DM-RIS: Deep Multimodel Rail Inspection System With Improved MRF-GMM and CNN

Xiating Jin, Yaonan Wang, Hui Zhang, Hang Zhong, Li Liu, Q. M. Jonathan Wu, *Senior Member, IEEE*, and Yimin Yang, *Member, IEEE*

***Abstract*— Rail inspection system (RIS) remains an emergent instrumentation for railway transportation, with its capacity of measuring surface defect on steel rail. However, detecting technique and interpretation of RIS constitute a challenging problem since traditional technologies are expensive and prone to errors. In this paper, a deep multimodel RIS (DM-RIS) is established for surface defect where fast and robust spatially constrained Gaussian mixture model is presented for segmentation proposal and Faster RCNN is utilized for objective location in a parallel structure. First, we incorporate spatial information between pixels into an improved Gaussian mixture model based on Markov random field (MRF) for accurate and rapid defect edge segmentation. Specifically, a direct parameter-learning in expectation–maximization (EM) algorithm is proposed. Meanwhile, to remove nondefect, numerous labeled samples with weak illumination, inequality reflection, external noise, rust, and greasy dirt are fed into Faster RCNN so that DM-RIS is robust environmentally to various light, angle, background, and acquisition equipment. Finally, the joint hit area refers to a real defect. The experimental results demonstrate that the proposed method performs well with 96.74% precision, 94.13% recall, 95.18% overlap, and 0.485 s/frame speed on average, and is robust compared with the related well-established approaches.**

***Index Terms*— Faster RCNN, improved Gaussian mixture model (GMM), Markov random field (MRF), rail inspection, surface defect, visual detection.**

Manuscript received October 28, 2018; revised March 19, 2019; accepted March 21, 2019. Date of publication April 16, 2019; date of current version March 10, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61601061, Grant 61841103, and Grant 81401490, in part by the National Key Research and Development Program of China under Grant 2018YFB1308200, in part by the Scientific Research Fund of Hunan Provincial Education Department under Grant 17C0046, in part by the Hunan Key Laboratory of Intelligent Robot Technology in Electronic Manufacturing under Grant IRT2018009, and in part by the Hunan Key Project of Research and Development Plan under Grant 2018GK2022 and Grant 2018JJ3079. The Associate Editor coordinating the review process was Huang-Chen Lee. *(Corresponding authors: Hui Zhang; Xiating Jin.)*

X. Jin, Y. Wang, H. Zhong, and L. Liu are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: xtchin@hnu.edu.cn; yaonan@hnu.edu.cn; zhonghang@hnu.edu.cn; liuli@hnu.edu.cn).

H. Zhang is with the College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410012, China, and also with the CVSS Laboratory, Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: zhanghuihby@126.com; zhanghui@uwindsor.ca).

Q. M. J. Wu is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B3P4, Canada (e-mail: jwu@uwindsor.ca).

Y. Yang is with the Department of Computer Science, Lakehead University, Thunder Bay, ON P7B5E1, Canada (e-mail: yyang48@lakeheadu.ca).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2019.2909940

I. Introduction

RAIL inspection counts a great deal for railway transportation while several traditional and contact measurement techniques are still applied for rail damage monitoring, resulting in severe misjudgment, omission, and low speed. The existing technologies are facing lots of challenging tasks, including early detection, reliability, and system cost. Recently, some nondestructive testing (NDT) methods, such as an eddy current method [1], acoustic emission (AE) method [2], and ultrasonic methods [3]–[5], have been developed to detect failures in the rail steel. A closed loop of eddy current pulsed thermography (ECPT) [1] enables to identify and measure rail flaw with preprocessing and quantitative assessment. However, inspection of the subsurface defect cannot be taken into account and multiple types of defects fail to display directly. Zhang *et al.* [2] present an improved AE rail defect detection method by multilevel adaptive noise canceling (ANC) with variable step-size least mean square, where the features of noise signals and defect signals are analyzed for effective detection. Furthermore, structural health monitoring (SHM) techniques by measuring full-field strain, structural deformation with ultrasonic piezoelectric transducer [3], and geometry profiles with 3-D digital image correlation (DIC) system [4] would detect potential hazards at an early stage meanwhile providing an efficient way to extend the operational life of railway structures, while not interfering with train operations. Mariani *et al.* [5] work on the development of noncontact (laser) rail inspection system (RIS), making it less sensitive to surface conditions with ultrasonic waves probing. In addition, Molodova *et al.* [6] present a new measuring method using axle box acceleration (ABA) to inspect squats. Similar work is published by Salvador *et al.* [7] on Metropolitan Rail Network of Valencia (Spain) to monitor track defects, singularities, and vibration modes.

Visual approaches [8], [9] play critical roles in rail component detection such as fastener [10], [11], train chassis [12], tie and tie plate [13], [14], and anchor [15]. An automatic visual detection system [10] is provided for partially worn and completely missing fastener. In particular, a novel probabilistic structure topic model (STM) is introduced to model fastener with different orientations and illumination conditions using unlabeled data. A fused approach [11] combining principal component analysis (PCA) with depth peeks (DPs) is utilized to analyze the depth image of hexagonal headed fastener. Lu *et al.* [12] propose time-scale normalization and image subtraction approach for multiple fault detection on China

0018-9456 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

Railway High-speed. The detection algorithm [13], [14] relies on signature tunes of the squats, which were identified from numerical simulations and validated by field measurements. Li *et al.* [15] first develop a set of image and video analytics and then integrate evidence from multiple cameras, global positioning system (GPS), and distance measurement instrument (DMI) in a global framework to monitor ties, tie plates, and anchors. The methods above seem to be encouraging; however, problems like heavy shadow and light overexposure cannot be addressed with ease.

Distinguished from component detection, more interest has grown in noise removal. Resendiz *et al.* [16] combine Gabor filters with spectral estimation to convert a filtered image into 1-D signals. Specifically, multiple signal classification (MUSIC) algorithm is proposed to extract periodic signals from the 1-D signal so component, turnout, and tie on track can be identified in the presence of noise. Several most relevant researches for rail surface maintenance are reported in [17]–[20]. Specifically, to extract the weak signals from strong background noises, a combined technique [20] of ANC and time frequency is provided for condition monitoring and fault diagnosis of railway wheel flat and rail surface defect. Other image processing methods are utilized to eliminate interference from illumination variation. In our initial work [21], we present a novel inverse Perona–Malik (P-M) diffusion model for rail surface enhancement and noise removal. In addition, rail image enhancement is also applied to detect headcheck, undulation, scour, and fracture defects [22]. In summary, many techniques are proposed to suppress noise and enhance the image for rail surface defect detection. However, a pixelwise gray feature is taken advantage of in a simple way, and moreover, rail image can be extremely complex with much high frequency, including noisy diffusion, weak illumination, and rust and greasy dirt, making it difficult to identify defect and extract edge.

At present, numerous updated research studies label numerous objectives like fastener, loose strand, split pins (SPs), surface squat, and crack for a trained classifier and neural network. Gibert *et al.* [23], [24] propose multiple detectors within a multitask deep learning framework to inspect ties and fasteners on a railway track. Faghih-Roohi *et al.* [25] present deep convolutional neural network (DCNN) of different architectures characterized by different sizes and activation functions for defect detection on steel rail. Chen *et al.* [26] cascade three DCNN-based detection stages in a coarse-to-fine manner, including two detectors to sequentially localize the cantilever joints and their fasteners and a classifier to diagnose the fasteners' defects. Liu *et al.* [27] use Faster RCNN-based image location method improved by a feature extraction network named ILNET, then apply a Markov random field (MRF)-based image segmentation method iteration condition model/maximization of the posterior marginal in an automatic fault diagnosis system for the detection of isoelectric line loose strands. Masci *et al.* [28] propose a Max-pooling CNN for steel defect classification on raw segmented images, which avoids time-consuming preprocessing. Zhong *et al.* [29] present a three-stage automatic defect inspection system for SPs mainly based on an improved DCNN, which is called PVANET++. In [30], CNNs are trained on a database of photometric stereo images of metal surface defects, i.e., rail defects. In addition, due to CNNs outstanding performance, several recent studies utilize it for defect detection including road cracks [31], concrete cracks [32], and cracks on nuclear power plant components [33], [34].

Rail surface defects almost come from wheel–rail contact stresses such as extrusion, impact, and abrasion as well as material degradation [35], [36]. The most common surface defects are squat, spalling, joint (broken rail), flaking, rolling contact fatigue crack, interior nuclear damage, crush, side wear, peel, corrosion, weld, and so on [37]. For instance, squat is a type of metal fatigue resulting from wheel impacts on the railhead, which "bruise" the steel and over time lead to degradation of the track. The cluster of base metal falling off railhead in series refers to spalling. Some hidden damage would further deteriorate metal and lead to rail joint if not controlled. Hence, periodical rail inspection is essential. Recently, a four-stage detection method based on curvature filter and improved Gaussian mixture model (GMM) is proposed in our previous work [38]. However, there are some problems and great challenges for RIS as follows.

1) RIS is supposed to be universal or robust environmentally. Steel rail is always exposed to the severe environment, making it low contrast by weak illumination, staggered bright and dark by reflection inequality, and noisy diffusion by rust and greasy dirt. In [38], quantitative analysis was only based on the limited and quite simple rail samples, without any big data analysis-based complex and diverse rail images.
2) RIS is in demand to obtain the edge of the defect with high accuracy. Although the method in [38] is more effective compared with other methods, many errors appear like tiny fuzzy defects that do not need to be inspected in project, rust and greasy dirt that are incorrectly segmented, and noisy background, which affect the judgment of real defect.
3) RIS ought to be real-time and highly efficient. In [38], the step-by-step image preprocessing containing region of interest (ROI) extraction, gray equalization, and curvature filter is much time- and memory-consuming, which affects the detection speed of the entire project.

To address these problems, a deep multimodel RIS (DM-RIS) is proposed and we summary main contributions in this paper as follows.

1) An end-to-end parallel architecture with multiple models is established for rail inspection.
2) DCNN framework achieves considerable defect detection speed and environmental robustness.
3) Fast and robust spatially constrained Gaussian mixture model (FRGMM) with direct parameter-learning is proposed for fast, robust, and precise segmentation.
4) Big data analysis is conducted based on more complex and diverse rail samples.

This paper is organized as follows: In Section II, imaging equipment and rail surface data set are explained.

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

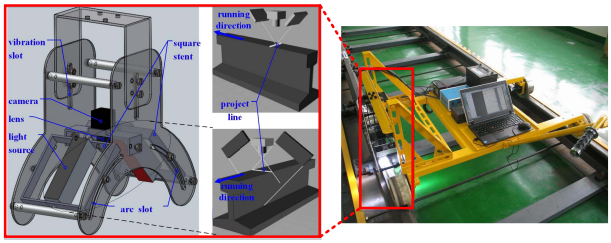

Fig. 1. 3-D illustration and physical map of imaging equipment.

In Section III, defect segmentation and detection algorithms are described briefly. Experimental results and quantitative analysis are provided in Section IV. In Section V, a reasonable discussion makes sense. Finally, the conclusion is given in Section VI.

II. System Overview

A. System Configuration

At the project site, the imaging quality of the rail scene is affected by many factors, such as illumination, exposure degree, incidence angle, camera motion, line speed, and vibration. For this purpose, an imaging mechanism is proposed for image acquisition. As shown in Fig. 1, an imaging system is equipped with a railway track inspection vehicle. One of the core components is a DalsaSpyder3 line scan camera, which can capture a series of consecutive video frames with 1024 pixels resolution and 68 000 lines/s maximum row speed. Note that image distortion would occur if camera velocity and line rate were mismatched. For synchronous data acquisition, a wheel encoder is utilized to trigger line scan camera, arc groove is applied to adjust the angle of incidence, and illumination is affected by a light controller as well as exposure degree. Panorama is generated based on a set of frames. Vibration slot is applied to reduce vibration with damping. In addition, rail surface images are inevitably affected by natural light, rust, and greasy dirt, bringing a great challenge to defect inspection. Also, the image is detected in real time by the processor.

B. Rail Surface Data Set

One concern of this paper is the scene segmentation on the complex and diverse rail surface. Thus, we sort out a rail data set of 6 categories, 450 samples for each category, and totally 2700 samples. All samples are RGB images of 250×160 and are acquired in Hunan province railways and [21], [25], and [38]. Due to the physical structure of rail, all images are staggered bright and dark by inequality reflection, which can be divided into two parts, namely, the simple and the complex. As shown in Table I, for a simple scene, C1 is smooth and clean rail surface, and C2 is C1 corrupted by Gaussian noise (smooth, noisy). For a complex scene, C3 is rough rail surface with rust and greasy dirt, and C4 is C3 corrupted by Gaussian noise (rough, noisy), and C5 is C3 corrupted by salt pepper noise (rough, noisy), and C6 is very dark due to weak illumination.

C. Framework of Proposed Detection Approach

As shown in Fig. 2, aiming to complex and diverse rail surface, a DM-RIS is proposed where a probabilistic graph model FRGMM generates the segmentation proposal and deep learning model Faster RCNN carries out the objective location. On the upward path, FRGMM performs on global image to achieve presegmentation, which is more accurate, robust, and faster due to a simpler MRF distribution and an efficient factor G_{ij} proposed to incorporate spatial information between neighboring pixels, so that the M-step of EM algorithm could be directly applied to prior distribution for the maximization of log-likelihood function. However, many errors occur like tiny fuzzy defect, rust and greasy dirt, and noisy background, which affect the judgment of real defect. On the downward path, to address this problem, Faster RCNN predicts the physical position of the defect in a bounding box. To make it universal and robust to various rail scenes, in-depth convolution features of the full image are shared for region proposal and regression correction, making detection much fast and accurate, without any repeated computation. Finally, the joint hit area refers to real defect with high accuracy and precision.

TABLE I
Rail Surface Data Set

Mnemonic	Scene	Characteristic
C1	Clean rail surface	smooth
C2	C1 + external noise	smooth, noisy
C3	Rusty rail surface	rough
C4	C3 + Gaussian noise	rough, noisy
C5	C3 + Salt pepper noise	rough, noisy
C6	Dark rail surface	dark, low contrast

III. Defects Detection Method

A. Segmentation Model FRGMM

An improved MRF-based GMM is proposed to segment defects more quickly, accurately, and robustly. In comparison to other mixture models with high computational expense, we incorporate spatial information between neighboring pixels, so that the M-step of expectation–maximization (EM) algorithm can be directly applied to the prior distribution. Hence, a new method is a FRGMM, which can be written as FRGMM. The following will be introduced into three parts.

1) Gaussian Mixture Model: A background modeling method based on pixel sample statistics, which is also the basis of the new method.

Let $x_i, i = (1, 2, \ldots, N)$, represents the ith pixel of a D-dimension image and ∂_i represents the neighborhood of x_i. Then, Ω_j, $j = (1, 2, \ldots, K)$, represents the jth label of image. To separate the image from N pixels into K labels, standard GMM [39], [40] supposes that each pixel x_i is independent of label Ω_j. For this, the density function corresponding to a pixel x_i is defined as

$$f(x_i|\Pi, \Theta) = \sum_{j=1}^{K} \pi_{ij} \phi(x_i|\Theta_j) \tag{1}$$

where $\Pi = \{\pi_{ij}\}, i = (1, 2, \ldots, N), j = (1, 2, \ldots, K)$, is the prior distribution set modeled from the probability of pixel x_i

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

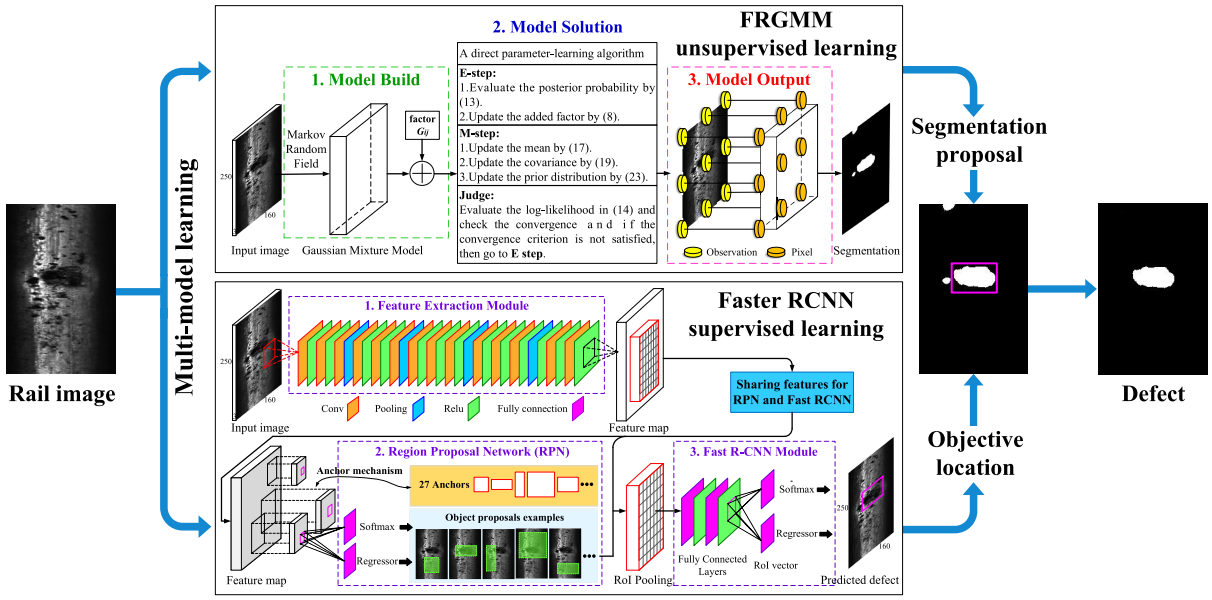

Fig. 2. Pipeline of rail surface defect detection approach.

being in label Ω_j, with a constraint condition as follows:

$$0 \le \pi_{ij} \le 1 \text{ and } \sum_{j=1}^{K} \pi_{ij} = 1 \tag{2}$$

and $\phi(x_i|\Theta_j)$ is the Gaussian distribution named a component of mixture model and specifically it denotes as

$$\phi(x_i|\Theta_j) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_j|^{1/2}} \times \exp\left\{-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j)\right\} \tag{3}$$

where $\Theta_j = \{\mu_j, \Sigma_j\}$, $j = (1, 2, \ldots, K)$, with the mean μ_j of D-dimension vector, the covariance Σ_j of $D \times D$ matrix, and $|\Sigma_j|$ is the determinant of Σ_j.

According to the assumption of GMM, pixel x_i is modeled independent in statistics, thus the joint density function of data set $X = (x_1, x_2, \ldots, x_N)$ denotes as

$$p(X|\Pi, \Theta) = \prod_{i=1}^{N} f(x_i|\Pi, \Theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_{ij} \phi(x_i|\Theta_j). \tag{4}$$

A major drawback is that pixel x_i is regarded as independent in the label so that the spatial consistency between adjacent pixels cannot be taken into account, making the standard GMM much sensitive to noise and illumination [40], which is not robust for complex rail surface segmentation.

2) Mixture Model Based on MRF: To improve 1), the introduced MRF distribution incorporates local information, but increases the difficulty of model solution.

MRF distribution [41] is utilized in such form

$$p(\Pi) = Z^{-1} \exp\left\{-\frac{1}{T} U(\Pi)\right\} \tag{5}$$

where $U(\Pi)$ represents the energy named as smoothing prior, Z and T represent normal constant and temperature constant, respectively. Having Bayessian rules, the posterior probability density function satisfies the following relations:

$$p(\Pi, \Theta|X) \propto p(X|\Pi, \Theta) p(\Pi). \tag{6}$$

Combining (5) with (6), log-likelihood function is given as

$$\begin{aligned} L(\Pi, \Theta|X) &= \log(p(\Pi, \Theta|X)) \\ &= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij} \phi(x_i|\Theta_j)\right\} + \log p(\Pi) \\ &= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{K} \pi_{ij} \phi(x_i|\Theta_j)\right\} - \log Z - \frac{1}{T} U(\Pi). \end{aligned} \tag{7}$$

Taking a close look at (7), the attribute of a model almost depends on the structure of energy $U(\Pi)$. For instance, in an MRF-based method [41], spatial information is taken into account with four considered directions, namely, horizontal, vertical, and two diagonal. In another Bayesian logic [42], two indexes Z and T in smoothing prior $U(\Pi)$ are set to 1 for image segmentation. For more details regarding the energy $U(\Pi)$ and mixture models, we refer the readers to [40] and [43]–[45].

As can be seen above, different structures of smoothing priors $U(\Pi)$ and MRF distribution increase the complexity of the model, although they are effective to integrate correlation between neighboring pixels. For this purpose, an iterative EM algorithm is utilized to maximize the probability of parameters Π and Θ in (7). However, the M-step of EM algorithm fails to solve the prior distribution π_{ij} directly, due to likelihood function with high computational expense. Moreover, the constraint condition in (2) deteriorates the solution. In summary, traditional models are time- and energy-consuming to optimize. For details, we refer the readers to [40], [43], and [44].

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

3) Proposed Model FRGMM: To improve 2), an efficient factor G_{ij} is proposed to simplify EM algorithm solution, making unsupervised learning easier to implement. In order to address problems above, we propose a novel factor G_{ij} defined as

$$G_{ij}^{(t)} = \exp\left[\frac{\beta}{2N_i}\sum_{m\in\partial_i}\left(z_{mj}^{(t)} + \pi_{mj}^{(t)}\right)\right] \tag{8}$$

where β is the temperature coefficient to control energy $U(\Pi)$, N_i is the amount of pixels in neighborhood ∂_i, z_{mj} is the posterior probability, π_{mj} is the prior distribution, t is the iteration step. In this study, we use a window of 5×5 size to traverse global image so that $N_i = 25$. An important consideration is that factor G_{ij} can be considered as a linear filter, which is defined as a multiplication of both posterior probability and prior probability. Thus, it enables to smooth image and filter out noises. In addition, it is known to all that there is a similarity among adjacent pixels in an image. For this, we replace each prior probability and posterior probability in a neighborhood with the average value of adjacent pixels, including itself. Note that factor G_{ij} only depends on the prior and posterior at previous iteration (at the t step). Hence, the proposed model is capable of incorporating spatial information of adjacent pixels in a simpler and more efficient way.

In particular, a new way is proposed to incorporate the spatial information between adjacent pixels into the smoothing prior $U(\Pi)$, which is defined as

$$U(\Pi) = -\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}. \tag{9}$$

It can be seen in (9) that, at current step ($t+1$ step), the derivative of smoothing prior $U(\Pi)$ with regard to prior distribution $\pi_{ij}^{(t+1)}$ is only related to the $\pi_{ij}^{(t+1)}$ term, i.e., the M-step of EM algorithm can be directly applied for π_{ij} in our model, which is much simpler and effective. From (5), the MRF distribution can be updated by

$$p(\Pi) = Z^{-1}\exp\left\{\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}\right\}. \tag{10}$$

From (7), the log-likelihood function is also updated by

$$L(\Pi, \Theta|X) = \sum_{i=1}^{N}\log\left\{\sum_{j=1}^{K}\pi_{ij}^{(t+1)}\phi\left(x_i\middle|\Theta_j^{(t+1)}\right)\right\} - \log Z + \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}. \tag{11}$$

In [40], complete data condition theory has proven that an objective function would increase as log-likelihood function rises in (11). The objective function $L(\Pi, \Theta|X)$ is defined as

$$J(\Pi, \Theta|X) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\left\{\log \pi_{ij}^{(t+1)} + \log\phi\left(x_i\middle|\Theta_j^{(t+1)}\right)\right\} - \log Z + \frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}. \tag{12}$$

In addition, the conditional expectation value z_{ij} of the hidden variables denotes as

$$z_{ij}^{(t)} = \frac{\pi_{ij}^{(t)}\phi\left(x_i\middle|\Theta_j^{(t)}\right)}{\sum_{k=1}^{K}\pi_{ik}^{(t)}\phi\left(x_i\middle|\Theta_k^{(t)}\right)}. \tag{13}$$

So far, (8)–(13) have derived the model estimation rule of the E-step of EM algorithm in FRGMM.

Next, the purpose of optimizing the parameters set $\{\Pi, \Theta\}$ is to further maximize the objective function $J(\Pi, \Theta|X)$ in (12). In this study, both Z and T is set to 1, so that the objective function can be updated as

$$J(\Pi, \Theta|X) = \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\left\{\log \pi_{ij}^{(t+1)} + \log\phi\left(x_i\middle|\Theta_j^{(t+1)}\right)\right\} + \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}. \tag{14}$$

Combining it with (3), the objective function denotes as

$$\begin{aligned} J(\Pi, \Theta|X) &= \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\left\{\log \pi_{ij}^{(t+1)} - \frac{D}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_j^{(t+1)}\right|\right\} \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{K} z_{ij}^{(t)}\left\{-\frac{1}{2}\left(x_i - \mu_j^{(t+1)}\right)^T\Sigma_j^{(t+1)}\left(x_i - \mu_j^{(t+1)}\right)\right\} \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{K} G_{ij}^{(t)} \log \pi_{ij}^{(t+1)}. \end{aligned} \tag{15}$$

Here, let us analyze the derivation of function $J(\Pi, \Theta|X)$ with regard to means μ_j at the $(t+1)$ iteration step, which is given by

$$\frac{\partial J}{\partial \mu_j^{(t+1)}} = \sum_{i=1}^{N} z_{ij}^{(t)}\left[-\frac{1}{2}\left(2\Sigma_j^{-1(t+1)} - 2\Sigma_j^{-1(t+1)}x_i\right)\right]. \tag{16}$$

As the differential equation satisfies $(\partial J/\partial \mu_j) = 0$, μ_j has the minimum value at $(t+1)$ step

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)} x_i}{\sum_{i=1}^{N} z_{ij}^{(t)}}. \tag{17}$$

Similarly, taking the derivative of function $J(\Pi, \Theta|X)$ with regard to $\sum_j^{-1}$ yields

$$\frac{\partial J}{\partial \sum_j^{-1(t+1)}} = \sum_{i=1}^{N} z_{ij}^{(t)}\left[\frac{1}{2}\sum_{j}^{(t+1)} -\frac{1}{2}\left(x_i - \mu_j^{(t+1)}\right)\left(x_i - \mu_j^{(t+1)}\right)^T\right] \tag{18}$$

then set it equal to zero and we have

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ij}^{(t)}\left(x_i - \mu_j^{(t+1)}\right)\left(x_i - \mu_j^{(t+1)}\right)^T}{\sum_{i=1}^{N} z_{ij}^{(t)}}. \tag{19}$$

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

An important consideration is that prior distribution should satisfy the constraints in (2). To enforce these constraints, we use Lagrange's multiplier η_i for each data point

$$\frac{\partial}{\partial \pi_{ij}^{(t+1)}}\left[J-\sum_{i=1}^{N}\eta_i\left(\sum_{j=1}^{K}\pi_{ij}^{(t+1)}-1\right)\right]=0. \tag{20}$$

Equation (20) can be rewritten in the following form:

$$\frac{z_{ij}^{(t)}}{\pi_{ij}^{(t+1)}}+\frac{G_{ij}^{(t)}}{\pi_{ij}^{(t+1)}}-\eta_i=0. \tag{21}$$

Lagrange multiplier η_i should be given as follows because of the constraint $\sum_{j=1}^{K}\pi_{ij}=1$

$$\eta_i=1+\sum_{j=1}^{K}G_{ij}^{(t)}. \tag{22}$$

Considering after $(t+1)$ step, the necessary condition for prior distribution π_{ij} denotes as

$$\pi_{ij}^{(t+1)}=\frac{z_{ij}^{(t)}+G_{ij}^{(t)}}{\sum_{k=1}^{K}\left(z_{ik}^{(t)}+G_{ik}^{(t)}\right)}. \tag{23}$$

So far, (15), (17)–(19), and (23) have derived the parameter $\{\Pi,\Theta\}$ optimization rule of the M-step of EM algorithm in FRGMM. Once the parameter-learning phase is completed, each pixel x_i is distributed to corresponding label Ω_j with a largest posterior probability z_{ij}

$$x_i\in\Omega_j:\ \text{IF } z_{ij}\geq z_{ik};\quad j,k=(1,2,\ldots,K). \tag{24}$$

B. Location Model Faster RCNN

A state-of-the-art detection algorithm Faster RCNN [46], [47] is studied to locate the defect on complex and diverse rail surface, making the proposed DM-RIS robust environmentally to meet different light, angle, background, and acquisition equipment.

1) Region Proposal Network: Distinguished from selective search (SS) [48] and EdgeBoxes [49], region proposal network (RPN) estimates any possible regions of rail defects with a pretrained CNN. VGG-16 is given as feature extractor. A 3×3 window (convolution kernel) is utilized to slide on feature map to generate a 512-D vector. Then, it is fed into two parallel fully connected (FC) layers, which are box classification layer (cls) and box regression layer (reg). For each window, $k=9$ anchors are proposed with three size and three length–width ratio. For this, the reg has 4×9-dimensional output for coordinate coding of nine estimated boxes, and cls has 2×9-dimensional output for score that each estimated box is target.

2) Fast RCNN: The Fast RCNN is provided for surface defect location in each rail image. ROI pooling layer uses spatial pyramid pooling. Then, feature vector is fed into Softmax classification layer and bounding regression layer, respectively, through FC layer for category prediction and bounding box correction [47]. Moreover, nonmaximal suppression (NMS) would eliminate repeated bounding boxes. The cost function for training Fast RCNN is given by

$$L(p,u,t^u,\upsilon)=L_{\text{cls}}(p,u)+\lambda[u\geq1]L_{\text{reg}}(t^u,\upsilon) \tag{25}$$

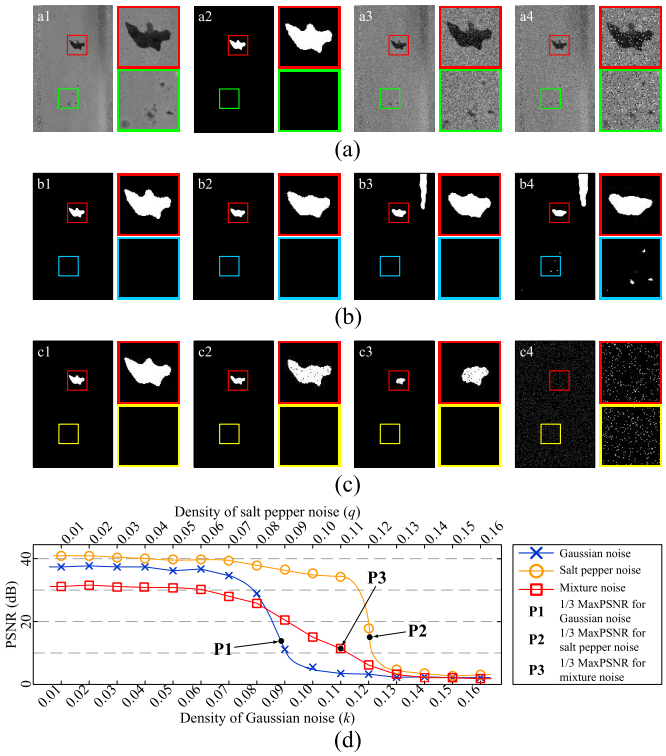

Fig. 3. Evaluation on noise robustness of FRGMM. (a) Original image, ground truth (GT), noise examples corrupted Gaussian and salt pepper in (a1)–(a4), respectively. (b) Results of Gaussian images with density $k=0.01$, 0.07, 0.09, 0.10 in (b1)–(b4). (c) Results of salt pepper images with density $q=0.01$, 0.11, 0.12, 0.13 in (c1)–(c4). (d) How denoising performance (measured by PSNR) changes as noise increases.

where L_{cls} is the classification cost, L_{reg} is the bounding regression cost, p is the classification prediction result, u is the u-class, t^u is the modified value of u-class bounding regression prediction, and υ is the actual modified value.

IV. Experiment and Analysis

A. Segmentation Proposal by FRGMM

1) Factor-Based Performance Evaluation: In (8), a novel factor G_{ij} is proposed to incorporate spatial information into the MRF-based GMM, making it robust to noise. To verify this, the first experiment is provided.

Fig. 3(b) presents several representative results under Gaussian noise of various densities k. Fig. 3(b1) is slightly polluted [$k=0.01$, peak signal-to-noise ratio (PSNR) = 39.12 dB] so that segmented defect has remarkable boundary. With moderate noise, some edge details are lost in Fig. 3(b2) ($k=0.07$, PSNR = 32.57 dB). So far, the Gaussian noise has been effectively filtered out; however, FRGMM wrongly segments the background in the challenging example shown in Fig. 3(b3) with PSNR = 12.88 dB. Note that FRGMM fails in Fig. 3(b4) once any noise appears in the blue box ($k=0.10$, PSNR = 6.68 dB).

Similar work is conducted on salt pepper images of various densities q. Fig. 3(c1) is slightly polluted ($q=0.01$, PSNR = 41.72 dB) so that segmented defect has sufficient edge details. In Fig. 3(c2), FRGMM performs, well although a little pepper noise remained around defect area ($q=0.11$,

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPUTATIONAL COST (IN SECONDS) OF SVFMM AND FRGMM

Algorithm	2th iteration		15th iteration		40th iteration	
	E step	M step	E step	M step	E step	M step
SVFMM	0.025	0.849	0.034	0.883	0.958	74.31
FRGMM	0.022	0.003	0.031	0.003	0.914	0.095

PSNR = 33.54 dB). With severe noise of $q = 0.12$, most of the edge details are lost in Fig. 3(c3) with PSNR = 18.22 dB. Note that FRGMM fails in Fig. 3(c4) because there is a large amount of noise remained in the global result and rail defect cannot be identified.

Comprehensively, Fig. 3(d) shows how the denoising performance measured by PSNR changes in various noise conditions. It can be seen that FRGMM works at one-third of the best performance at P1, P2, and P3, which indicates that it is not competent for segmentation and denoising tasks at such noise levels. Here, we consider one-third of MaxPSNR value as a threshold for judging good or poor robustness. Specifically, for the Gaussian noise, FRGMM performs well up to $k = 0.09$, whereas for salt and pepper noise, it works up to $q = 0.12$. For mixture noise which is mixed by the Gaussian and salt pepper at a ratio of 1:1, it is competent up to $k(q) = 0.11$. An important consideration is that, in our rail surface imaging system, the noise is with a density of 0.03–0.06 in usual, thus it can be seen that FRGMM is not only an effective segmentation tool but also a good filter.

In (9), an effective smoothing prior $U(\Pi)$ is defined to make the derivative with respect to prior distribution π_{ij} at the current step (at $t+1$ step) only dependent to term $\pi_{ij}^{(t+1)}$, so that the M-step of EM algorithm is much simpler and easy to implement. Next experiment is provided to verify it.

We perform our method on rail image (Fig. 3) to compare with spatially variant finite mixture model (SVFMM) [44]. Both methods are initialized under the same initial condition and log-likelihood functions iterate toward convergence. Results show that FRGMM converts at the 15th iteration, whereas SVFMM converts slowly after the 25th iteration. In addition, we look inside each iteration of SVFMM and our method in Table II. At each iteration, the cost of E-step is almost the same in both methods. However, since the M-step of SVFMM cannot be directly applied to prior distribution, more time is consumed to maximize the log-likelihood function, proving that our FRGMM is faster and simpler.

2) Sample-Based Performance Evaluation: Next, we take into account the pervasiveness or environmental robustness of FRGMM. Fig. 4 shows the results of different mixture models on various rail samples. Note that rail surface C1 and C2 belong to simple scene, which has been tested before [38]. However, the performance on complex rail surface such as C3, C4, C5, and C6 has not ever been discussed, which is a concern of this paper.

An important consideration is that several advanced mixture models are used for comparison, e.g., K-means algorithm, standard GMM [39], SVFMM [44], Student-T mixture model, and two MRF-based mixture models, namely, simulated field algorithm (SIMF) and mean field algorithm (MEANF) [50], [51]. The experiment is implemented in MATLAB environment of MacBook Pro with 3.1 GHz, Intel Core i5, and 8 GB. All algorithms run until the convergence.

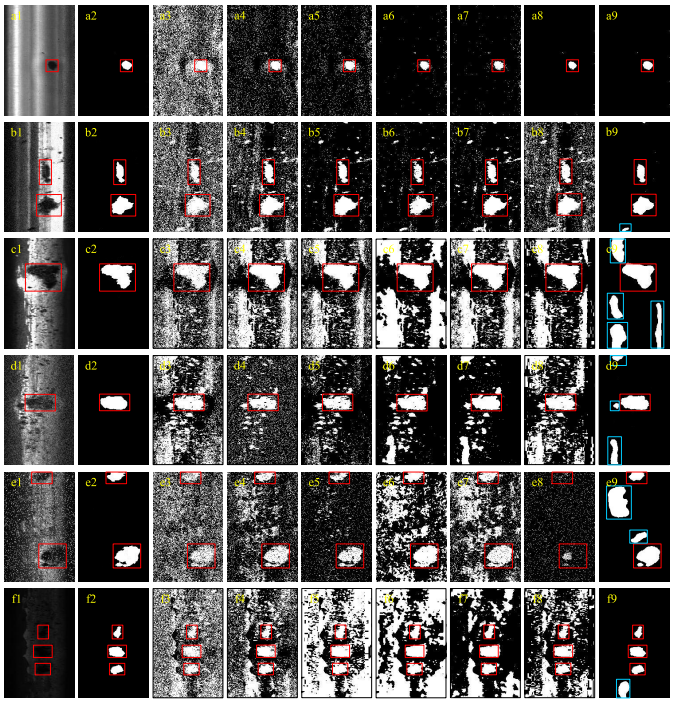

Fig. 4. Results of different mixture models on various rail samples (a1)–(a9) for C1, (b1)–(b9) for C2, (c1)–(c9) for C3, (d1)–(d9) for C4, (e1)–(e9) for C5, and (f1)–(f9) for C6, horizontally, and (a1)–(f1) original image, (a2)–(f2) GT, (a3)–(f3) K-means, (a4)–(f4) GMM, (a5)–(f5) SVFMM, (a6)–(f6) SIMF, (a7)–(f7) MEANF, (a8)–(f8) Student-T, and (a9)–(f9) proposed FRGMM, vertically. Red box: real defect. Blue box: false segmentation.

The quantitative results are recorded in Table III, where the misclassification ratio (MCR), PSNR, and mean squared error (MSE) are utilized to evaluate the corrupted original image with noise, the rand index (RI), variation of information (VI), and global consistency error (GCE) are utilized to evaluate real image. Additionally, values in red are the best while the ones in blue are the second best results. The following will be discussed in two parts.

1) *Simple Rail Samples:* For smooth rail C1 with reflection inequality, K-means (MCR = 71.84%, Time = 0.43), GMM (MCR = 53.46%, Time = 0.32), and SVFMM (MCR = 46.62%, Time = 80.31) present noisy diffusion, indicating that these algorithms are very sensitive to strong light so severe misclassification occurs. In comparison, MRF-based mixed model SIMF (MCR = 10.31%, Time = 163.25), MEANF (MCR = 14.18%, Time = 121.80), and Student-T (MCR = 2.56%, Time = 15.36) can suppress most noise. Moreover, the proposed FRGMM (MCR = 2.00%, Time = 0.81) performs well with little noise or false segmentation. For smooth rail C2 with the Gaussian noise and reflection inequality, K-means (MCR = 69.51%, Time = 0.31), GMM (MCR = 60.25%, Time = 0.12),

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT MODELS ON VARIOUS SAMPLES

Image	Algorithm	MCR(%)	RI(%)	PSNR(dB)	MSE	VI(pixel)	Time(s)	GCE(%)
Rail surface C1	k-means	71.84	18.13	3.65	2.80E+04	4.13	0.43	4.20
	GMM [39]	53.46	26.56	9.85	6.73E+03	3.65	0.32	4.10
	SVFMM [44]	46.62	32.99	12.29	3.84E+03	3.28	80.31	4.02
	SIMF [50]	10.31	83.18	21.76	434.09	1.00	163.25	3.49
	MEANF [51]	14.18	76.59	20.35	599.76	1.30	121.80	3.58
	Student-T	2.56	96.72	25.72	174.19	0.27	15.36	2.18
	FRGMM	2.00	98.85	32.95	33.00	0.16	0.81	1.57
Rail surface C2	k-means	69.51	25.90	4.59	2.26E+04	4.41	0.31	13.97
	GMM [39]	60.25	30.12	7.12	1.26E+04	4.09	0.12	13.13
	SVFMM [44]	55.66	33.62	7.98	1.04E+04	3.88	117.14	12.73
	SIMF [50]	29.24	62.22	15.04	2.04E+03	2.40	108.26	10.43
	MEANF [51]	35.67	54.13	12.89	3.35E+03	2.74	97.72	10.71
	Student-T	61.16	29.63	6.62	1.41E+04	4.12	9.28	13.28
	FRGMM	8.44	94.65	22.33	380.00	0.64	0.56	5.92
Rail surface C3	k-means	70.50	28.40	4.53	2.29E+04	4.51	0.19	17.47
	GMM [39]	72.00	27.91	3.93	2.63E+04	4.37	0.23	17.26
	SVFMM [44]	68.78	28.86	4.59	2.26E+04	4.32	116.49	17.11
	SIMF [50]	66.56	34.27	3.70	2.77E+04	3.61	112.17	17.55
	MEANF [51]	71.11	28.86	3.93	2.63E+04	4.25	96.84	17.31
	Student-T	56.39	36.83	5.50	1.83E+04	3.67	16.49	16.90
	FRGMM	24.81	70.15	9.09	8.02E+03	1.78	0.47	15.68
Rail surface C4	k-means	67.58	24.02	4.42	2.35E+04	4.17	0.28	10.18
	GMM [39]	60.74	26.57	7.17	1.25E+04	4.08	0.34	9.91
	SVFMM [44]	46.87	38.22	9.16	7.89E+03	3.44	131.28	9.42
	SIMF [50]	32.32	55.38	9.45	7.39E+03	2.51	97.54	9.04
	MEANF [51]	31.14	57.28	9.09	7.53E+03	2.61	148.48	9.58
	Student-T	53.96	33.53	5.43	1.86E+04	3.48	13.48	10.04
	FRGMM	11.70	87.23	14.88	2.11E+03	0.96	0.59	6.93
Rail surface C5	k-means	70.16	26.98	4.57	2.27E+04	4.52	0.22	15.74
	GMM [39]	63.15	29.79	6.77	1.37E+04	4.32	0.18	14.98
	SVFMM [44]	53.92	36.15	11.22	4.91E+03	3.87	140.65	13.61
	SIMF [50]	67.46	27.95	5.23	1.95E+04	4.43	113.48	15.30
	MEANF [51]	64.13	29.65	5.48	1.84E+04	4.21	152.73	14.97
	Student-T	52.71	35.24	10.24	6.15E+03	3.92	42.51	15.98
	FRGMM	17.88	80.57	10.83	5.38E+03	1.30	0.40	12.32
Rail surface C6	k-means	73.44	22.90	3.32	3.03E+04	4.38	0.51	10.93
	GMM [39]	67.61	24.11	4.69	2.21E+04	4.25	0.44	10.74
	SVFMM [44]	52.96	32.55	1.96	4.14E+04	3.60	128.03	10.98
	SIMF [50]	59.08	31.45	2.69	3.50E+04	3.49	124.16	10.93
	MEANF [51]	45.18	41.68	6.23	1.55E+04	3.11	142.51	10.52
	Student-T	56.90	31.38	5.25	1.94E+04	3.69	35.80	10.69
	FRGMM	7.60	93.83	17.50	1.16E+03	0.58	0.76	6.60

SVFMM (MCR = 55.66%, Time = 117.14), SIMF (MCR = 29.24%, Time = 108.26), MEANF (MCR = 35.67%, Time = 97.72), and Student-T (MCR = 61.16%, Time = 9.28) present noisy diffusion, indicating that these algorithms fail to suppress the Gaussian noise so severe misclassification occurs. In comparison, the proposed method (MCR = 8.44%, Time = 0.56) can filter out external noise and defect edge is segmented accurately and quickly, although one error in blue box shown in Fig. 4(b9). To conclude, with the smallest MCR, VI, and GCE and the biggest RI and PSNR, FRGMM could obtain defect edge fast and accurate, and is robust to reflection inequality and external noise.

2) *Complex Rail Samples:* For rough rail C3 with rust, greasy dirt, and reflection inequality, K-means (MCR = 70.50%, Time = 0.19), GMM (MCR = 72.00%, Time = 0.23), SVFMM (MCR = 68.78%, Time = 116.49), SIMF (MCR = 66.56%), Time = 112.17), MEANF (MCR = 71.11%, Time = 96.84), Student-T (MCR = 56.39%, Time = 16.49) perform badly. In comparison, ours (MCR = 24.81%, Time = 0.47) gets rid of noise and segment edge rapidly and precisely. Note that four errors remain in blue box of Fig. 4(c9). For rough rail C4 with the Gaussian noise, rust, greasy dirt, and reflection inequality, K-means (MCR = 67.58%, Time = 0.28), GMM (MCR = 60.74%, Time = 0.34), SVFMM (MCR = 46.87%, Time = 131.28), Student-T (MCR = 53.96%, Time = 13.48) present severe noisy diffusion. SIMF (MCR = 32.32%, Time = 97.54) and MEANF (MCR = 31.14%, Time = 148.48) present severe false segmentation. In comparison, the proposed FRGMM (MCR = 11.70%, Time = 0.59) gets rid of the Gaussian noise and segment defect edge accurately and quickly. Note that three errors remain in blue box of Fig. 4(d9). For rough rail C5 with salt pepper noise, rust, greasy dirt, and reflection inequality, K-means (MCR = 70.16%, Time = 0.22), GMM (MCR = 63.15%, Time = 0.18), SVFMM (MCR = 53.92%, Time = 140.65), SIMF (MCR = 67.46%, Time = 113.48), MEANF (MCR = 64.13%), Time = 152.73), Student-T (MCR = 52.71%, Time = 42.51) present noisy diffusion and false segmentation. In comparison, our method (MCR = 17.88%, Time = 0.40) gets rid of salt pepper noise and segment the edge rapidly and precisely. Note that two errors in blue box of Fig. 4(e9) for the reason of rust and greasy dirt. For rough rail C6 with weak illumination, and rust and greasy dirt, K-means (MCR = 73.44%, Time = 0.51), GMM (MCR = 67.61%, Time = 0.44), SVFMM (MCR = 52.96%, Time = 128.03), SIMF

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

(MCR = 59.08%, Time = 124.16), MEANF (MCR = 45.18%, Time = 142.51) perform badly for the reason of low contrast. However, ours (MCR = 7.60%, Time = 0.76) segments the defect edge rapidly and accurately. Note that one error remains in blue box of Fig. 4(f9).

To conclude, with the smallest MCR, VI, and GCE and the biggest RI and PSNR, FRGMM could obtain defect edge fast and accurate, and is robust to reflection inequality, various external noise, and weak illumination. However, besides real defects, many errors still appear in the results, showing FRGMM is not robust environmentally and it can only achieve a presegmentation on complex and diverse rail image.

B. Objective Location by Faster RCNN

Section IV-A2 illustrates that FRGMM is not widely effective, especially for complex rail surface like C3, C4, C5, and C6, where many errors and noisy diffusion frequently occur, although the defect edge is segmented accurately. To solve this problem, a deep multimodel system is proposed with a parallel structure, i.e., FRGMM generates the segmentation proposal while Faster RCNN carries out the objective location, then the joint hit area refers to real defect.

To make Faster RCNN equipped with good capacity, the performance with different training steps and different learning rates will be discussed below. The model is trained on Tensorflow with Ubuntu 16.04, Intel Xeon 24 core CPU, two GTX Titan xp GPU and 64-GB memory.

1) Test on Training Step: In this experiment, we design 12-step scales in the process of model training, with a learning rate of 0.001, a dropout rate of 0.98, and batch size of 128. To detect the surface defects on complex and diverse scenes, the train set contains rail samples C1–C6, which accounts for 70% of the whole data set, and the validation set takes up 20%. Results show that at 4k (L1) step, Faster RCNN outputs the mean prediction probability of only 0.4287 and mean recognition accuracy of only 0.5195. As step goes up, the probability and accuracy are gradually increasing. At 12k (L9) step, Faster RCNN predicts the maximum probability of 0.8120 and maximum accuracy of 0.9524. When training step continues to 20k (L12), mean probability drops to 0.4928 and accuracy is only 0.4083. In addition, single-graph inference time is between 2.1 and 2.5 ms, showing that defect detection is real time.

Specifically, a box plot of prediction probability is drawn in Fig. 5 to search for the best training step. At 4k–7k (L1–L4) step, a very small median and very long rectangle with few outliers demonstrate that model outputs a large number of estimates with small prediction probability, that is not accurate and stable. During 8k–10k (L5–L7) step, the median is a bit large while the rectangle is still long with some outliers at nearly zero, indicating that the model could only output a small number of estimates with large probability. During 11k–13k (L8–L10) step, the rectangle is very short and the median is very large; however, the model performs unsteadily with many outliers. At 12k (L9) step, the rectangle is the shortest and median is the largest, although for some outliers, the extreme outlier is much higher than all the lower quartiles

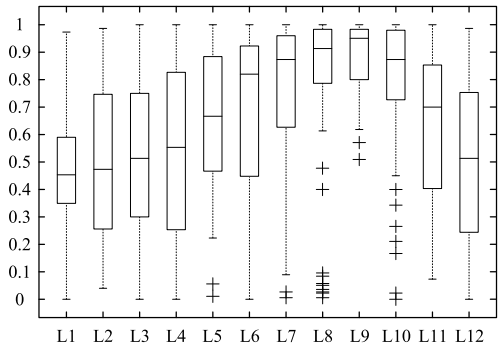

Fig. 5. Box plot of prediction probability at training step. L1–L10 refers to training steps, namely, 4k, 5k, 6k, 7k, 8k, 9k, 10k, 11k, 12k, 13k, 15k, and 20k.

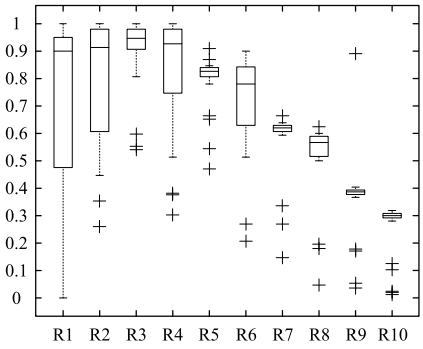

Fig. 6. Box plot of prediction probability with learning rate. R1–R10 refers to learning rates, namely 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}, and 10^{-10}.

by steps of 4k, 5k, 6k, and 7k. Taking a close look, only two outliers of not less than 0.50 probability show that the model performs well with large prediction probability in most cases. At 15k and 20k (L11, L12) step, model gradually deteriorates.

2) Test on Learning Rate: In next experiment, we design ten learning rates in the process of model training, with the step of 12k (L9), a dropout rate of 0.98, and batch size of 128. To detect the surface defects on complex and diverse scenes, the train set contains rail samples C1–C6, which accounts for 70% of the whole data set, and the validation set takes up 20%. The results show that with a learning rate greater than 10^{-2} (R2), the mean prediction probability and recognition accuracy gradually increase as it falls. At 10^{-3} (R3) rate, Faster RCNN outputs a maximum probability of 0.9146 and a maximum accuracy of 0.9989. When the rate ranges from 10^{-4} to 10^{-6} (R4–R5), over 0.70 probability and about 0.93 accuracy demonstrate that the model is stable with any variation of learning rate. In addition, the performance grows worse significantly during $10^{-7} - 10^{-10}$ (R7–R10).

Specifically, a box plot of prediction probability is drawn in Fig. 6 to search for the best learning rate. During $10^{-1} - 10^{-2}$ (R1–R2) rate, a very large median and very long rectangle with some outliers demonstrate that model could predict several good estimates while it is unstable. With 10^{-3} (R3) rate, the rectangle is very short and the median is the largest. Taking a close look, there are only three extreme outliers higher than both the lower quartiles by rates of 10^{-1}

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

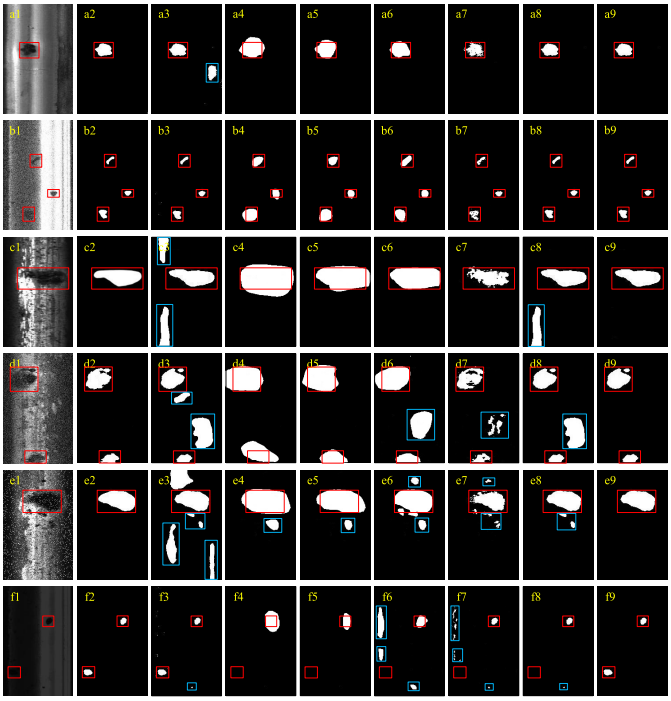

Fig. 7. Results of different methods on various rail samples (a1)–(a9) for C1, (b1)–(b9) for C2, (c1)–(c9) for C3, (d1)–(d9) for C4, (e1)–(e9) for C5, and (f1)–(f9) for C6, horizontally, and (a1)–(f1) original image, (a2)–(f2) GT, (a3)–(f3) Previous, (a4)–(f4) FCN-32s, (a5)–(f5) FCN-8s, (a6)–(f6) SegNet, (a7)–(f7) SegNet + CRF, (a8)–(f8) YOLO + FRGMM, (a9)–(f9) proposed system, vertically. Red box: real defect. Blue box: false segmentation.

and 10^{-2}, indicating that the model performs well in most cases. When $10^{-4} - 10^{-10}$ (R4–R10) rate, median goes down and rectangle becomes longer with a large number of outliers, showing that model gradually deteriorates.

To conclude, Faster RCNN performs well with a training step of 12k (L9) and learning rate of 10^{-3} (R3).

C. Multimodel Parallel System

1) Sample-Based Evaluation: The experiments above have separately discussed two modules of our system, namely, the segmentation proposal and objective location. Next, the performance of the overall system will be taken into account. Fig. 7 shows the results of different methods on various rail samples. Note that rail surfaces C1 and C2 belong to simple scenes, while C3–C6 belong to complex scenes, which is a concern of this study.

An important consideration is that advanced semantic segmentation methods are used for comparison, which is divided into type I, II, and III, namely, probability graph model (PGM)-based method, pure CNN-based method, and combined method. For type I, our previous work is utilized as PGM-based method. For type II, two fully convolutional network (FCN)-based methods [52] FCN-32s, FCN-8s, and an Encoder–Decoder method SegNet [53] are utilized as pure CNN-based methods. For type III, SegNet with conditional random field (CRF) algorithm [54], our method and its variant YOLO + [55] FRGMM are utilized as combined methods. In addition, CNN train and test are carried on Caffe with Ubuntu 16.04, Intel Xeon 24 core CPU, two GTX Titan xp GPU and 64-GB memory. Other tests are implemented in MATLAB of MacBook Pro with 3.1 GHz, Intel Core i5, and 8 GB. All algorithms run until the convergence.

The quantitative results are recorded in Table IV, where error positive (EP), precision (Pr), and recall (Rc) are utilized to evaluate the pixelwise accuracy, Dice and Jaccard (Jacc) are utilized to evaluate the edgewise accuracy, conformity (Confm) is a coefficient of consistency, inference time (Infer) also gives the running time of YOLO and Faster RCNN in brackets, probability (Prob) is for predicted bounding box, and memory (Memo) is the size of model on disk. Additionally, values in red are the best while the ones in blue are the second best results. The following will be discussed in two parts.

1) *PGM-Based Method:* It is random for previous work. In general, it performs well on the simple rail with robust to noise and strong light. Although defect edge is obtained accurately for complex rail, it is very sensitive to salt pepper noise, rust, greasy dirt, and weak illumination, making many errors in global segmentation proposal. Two exceptions are C1 (Dice = 72.57%, false positive (FP) = 40.65%, Infer = 6300 ms) and C6 (Dice = 96.13%, FP = 37.96%, Infer = 5950 ms). In addition, the step-by-step image preprocessing is time-consuming, which is ROI extraction and curvature filter.
2) *CNN-Based Method:* FCN-32s, FCN-8s, and SegNet learn the deep convolution features on complex images and then rail defects are correctly classified. In precision, SegNet is slightly better than FCN-8s followed by FCN-32s. However, CNN fails to take into account the spatial information between neighboring pixels, so that results are much rough, fuzzy, and smooth for the lack of spatial consistency. For instance, many errors occur by FCN-32s (Dice = 64.50%, FP = 58.50%, Infer = 25.6 ms), FCN-8s (Dice = 71.13%, FP = 53.46%, Infer = 120.3 ms), SegNet (Dice = 75.30%, FP = 49.41%, Infer = 48.7 ms) on complex rail C5, showing CNN-based methods are sensitive to salt pepper noise. In addition, the deep learning framework brings high speed but large cost (Memo = 519, 538, 106 MB, respectively).
3) *Combined Method:* CNN and PGM are well incorporated into method III, where CNN extracts deep features to make it robust to complex and diverse rail scene, and PGM considers spatial information between neighboring pixels to obtain defect edge with high precision. They perform differently on rail samples.

For instance, SegNet + CRF (Dice = 82.96%, FP = 7.41%, Infer = 2450 ms) oversegments defects on sample C3, making the edge much noisy and unsmooth. Also, the CRF algorithm optimizes parameter too slowly to converge. YOLO + FRGMM (Dice = 72.09%, FP = 42.24%, Infer = 530(1.0) ms, Prob = 86.62%) applies the detection strategy of Ours to obtain defect accurately, which will be discussed in Section V. However, lower probability of bounding box balances the advantage of high speed, i.e., YOLO could predict rust and tiny defects.

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON VARIOUS SAMPLES

Image	Type	Method	Dice(%)	Jacc(%)	FP(%)	Confm	Pr(%)	Rc(%)	Infer(ms)	Prob(%)	Memo(MB)
C1	I	Method in [38]	72.57	56.95	40.65	0.24	58.14	96.53	≈6300	N/A	6.5
	II	FCN-32s [52]	57.00	39.86	69.95	-0.51	39.92	99.61	24.6		519
		FCN-8s [52]	68.40	51.98	60.51	0.08	52.56	97.94	120.5		538
		SegNet [53]	77.66	63.47	50.15	0.42	63.58	99.74	52.3		106
		SegNet+CRF [54]	87.90	78.41	18.14	0.72	87.01	88.80	≈1900		121
	III	YOLO+FRGMM [55]	95.46	91.56	6.23	0.91	95.40	96.35	450(1.1)	84.29	79
		DM-RIS	95.77	91.89	6.23	0.91	95.29	96.27	450(2.4)	95.63	452
C2	I	Method in [38]	95.87	92.07	35.08	0.91	94.45	97.33	≈5700	N/A	N/A
	II	FCN-32s [52]	51.34	34.54	70.91	-0.90	34.54	100.0	22.3		
		FCN-8s [52]	61.33	44.23	65.57	-0.26	44.23	100.0	114.9		
		SegNet [53]	65.46	48.66	60.18	-0.06	48.66	100.0	50.6		
		SegNet+CRF [54]	87.24	77.37	34.30	0.71	85.95	88.57	≈1750		
	III	YOLO+FRGMM [55]	96.23	92.74	8.23	0.92	95.16	97.33	420(1.0)	81.36	
		DM-RIS	96.27	92.81	8.23	0.92	95.25	97.84	420(2.1)	96.48	
C3	I	Method in [38]	63.64	46.67	53.23	-0.14	49.79	88.18	≈5900	N/A	
	II	FCN-32s [52]	56.89	39.75	69.32	-0.52	39.75	100.0	24.8		
		FCN-8s [52]	69.81	53.62	57.13	0.13	53.62	100.0	118.6		
		SegNet [53]	77.54	63.32	50.29	0.42	63.32	100.0	55.8		
		SegNet+CRF [54]	82.96	70.88	7.41	0.59	94.70	73.80	≈2100		
	III	YOLO+FRGMM [55]	72.09	56.36	42.24	0.23	60.97	88.18	530(1.0)	86.62	
		DM-RIS	92.50	86.04	11.19	0.84	97.25	88.56	530(2.2)	94.70	
C4	I	Method in [38]	57.20	40.06	68.09	-0.50	42.03	89.55	≈6800	N/A	
	II	FCN-32s [52]	50.02	33.35	79.57	-1.00	34.38	91.75	27.5		
		FCN-8s [52]	64.32	47.40	72.45	-0.11	47.40	100.0	129.4		
		SegNet [53]	53.67	36.68	78.94	-0.73	36.68	100.0	60.5		
		SegNet+CRF [54]	81.68	69.04	27.12	0.55	75.42	89.07	≈2450		
	III	YOLO+FRGMM [55]	61.04	43.92	24.15	-0.28	46.30	89.55	650(1.3)	82.74	
		DM-RIS	93.09	87.08	9.50	0.85	96.92	89.55	650(2.4)	98.52	
C5	I	Method in [38]	54.22	37.19	58.98	-0.69	38.03	94.41	≈6200	N/A	
	II	FCN-32s [52]	64.50	47.60	58.50	-0.11	47.60	100.0	25.6		
		FCN-8s [52]	71.13	55.19	53.46	0.19	55.20	99.96	120.3		
		SegNet [53]	75.30	60.38	49.41	0.34	60.41	99.92	48.7		
		SegNet+CRF [54]	82.71	70.51	11.91	0.59	86.68	79.09	≈2300		
	III	YOLO+FRGMM [55]	93.05	87.00	10.09	0.85	91.65	94.49	390(1.1)	80.44	
		DM-RIS	97.17	94.49	4.25	0.94	99.96	94.53	390(2.2)	95.18	
C6	I	Method in [38]	96.13	92.56	37.96	0.92	92.79	99.73	≈5950	N/A	
	II	FCN-32s [52]	20.07	11.16	86.69	-6.96	12.99	44.12	26.4		
		FCN-8s [52]	36.26	22.15	70.68	-2.52	30.78	44.12	118.3		
		SegNet [53]	16.56	9.03	89.26	-9.08	10.19	44.12	52.6		
		SegNet+CRF [54]	42.54	27.01	47.66	-1.70	42.04	43.05	≈2260		
	III	YOLO+FRGMM [55]	60.11	42.97	2.90	-0.33	94.29	44.12	470(1.0)	83.27	
		DM-RIS	96.27	93.80	0.35	0.98	95.64	98.05	470(2.2)	97.39	

In comparison, ours (Dice = 92.50%, FP = 92.50%, Infer = 530(1.0) ms, Prob = 94.70%) proposes a new smoothing prior to incorporate spatial information between neighboring pixels into MRF-GMM, thus the M-step of EM algorithm can be directly used to prior distribution. Also, the mixture model satisfies the constraint (20), making it not over-smooth the edge. In addition, Faster RCNN can output high prediction probability fast and steadily.

To sum up, the combined methods are generally superior to pure CNN-based methods followed by PGM-based method. Our system (mDice = 95.18%, mJacc = 91.02%, mPr = 96.74%, mRc = 94.13%, mInfer = 485 ms, mProb = 96.32%, Memo = 452 MB) is very accurate, robust for various rail scenes, and fast with reasonable model cost.

Fig. 8 evaluates the accuracy in view of pixel classification and edge overlap. It can be seen that: 1) generally, two indexes maintain the same trend, although mAP of YOLO + FRGMM is over 90% and Jacc is below 40% in C6, and mAP of SegNet + CRF is over 90% and Jacc is only 70% in C3; 2) ours ranks top one steadily with mAP and Jacc over 90%; 3) CNN-based methods are moderate with two indexes around 60% and 50%; and 4) previous and SegNet + CRF are unstable, with exception in C1, and errors in C3, C4, C5, C6 caused by rust, external noise, and weak illumination.

2) Data Set-Based Evaluation: Fig. 9 uses the whole test set to evaluate the performance of different methods. Note that it is

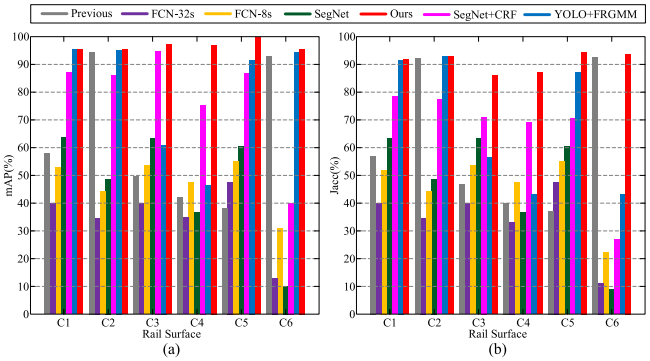

Fig. 8. Accuracy of different methods on various rail samples. (a) Pixelwise accuracy. (b) Edgewise accuracy.

drawn by Curve Fitting Tool in MATLAB and data here also include samples from other data sets [21], [25], [38]. It can be seen that: 1) for all types of rail surface, the multimodel system proposed is outstanding and stable. For instance, once only 10% of defects are detected (Recall = 10%) in data set C3, ours (Precision = 98%) is 10 times better than YOLO (Precision = 80%), and 20 times better than Previous (Precision = 60%), and 30 times better than CNN-based method (Precision = 40–60%); 2) CNN-based methods are also steady but lose local information on defect edge, making a moderate accuracy from 50% to 70% roughly; 3) several

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

TABLE V
STATISTIC AND COMPARISON OF DCNN-BASED RAILWAY FAILURE DIAGNOSIS METHODS

Method	Year	Components	Defect	Characteristic	Accuracy (%)	Speed (s/frame)
MPCNN [28]	2012	Rail	Surface defect	Max-pooling DCNN	92.42	0.00875
Method in [30]	2014	Rail	Surface defect	Photometric stereo	92.89	N/A
Method in [24]	2015	Concrete ties	Tie condition	Material classification	93.35	N/A
Method in [25]	2016	Rail	Surface defect	DCNN for classification	90.77	1.375
MTL [23]	2017	Ties / Fastener	Missing / Defective	Multitask learning	96.74	N/A
ILNET [27]	2017	Isoelectric line	Loosen	DCNN with ICM/MPM	96.54	N/A
Method in [26]	2018	Fastener	Missing	Three-stage cascaded DCNN	92.78	0.00157
PVANET++ [29]	2018	Split pins	Missing / Loosen	Condition-based detector	95.28	2.325
DM-RIS	2018	Rail	Surface defect	Deep multi-model parallel system	96.74	0.485

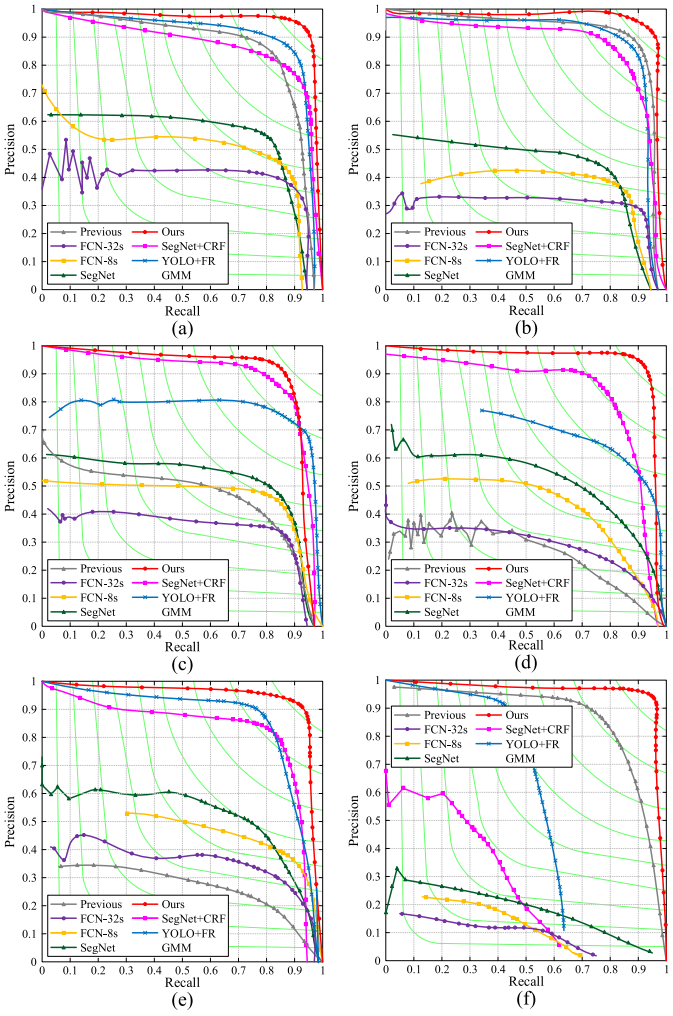

Fig. 9. Performance of different methods on various data sets. (a) Rail data set C1. (b) Rail data set C2. (c) Rail data set C3. (d) Rail data set C4. (e) Rail data set C5. (f) Rail data set C6.

"divings" occur in data set C6, indicating Previous, SegNet + CRF, and YOLO + FRGMM perform badly. To conclude, our system has combined the advantage of PGM and CNN, with high precision and environmental robustness.

3) Further Comparison With Other Studies: Table V presents recently published research studies where DCNN is successfully applied for railway failure diagnosis. Note that those working on rail surface defect detection are marked in bold, and remarkable results are marked in red. "N/A" means there is no record in the original paper. Quantitative

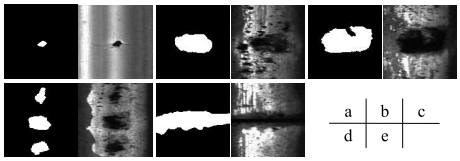

Fig. 10. Type of defects inspected with our system. (a) Light squat. (b) Moderate squat. (c) Severe squat. (d) Spalling. (e) Joint.

indexes "Accuracy" and "Speed" are in a broad sense for the reason that the experimental condition, subject and evaluation criteria are quite different in each research. For instance, the speed of [26] is extremely fast because three DCNNs are cascaded for the target location task, which differs from the semantic segmentation task in this paper. In addition, various data sets are utilized in [30] and more interest is grown in classification challenge in [25], although all of us work for rail steels. An important consideration is that direct numerical comparison is not the genuine purpose of Table V, as various subjects, objectives, and conditions would not support this. Instead, the authors wish to show that proposed DM-RIS with 96.74% accuracy and 0.485 s/frame speed is fairly promising and competent. In addition, the proposed system can effectively detect various types of defects as shown in Fig. 10.

V. DISCUSSION

Various frameworks lead to diverse detection strategies and results. For instance, in a cascaded architecture, front-end Faster RCNN predicts a bounding box and then FRGMM performs on the local image within it, which is a step-by-step strategy. Instead, in the proposed DM-RIS, the global image is utilized by FRGMM, and meanwhile, an estimation is made by Faster RCNN, which is an end-to-end parallel strategy. Furthermore, some representative results are shown as follows.

A. Robustness Comparison

Fig. 11 provides results as a wrong location (FP) and a missing location (true negative) occur in Faster RCNN. It can be seen in Fig. 11(a1) that DM-RIS not only segments the rail defect effectively but also eliminates the rusty background robustly, due to that FRGMM has filtered out most noise in such a parallel architecture. When a missing situation occurs in Fig. 11(b1), the center of bounding box falls right within segmentation proposal so that complete defect is still hit, although an incomplete estimation is made in the path of objective location. To conclude, when one model fails, the whole system remains robust in DM-RIS.

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

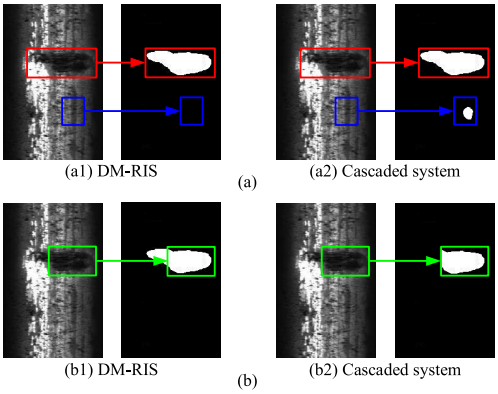

Fig. 11. Robustness comparison of two systems. (a) FP. (b) True negative. (a1) and (b1) Detection process of proposed DM-RIS. (a2) and (b2) Detection process of cascaded system. Red, blue, and green boxes: bounding boxes predicted by Faster RCNN.

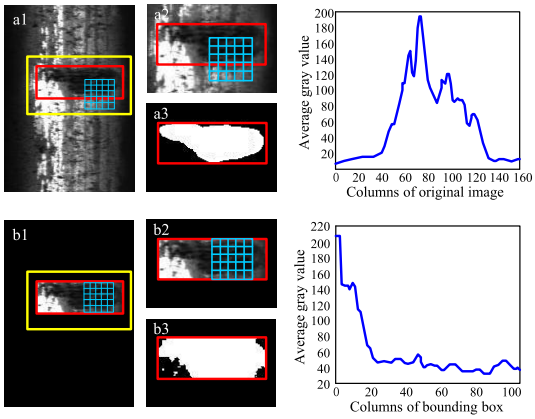

Fig. 12. Precision comparison of two systems. (a) Proposed DM-RIS. (b) Cascaded system. (a1) and (b1) Bounding boxes (red) in Faster RCNN. (a2) and (b2) zoomed-in view of the yellow boxes in (a1) and (b1). (a3) and (b3) Final results. Blue grid is FRGMM's window. Curves are vertical projection of global image and bounding box image, respectively.

B. Precision Comparison

In proposed DM-RIS, a 5×5 window of FRGMM performs through the whole global image, including defect edge and noise area, so that the surface defect is segmented with high precision and noise is filtered out. More details are shown in Fig. 12(a1)–(a3). In addition, the curve presents a narrow and high crest in columns [60,120], indicating a high contrast between defect and background.

In a cascaded system, the 5×5 window is restricted within the bounding box so that the local information on the edge is lost, making a rough, fuzzy, and noisy result. More details are shown in Fig. 12(b1)–(b3). In addition, the curve presents quite low and stable gray values in [20,110], indicating a very low contrast between defect and background, because defect almost has taken up the entire local image.

VI. Conclusion

In this paper, DM-RIS is proposed for complex and diverse rail surface where FRGMM generates the segmentation proposal and Faster RCNN carries out the objective location. In such a system, the two algorithms are well taken advantage of: 1) in order to obtain defect edge accurately, robustly, and fastly, we propose a new way to incorporate spatial information between neighboring pixels into the MRF-based GMM, so that the M-step of EM algorithm could be directly applied to prior distribution for the maximization of log-likelihood function; 2) to make it pervasive and robust environmentally, various rail scenes like weak illumination, inequality reflection, external noise, rust, and greasy dirt are fed into Faster RCNN so that our system meets different light, angle, background, and acquisition equipment. In the experiment, the proposed FRGMM achieves higher PSNR and precision (91.01%) and is 100–1000 times faster than other mixture models, the trained Faster RCNN achieves mean recognition accuracy (99.89%) and inference time (2.35 ms), and multi-model system performs well with overlap (95.18%), precision (96.74%), recall (94.13%), and speed (0.485 s/frame), which is, respectively, 7.94%, 2.13%, and 0.555 s/frame superior to previous work. The results demonstrate that RIS is highly effective and efficient in railway defects detection.

References

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

[2] X. Zhang, Y. Cui, Y. Wang, M. Sun, and H. Hu, "An improved AE detection method of rail defect based on multi-level ANC with VSS-LMS," *Mech. Syst. Signal Process.*, vol. 99, pp. 420–433, Jan. 2018.

[3] D. A. Ramatlo, D. N. Wilke, and P. W. Loveday, "Development of an optimal piezoelectric transducer to excite guided waves in a rail Web," *NDT E Int.*, vol. 95, pp. 72–81, Apr. 2018.

[4] A. Sabato and C. Niezrecki, "Feasibility of digital image correlation for railroad tie inspection and ballast support assessment," *Measurement*, vol. 103, pp. 93–105, Jun. 2017.

[5] S. Mariani, T. Nguyen, X. Zhu, and F. L. di Scalea, "Field test performance of noncontact ultrasonic rail inspection system," *J. Transp. Eng. A, Syst.*, vol. 143, no. 5, 2017, Art. no. 04017007.

[6] M. Molodova, Z. Li, A. Núñez, and R. Dollevoet, "Automatic detection of squats in railway infrastructure," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1980–1990, Oct. 2014.

[7] P. Salvador, V. Naranjo, R. Insa, and P. Teixeira, "Axlebox accelerations: Their acquisition and time–frequency characterisation for railway track monitoring purposes," *Measurement*, vol. 82, pp. 301–312, Mar. 2016.

[8] J. Wang, T. Li, Y.-Q. Shi, S. Lian, and J. Ye, "Forensics feature analysis in quaternion wavelet domain for distinguishing photographic images and computer graphics," *Multimedia Tools Appl.*, vol. 76, no. 22, p. 23721–23737, 2017.

[9] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015.

[10] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.

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

[12] S. F. Lu, Z. Liu, and Y. Shen, "Automatic fault detection of multiple targets in railway maintenance based on time-scale normalization," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 4, pp. 849–865, Apr. 2018.

[13] Z. Zhou, Y. Wang, Q. M. J. Wu, C.-N. Yang, and X. Sun, "Effective and efficient global context verification for image copy detection," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 48–63, Jan. 2017.

[14] H. Liu, J. Qin, F. Sun, and D. Guo, "Extreme kernel sparse learning for tactile object recognition," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4509–4520, Dec. 2017.

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

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

[16] E. Resendiz, J. M. Hart, and N. Ahuja, "Automated visual inspection of railroad tracks," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 751–760, Jun. 2013.
[17] M. Sun, X. Lin, Z. Wu, Y. Liu, Y. Shen, and N. Feng, "Non-destructive photoacoustic detecting method for high-speed rail surface defects," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2014, pp. 896–900.
[18] S. Grassie, "Squats and squat-type defects in rails: The understanding to date," *Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit*, vol. 226, no. 3, pp. 235–242, 2012.
[19] H.-M. Thomas, T. Heckel, and G. Hanspach, "Advantage of a combined ultrasonic and eddy current examination for railway inspection trains," *Insight-Non-Destructive Test. Condition Monitor.*, vol. 49, no. 6, pp. 341–344, 2007.
[20] B. Liang, S. Iwnicki, A. Ball, and A. E. Young, "Adaptive noise cancelling and time–frequency techniques for rail surface defect detection," *Mech. Syst. Signal Process.*, vol. 54, pp. 41–51, Mar. 2015.
[21] Z. He, Y. Wang, F. Yin, and J. Liu, "Surface defect detection for high-speed rails using an inverse PM diffusion model," *Sensor Rev.*, vol. 36, no. 1, pp. 86–97, 2016.
[22] C. Tastimur, M. Karaköse, E. Akin, and I. Aydin, "Rail defect detection with real time image processing technique," in *Proc. IEEE 14th Int. Conf. Ind. Inform. (INDIN)*, Jul. 2016, pp. 411–415.
[23] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.
[24] X. Gibert, V. M. Patel, and R. Chellappa, "Material classification and semantic segmentation of railway track images with deep convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 621–625.
[25] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter, "Deep convolutional neural networks for detection of rail surface defects," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2584–2589.
[26] J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 257–269, Feb. 2018.
[27] Z. Liu, L. Wang, C. Li, and Z. Han, "A high-precision loose strands diagnosis approach for isoelectric line in high-speed railway," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1067–1077, Mar. 2018.
[28] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout, "Steel defect classification with max-pooling convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–6.
[29] J. Zhong, Z. Liu, Z. Han, Y. Han, and W. Zhang, "A CNN-based defect inspection method for catenary split pins in high-speed railway," *IEEE Trans. Instrum. Meas.*, to be published.
[30] D. Soukup and R. Huber-Mörk, "Convolutional neural networks for steel surface defect detection from photometric stereo images," in *Proc. Int. Symp. Vis. Comput.*, Dec. 2014, pp. 668–677.
[31] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3708–3712.
[32] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, 2017.
[33] F.-C. Chen and M. R. Jahanshahi, "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392–4400, May 2018.
[34] S. J. Schmugge *et al.*, "Detection of cracks in nuclear power plant using spatial-temporal grouping of local patches," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–7.
[35] J. W. Palese and T. W. Wright, " Risk based ultrasonic rail test scheduling on Burlington Northern Santa Fe," *Permanent Way Inst. J. Rep.*, vol. 119, no. 2, pp. 164–178, 2001,
[36] D. F. Cannon, K.-O. Edel, S. L. Grassie, and K. Sawley, "Rail defects: An overview," *Fract. Eng. Mater. Struct.*, vol. 26, no. 10, pp. 865–886, 2003.
[37] N. O. Attoh-Okine, *Big Data and Differential Privacy: Analysis Strategies for Railway Track Engineering*. Hoboken, NJ, USA: Wiley 2017.
[38] H. Zhang, X. Jin, Q. M. J. Wu, Y. Wang, Z. He, and Y. Yang, "Automatic visual detection system of railway surface defects with curvature filter and improved Gaussian mixture model," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1593–1608, Jul. 2018.
[39] G. McLachlan and D. Peel, *Finite Mixture Models*. New York, NY, USA: Wiley, 2000.
[40] S. Sanjay-Gopal and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Trans. Image Process*, vol. 7, no. 7, pp. 1014–1028, Jul. 1998.
[41] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York, NY, USA: Springer, 2009.
[42] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. B, Methodol.*, vol. 36, no. 2, pp. 192–236, 1974.
[43] C. Nikou, N. P. Galatsanos, and C. L. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1121–1130, Apr. 2007.
[44] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 494–498, Mar. 2005.
[45] T. M. Nguyen and Q. M. J. Wu, "Fast and robust spatially constrained Gaussian mixture model for image segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 621–635, Apr. 2013.
[46] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, May 2015, pp. 91–99.
[47] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[48] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.
[49] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Proc. Eur. Conf. Comput. Vis.*, Apr. 2014, pp. 391–405.
[50] C. S. Won, "A block-based MAP segmentation for image compressions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 592–601, Sep. 1998.
[51] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognit.*, vol. 36, no. 1, pp. 131–144, Jan. 2003.
[52] L. Jonathan, S. Evan, and D. Trevor, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2015, pp. 3431–3440.
[53] V. Badrinarayanan, A. Kendall, and R. Cipolla, (2015). "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling." [Online]. Available: https://arxiv.org/abs/1505.07293
[54] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," vol. 20, pp. 1–28, Nov. 2001.
[55] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Feb. 2016, pp. 779–788.

Xiating Jin was born in Wenzhou, China, in 1995. He received the B.S. degree in automation from the Changsha University of Science and Technology, Changsha, China, in 2017. He is currently pursuing the master's degree with the College of Electrical and Information Engineering, Hunan University, Changsha.

His current research interests include machine learning, deep learning, and industry inspection.

Yaonan Wang received the Ph.D. degree in electrical engineering from Hunan University, Changsha, China, in 1994.

Since 1995, he has been a Professor with the College of Electrical and Information Engineering, Hunan University. From 1994 to 1995, he was a Post-Doctoral Research Fellow with the Normal University of Defense Technology, Changsha. From 1998 to 2000, he was supported as a Senior Humboldt Fellow by the Federal Republic of Germany, in the University of Bremen, Bremen, Germany. From 2001 to 2004, he was a Visiting Professor with the University of Bremen. His current research interests include robotics and image processing.

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.

Hui Zhang received the B.S., M.S., and Ph.D. degrees in pattern recognition and intelligent system from Hunan University, Changsha, China, in 2004, 2007, and 2012, respectively.

He is currently an Assistant Professor with the College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha, and also a Visiting Scholar with the CVSS Laboratory, Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. His current research interests include machine vision, sparse representation, and visual tracking.

Hang Zhong received the M.S. and B.S. degrees in automation from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree in control theory and application.

His current research interests include robotics modeling and control, visual servo control, and path planning of the aerial robots.

Li Liu was born in Changsha, China, in 1984. He received the B.S. degrees in measurement and control technology and instrument from Southeast University, Nanjing, China, in 2006. He is currently pursuing the Ph.D. degree with the College of Electrical and Information Engineering, Hunan University, Changsha, China.

His current research interests include robot vision measurement, robot path planning, and intelligent control.

Q. M. Jonathan Wu (M'92–SM'09) received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990.

Since 1995, he has been with the National Research Council of Canada, Ottawa, ON, USA, where he became a Senior Research Officer and a Group Leader. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He has authored more than 250 peer-reviewed papers in computer vision, image processing, intelligent systems, robotics, and integrated microsystems. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.

Dr. Wu has served on the Technical Program Committees and International Advisory Committees for many prestigious conferences. He is the Tier 1 Canada Research Chair in Automotive Sensors and Information Systems. He is currently an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the *Cognitive Computation*, and the *International Journal of Robotics and Automation*.

Yimin Yang (S'10–M'13) received the M.Sc. and Ph.D. degrees in electrical engineering from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2009 and 2013, respectively.

From 2014 to 2018, he was a Post-Doctoral Fellow with the University of Windsor, Windsor, ON, Canada. He is currently an Assistant Professor with the Computer Science Department, Lakehead University, Thunder Bay, ON, Canada. He has authored or coauthored more than 20 refereed papers. His current research interests include computer vision and robotics, machine learning.

Authorized licensed use limited to: HUNAN UNIVERSITY. Downloaded on August 03,2020 at 07:35:46 UTC from IEEE Xplore. Restrictions apply.