intelligence.



Shi Bao received the B.E. degree in science from Qinghai University, China, in 2005, and the M.E. and Ph.D. degrees in science from Nagoya City University, Japan, in 2011 and 2015, respectively. He is currently with the School of Information Engineering, Inner Mongolia University of Technology, China, where he is an assistant professor. He is interested in image processing.