Optics and Lasers in Engineering 121 (2019) 397–405

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

A deep-learning-based approach for fast and robust steel surface defects classification

Guizhong Fu[a,b], Peize Sun[a,b], Wenbin Zhu[a,b], Jiangxin Yang[a,b], Yanlong Cao[a,b], Michael Ying Yang[c], Yanpeng Cao[a,b,*]

[a] State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China

[b] Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, School of Mechanical Engineering, Zhejiang University, Hangzhou, China

[c] Scene Understanding Group, University of Twente, Hengelosestraat 99, AE Enschede, 7514 The Netherlands

ARTICLE INFO

Keywords:
Surface inspection
Defect classification
Convolutional neural network
Feature extraction
Multi-receptive field

ABSTRACT

Automatic visual recognition of steel surface defects provides critical functionality to facilitate quality control of steel strip production. In this paper, we present a compact yet effective convolutional neural network (CNN) model, which emphasizes the training of low-level features and incorporates multiple receptive fields, to achieve fast and accurate steel surface defect classification. Our proposed method adopts the pre-trained SqueezeNet as the backbone architecture. It only requires a small amount of defect-specific training samples to achieve high-accuracy recognition on a diversity-enhanced testing dataset of steel surface defects which contains severe non-uniform illumination, camera noise, and motion blur. Moreover, our proposed light-weight CNN model can meet the requirement of real-time online inspection, running over 100 fps on a computer equipped with a single NVIDIA TITAN X Graphics Processing Unit (12G memory). Codes and a diversity-enhanced testing dataset will be made publicly available.

1. Introduction

Steel strips typically contain different categories of defects on the surface (e.g., crazing, inclusion, scratches, and rolled-in scale) due to the imperfect manufacturing process. These visually observable defects will cause changes in steel material properties such as corrosion resistance, wear resistance, and fatigue strength, and significantly decreases the quality of the final product [1,2]. The major objective of steel surface inspection is to accurately predict defect categories, providing important information for identifying and subsequently correcting causative factors [3]. Traditionally, the surface quality of steel strips is manually inspected by human experts. However, the manual inspection process is highly subjective, labor-intensive and too slow to facilitate real-time inspection tasks [3]. Thus, there is an urgent need to develop accurate and fully automatic machine vision-based inspection solutions as a critical component to guarantee steel products defect-free

In the last few decades, machine vision-based surface defect inspection technology has received extensive attention as a non-contact, non-destructive and fully automatic solution to assist or replace human inspectors [4–7]. The most commonly used approach is to firstly build the high-level representation of defects using various feature extraction techniques (e.g, Gabor filters [8,9], wavelet filters [10,11], contrast filters [12,13]), and then use classifiers (e.g., thresholding methods [5,14], support vector machines[10,15,16] and neural networks [17,18]) to determine the defect categories. However, it remains an unsolved problem to develop the optimal hand-crafted features to achieve high-accuracy recognition of steel surface defects which exhibit both “inter-class” similarity and “intra-class” diversity [1,3].

Recently, deep learning (DL) techniques have been successfully applied to solve many challenging computer vision tasks such as object classification [19–21] and target detection [22,23], and achieved breakthrough improvements. AlexNet is the first CNN model which is successfully implemented in Graphics Processing Unit (GPU) to accelerate the training and testing process in large-scale datasets [24]. Simonyan et al. proposed to make use of 3 × 3 convolutional filters to construct a very deep CNN model VGG [25], which provides a backbone architecture to facilitate other computer vision tasks. He et al. proposed a residual learning framework ResNet to ease the training process of deep CNN models and achieved significant performance improvement [26]. However, these deep CNN models contain a large number of parameters and cannot deliver real-time speed. In an attempt to make a good balance between performance and efficiency, Iandola et al. proposed a compact

[*] Corresponding author at: State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China.

E-mail address: caoyp@zju.edu.cn (Y. Cao).

https://doi.org/10.1016/j.optlaseng.2019.05.005
Received 2 February 2019; Received in revised form 3 May 2019; Accepted 6 May 2019
Available online 15 May 2019

SqueezeNet CNN architecture which produces high-accuracy recognition with significantly fewer parameters [27].

Inspired by the recent success of CNN models for object recognition, some researchers attempt to adopt deep convolutional neural network (CNN) architectures to solve the surface defects classification problem, directly learning defect-related features from a number of training samples instead of being hand-crafted before. Li et al. designed a simple end-to-end CNN model (consists of 7 layers) for steel surface defects recognition [28]. The parameters of this model are randomly initialized and trained from scratch using a small defect-specific image dataset, thus its performance is not satisfactory. Ren et al. proposed a deep-learning-based approach for automatic and high-accuracy surface defect inspection [29]. However, they directly apply the pre-trained Decaf model [30] for feature extraction which is sub-optimal to depict defect-specific characteristics. It is not a trivial task to develop deep-learning-based steel surface defects recognition approaches working reliably in real-world inspection situations. The major challenge is three-folder. First, steel surface defects occur in low probability and are not visually distinctive during the inspection process, therefore it is difficult and time-consuming to build a large scale defect dataset to train deep CNN models. Second, the appearance of steel surface defects depends on the setup of the image acquisition system [31], thus the captured images typically contain severe non-uniform illumination, camera noise, and motion blur. Third, deep CNN models typically contain a large number of network parameters and require heavy computational loads [24–26]. Even with GPU acceleration, the running time is still not suitable for real-time detection tasks.

To overcome the problems mentioned above, we propose a compact yet effective CNN model to achieve fast and robust steel surface defect classification. Our proposed method adopts the pre-trained SqueezeNet as the backbone architecture and only requires hundreds of defect-specific training samples to fine-tune the model and achieve accurate recognition results. We set higher learning rates for shallower layers to emphasize the fine-tuning of low-level features in the pre-trained model to better characterize steel surface defects. Moreover, we incorporate a multi-receptive field (MRF) module to generate scale-dependent high-level features for accurate classification of surface defects of various sizes. Evaluated on a diversity-enhanced testing dataset of steel surface defects, our proposed defect recognition model remain very robust against non-uniform illumination, severe CCD (Charge-coupled device) camera noise, and motion blur, which inevitably occur during the on-site image acquisition process. Our proposed light-weight CNN model runs over 100 fps on a PC equipped with a single NVIDIA TITAN X GPU (12G memory) and can meet the requirement of real-time online defect inspection tasks. Overall, the contributions of this paper are summarized as follows:

- We propose an end-to-end SqueezeNet-based model, which is capable of learning defect-related features from a number of training samples instead of being hand-crafted before. Moreover, the proposed method only requires a small amount of defect-specific training samples to fine-tune the pre-trained model to achieve accurate steel surface defects recognition results.
- We present two effective techniques to improve defect recognition accuracy of our proposed CNN model. First, we emphasize the fine-tuning of low-level features in the pre-trained model to better characterize steel surface defects. Second, we incorporate multiple receptive fields to improve the scale invariance of high-level features for accurate classification of defects of various sizes.
- Our method achieves significantly higher recognition accuracy compared with the state-of-the-art defect classifiers [1,28,29,32,33], particularly when the testing images contain severe non-uniform illumination, camera noise, and motion blur. Moreover, this light-weight model can process more than 100 images per second on a single NVIDIA Geforce Titan X GPU to facilitate real-time inspection tasks.

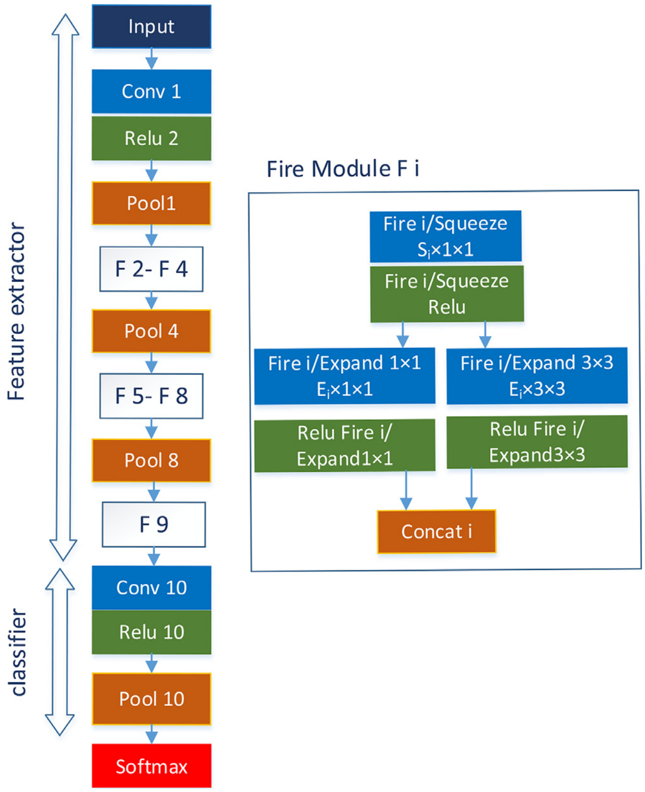

Fig. 1. The architecture of the pre-trained SqueezeNet model [27].

The remainder of our paper is structured as follows. Our proposed SqueezeNet-based model and two improving techniques are presented in Section 2. Details of a diversity-enhanced testing dataset of steel surface defects are provided in Section 3. An extensive evaluation of our method and experimental comparison with state-of-the-art alternatives are provided in Section 4. We conclude our paper in Section 5.

2. Proposed method

In this section, we proposed an end-to-end SqueezeNet-based model for steel surface defect recognition. Moreover, we present two effective techniques to improve defect recognition accuracy including (1) emphasizing the fine-tuning of low-level features and (2) incorporating a MRF module.

2.1. Squeezenet-based defect classification

Many researchers attempted to achieve higher recognition accuracy by either increasing the depth of the network or adopting more complex architectures [25,26,34]. However, these deep CNN models contain a large number of network parameters and cannot produce real-time processing speed (>30fps) even through GPU acceleration, which are not suitable for high-speed online steel strip surface inspection tasks (e.g., the rolling speed of the steel strip is up to 18 m/s [1]). To achieve a good balance between recognition accuracy and computational efficiency, Iandola et al. proposed the SqueezeNet CNN architecture to produce high-accuracy recognition on ImageNet with significantly fewer parameters [27]. Using such a light-weight CNN architecture has several advantages including more efficient model training, less prone to small dataset over-fitting, and feasible deployment in embedded systems (e.g., field-programmable gate array).

We adopt the pre-trained SqueezeNet 1.0 model as the backbone architecture, which contains nine fire modules as illustrated in Fig. 1. The i^{th} fire module contains a squeeze convolution layer (using S_i 1×1

Table 1
The detailed configurations of individual layers/modules in the baseline SqueezeNet model (SDC-SN-baseline) for defect classification on the NEU benchmark dataset. The filter parameters are indicated as $C \times W \times L$ where C is the channel number, W is the kernel width, and L is the kernel length.

Layers	Output size	Filter size	Depth
Input	224 × 224×3	–	
Conv 1	109 × 109×96	96 × 7×7, stride 2	1
Pool 1	54 × 54×96	3 × 3 Pooling, stride 2	0
Fire 2	54 × 54×128	16 × 1×1,64 × 1×1,64 × 3×3	2
Fire 3	54 × 54×128	16 × 1×1,64 × 1×1,64 × 3×3	2
Fire 4	54 × 54×256	32 × 1×1,128 × 1×1,128 × 3×3	2
Pool 4	27 × 27×256	3 × 3 Pooling, stride 2	0
Fire 5	27 × 27×256	32 × 1×1,128 × 1×1,128 × 3×3	2
Fire 6	27 × 27×384	48 × 1×1,192 × 1×1,192 × 3×3	2
Fire 7	27 × 27×384	48 × 1×1,192 × 1×1,192 × 3×3	2
Fire 8	27 × 27×512	64 × 1×1,256 × 1×1,256 × 3×3	2
Pool 8	13 × 13×128	3 × 3 Pooling, stride 2	0
Fire 9	13 × 13×512	64 × 1×1,256 × 1×1,256 × 3×3	2
Conv 10	13 × 13×6	6 × 13×13, stride 1	1
Pool 10	1 × 1×6	13 × 13 Pooling	0

filters) and an expand layer (using E_i 1×1 filters and E_i 3×3 filters). The squeeze layers are designed to have fewer channels than the expand layers ($S_i < 2\times E_i$) to limit the number of input channels to the 3×3 filters. To utilize the SqueezeNet model for steel surface defect classification task, we need to modify the number of output channels in the Conv-10 layer. More specifically, we set the number of channels in the Conv-10 layer to six for the defect classification on the Northeastern University (NEU) benchmark dataset of steel surface defects (it contains six categories of surface defects existing in hot-rolled steel strip including crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches). A global average pooling (GAP) layer is used to replace the full connected layer, which is commonly adopted in many CNN based classification architectures including AlexNet [24] and VGG [25], to further reduce the number of model parameters. The GAP layer computes the average over the 13×13 slices to generate $1\times1\times6$ tensors. The detailed configurations of individual layers/modules in this baseline SqueezeNet model for surface defect classification (SDC-SN-baseline) on the NEU benchmark are shown in Table 1.

The parameters of steel surface defect classifier are updated by minimizing a multi-class loss function which is defined as

$$\mathcal{L} = -\sum_{k=1}^{6} t_k \log \Pr(y = k), \tag{1}$$

where $t_k = 1$ when the ground-truth label of an input image is k, else $t_k = 0$. $\Pr(y = k) \in [0, 1]$ is the confidence score which is calculated utilizing the softmax function as

$$\Pr(y = k) = \frac{e^{G_k}}{\sum_{j=1}^{6} e^{G_j}}, \tag{2}$$

where G_k denotes the channel-k output of GAP layer. Note that the confidence score $\Pr(y = k)$ predicts the existence of a defect of k*th* category in an image (e.g., the NEU benchmark dataset contains six categories of surface defects om the surface of hot-rolled steel strip).

2.2. Model optimization

After building the baseline SqueezeNet model for steel surface defect classification, we further present two effective techniques to improve recognition accuracy. First, higher learning rates are set for shallower layers, fine-tuning the low-level features to better characterize texture-related defects. Second, a MRF module is utilized to improve the distinctness of high-level features for accurate classification of surface defects of various sizes.

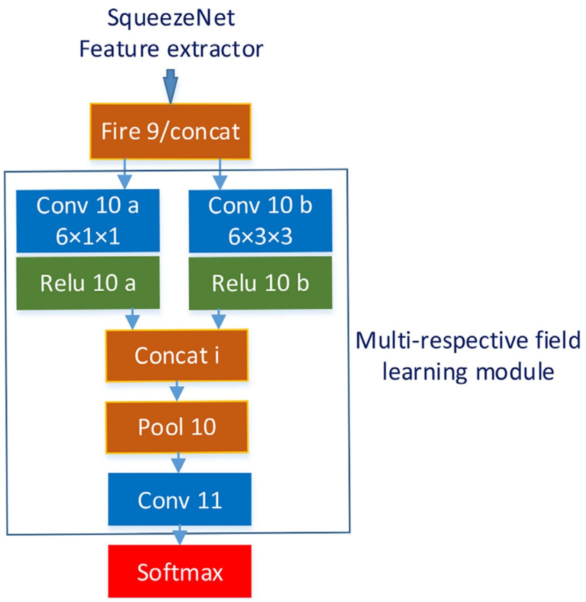

Fig. 2. The architecture of the proposed multi-respective field module.

The SqueezeNet model is pre-trained on the ImageNet dataset which contains more than 1,300,000 images of 1000 common object categories [21]. To achieve accurate steel surface defect recognition results, it is important to fine-tune the parameters of SqueezeNet model using defect-specific training samples. It is noted that the steel surface defects are typically presented as local abnormalities in texture patterns [29]. As a result, images of common objects (e.g., bird, flower, person and vehicle) and steel surface defects (e.g., rolled-in scale, patches, crazing, and scratches) present very different low-level features, which are extracted using the shallower layers in CNN architectures. Instead of using the same learning rate for all different layers in a CNN model [35,36], we propose to use higher learning rates for shallower layers to emphasize the fine-tuning of low-level features in the pre-trained model to better characterize texture-related features of surface defects. More specifically, the weight and bias learning rates of shallower layers/modules including Conv-1, Fire module 2 and Fire module 3 are set to 9, while learning rates of deeper layers/modules are equally set to 1.

It is observed that different types of steel surface defects present significant variations in appearance such as size and shape. For instance, scratches are typically shown as locally-defined defects while crazing defects will cause abnormal texture patterns in a large image region. Based on this observation, we proposed to incorporate a MRF module to improve the scale-invariance of high-level features for accurate classification of surface defects of various sizes. In CNN models, receptive field refers to the receptive range in the current convolutional layer to infer information for the following layers, determining how many neighboring pixels are considered to perform recognition tasks. The receptive field RF_i of the i^{th} convolutional layer is calculated as

$$RF(i) = RF(i-1) + kernel_{size}(i-1) \times stride(i), \tag{3}$$

$$stride(i) = \prod_{i=1}^{i-1} s(i), \tag{4}$$

where $kernel_{size}(i)$ is the kernel size of convolutional layer i, and $stride(i)$ is the accumulated stride number which is calculated by multiplying the stride values of proceeding layers before layer i. As shown in Fig. 2, we add a MRF module after Fire module 9. The outputs of two layers with different receptive fields (Conv 10a and Conv 10b) are concatenated to capture multi-scale information and generate distinctive high-level features for accurate surface defect classification.

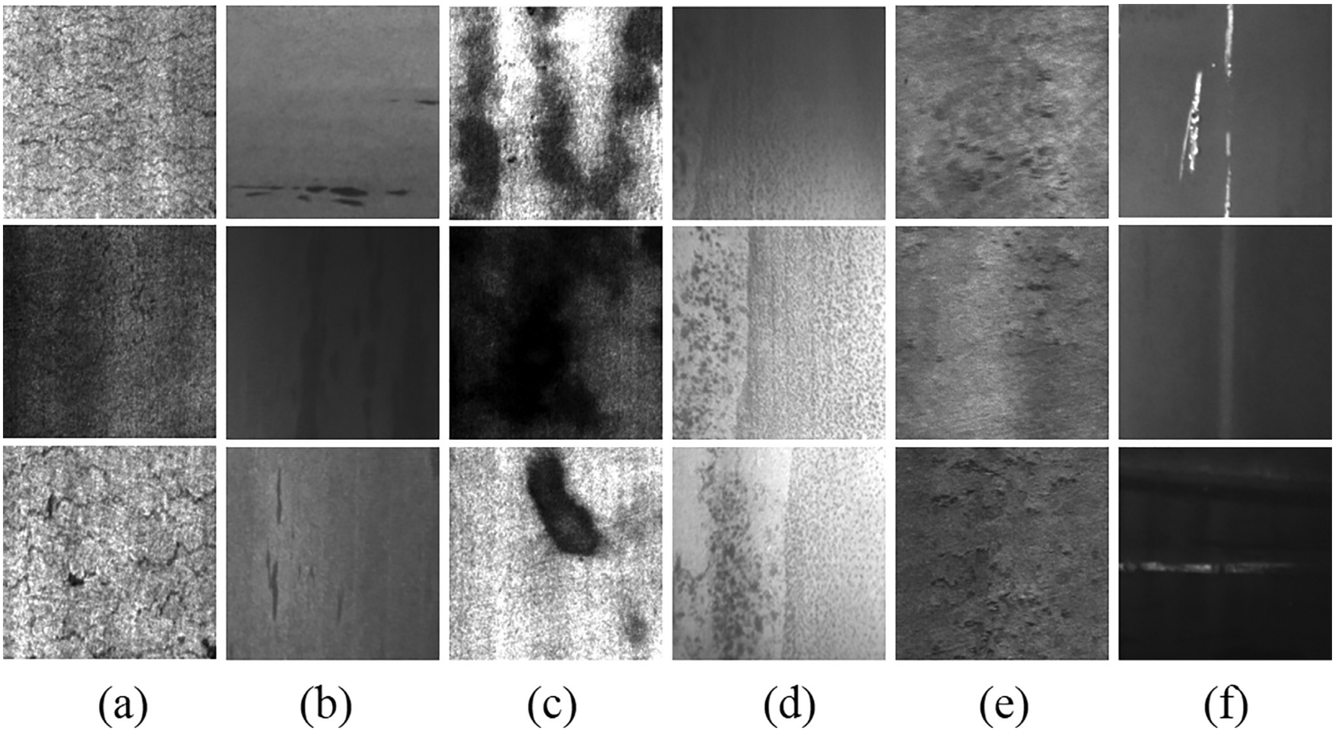

Fig. 3. Samples images of six typical surface defects in the NEU surface defect database including (a) Crazing; (b) Inclusion; (c) Patches; (d) Pitted surface (e) Rolled-in scale; (f) Scratches.

To sum up, we present two effective techniques to improve defect recognition accuracy including (1) incorporating a MRF module and (2) emphasizing the fine-tuning of low-level features in the pre-trained model. The effectiveness of the proposed techniques is systematically evaluated in Section 4.2.

3. Steel surface defect dataset

Based on the publicly available NEU steel surface defects benchmark dataset, we further present a diversity-enhanced testing dataset to evaluate the robustness of various defect classification models against non-uniform illumination, sensor noise, and camera motion blur, which often occur during the on-site image acquisition process.

3.1. Training database

The NEU surface defect database contains six categories of surface defects of the hot-rolled steel strip including crazing (Cr), inclusion (In), patches (Pa), pitted surface (PS), rolled-in scale (RS), and scratches (Sc) [1]. In total, the database consists of 1800 200×200 grayscale images (300 samples for each surface defect category). Sample images of some typical defects are shown in Fig. 3. It is observed that the "intra-class" defects show large differences in appearance while "inter-class" defects sometimes present similar characteristics. We randomly select 80% of images (240 images for each defect category) in the NEU dataset to form the training data. Here we did not apply any data augmentation technique to increase the size of training data in an attempt to evaluate whether our proposed model works well using a small training dataset. The rest NEU images are utilized to generate a diversity-enhanced testing dataset.

3.2. Diversity-enhanced testing dataset

The visual appearance of steel surface defects is affected by the setup of the image acquisition system [28]. For instance, non-uniform illumination will cause changes of intensity values in the acquired defect images. Also, a variety of noise sources exist in image data acquired by a CCD camera [37]. Moreover, multiple sources of vibrations in the production line and fast rolling speed of the steel strips may cause motion blurring effects in the captured images. The disturbances mentioned above inevitably increase the difficulty of defect recognition task. To evaluate the robustness of the proposed defect classifiers, we construct a diversity-enhanced testing dataset of steel surface defects which contains severe non-uniform illumination, camera noise, and motion blur.

Non-uniform illumination: During the on-site image acquisition process, it is difficult to set up light sources which provide perfect illumination to make the defects visible. Therefore, it is important to evaluate the performance of defect classifiers under various illumination conditions. In our experiments, a 3rd order polynomial model is used to generate slow-varying intensity bias fields [38], which are further combined with the original images to simulate images with various non-uniform illumination effects. We set the luminance range α of bias fields to ± 0.4 and ± 1. For each intensity range, we obtain 1800 images with obvious spatially-varying intensity distortions (300 for each defect category).

Camera noise: Following the research work presented by Song et al. [1], we add Gaussian noise to the original image to simulate a number of noisy images with different signal-to-noise ratio (SNR) values (20dB and 35dB). Here SNR is defined as

$$SNR = 10\log_{10}\left[\frac{\sum_{i=1}^{M}\sum_{j=1}^{N} g(i,j)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N} r(i,j)^2)}\right], \quad (5)$$

where M and N indicate the width and height of input image, respectively. $g(i, j)$ and $r(i, j)$ are the intensity value of original signal and Gaussian noise at pixel (i, j). For each SNR value, we randomly generate 1800 noisy images (300 for each defect category).

Motion blur: Vibrations are inevitable during the steel manufacturing process. Moreover, the rolling speed of the steel strip is fast (up to 18 m/s [1]). Therefore, the captured defect images typically contain obvious motion blurring effects. The blurred image is computed by convolving the original image with a motion filter which approximates

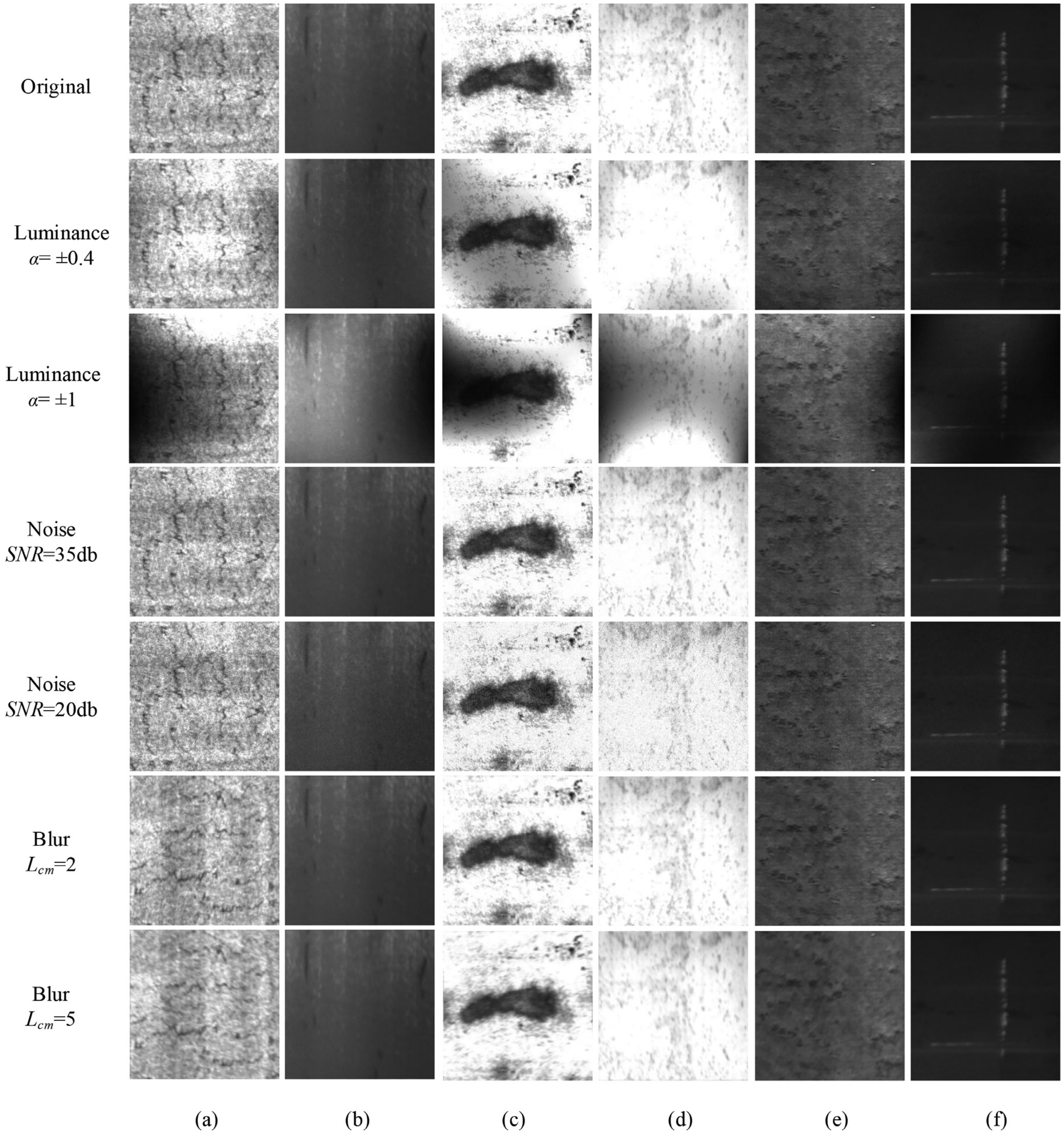

Fig. 4. Sample images of six typical surface defects in both the NEU dataset and a diversity-enhanced testing dataset including (a) Crazing; (b) Inclusion; (c) Patches; (d) Pitted surface (e) Rolled-in scale; (f) Scratches. Note we add severe disturbances such as non-uniform illumination (Luminance $\alpha = \pm 0.4$ and $\alpha = \pm 1$), camera noise (Noise SNR=20db and SNR=35db), and motion blur (Motion L_{cm}=2 and L_{cm}=5) to images in the diversity-enhanced testing dataset.

camera trajectory. We set the length of camera motion L_{cm} to 2 and 5. For each motion length, we generate 1800 motion blurred images using random motion angles between 0° and 360° (300 images for each defect category).

In experiments, defect classifiers are trained using original images in the NEU dataset, while evaluated on images in both the NEU dataset and a diversity-enhanced testing dataset (some sample images are shown in Fig. 4). The constructed diversity-enhanced testing dataset will be made publicly available in the future.

4. Experiments

4.1. Implementation details

Ou r method is implemented using the publicly available Caffe framework. We use 80% of images (240 images for each defect category) in the NEU dataset as the training data. Performance evaluation is done on images in both the NEU dataset and the diversity-enhanced testing dataset. The convolutional layer parameters are initialized using

Table 2
The classification accuracy(%) of different SqueezeNet-based CNN models on images in both the NEU dataset and a diversity-enhanced testing dataset.

Method	Original	Luminance		Noise		Blur	
		$\alpha = \pm 0.4$	$\alpha = \pm 1$	SNR=20db	SNR=35db	L_{cm}=2	L_{cm}=5
SDC-SN-baseline	99.7	96.5	84.9	82.9	99.4	99.7	73.5
SDC-SN-ELF	100.0	99.7	89.7	92.9	99.9	100.0	87.6
SDC-SN-ELF+MRF	100.0	99.9	92.9	94.8	100.0	100.0	97.2

Table 3
The classification accuracy(%) of various state-of-the-art steel surface defect classifiers on images in both the NEU dataset and a diversity-enhanced testing dataset. The top three results are highlighted in **red**, *green*,and ***blue***, respectively.

Method	Original	Luminance		Noise		Blur	
		$\alpha = \pm 0.4$	$\alpha = \pm 1$	SNR=20db	SNR=35db	L_{cm}=2	L_{cm}=5
GLCM+SVM [32]	88.1	69.0	43.8	67.8	87.2	54.6	51.2
GLCM+NNC [32]	89.7	68.7	46.1	***71.0***	89.2	73.7	56.7
GLCM+MLR [32]	94.7	55.7	45.3	52.3	92.0	31.8	25.9
AELTP+SVM [33]	76.1	31.4	26.3	46.6	70.5	51.2	44.6
AELTP+NNC [33]	96.4	47.9	47.9	64.4	***92.7***	54.5	37.6
AELTP+MLR [33]	98.6	44.4	44.4	48.8	82.6	46.3	29.3
AECLBP+SVM [1]	***98.9***	47.9	42.9	39.9	74.6	70.2	53.2
AECLBP+NNC [1]	98.3	41.3	33.2	42.7	80.8	67.7	47.9
AECLBP+MLR [1]	98.3	68.4	45.8	43.7	77.2	85.5	72.4
ETE [28]	95.8	***92.8***	*84.3*	47.6	84.6	***93.4***	***79.1***
DECAF+MLR [29]	*99.7*	*98.9*	***79.6***	*89.7*	*99.7*	*99.7*	*79.2*
SDC-SN-ELF+MRF	**100.0**	**99.9**	**92.9**	**94.8**	**100.0**	**100.0**	**97.2**

Xavier's method [39]. The batch size is set to 18. We use "step" learning policy, setting the base learning rate to 0.01, gamma to 0.1, step size to 800, momentum to 0.9 and $iter_{size}$ to 10. Clip gradient is set to 35 to avoid gradient exploding. Our model is trained using Adaptive Moment Estimation (Adam) [40]. Due to the compactness of our proposed method, it only takes about 20 minutes to complete the training processing using a single NVIDIA TITAN X GPU (12G memory). In the testing phase, our model outputs six confidence scores corresponding to six categories of surface defects in the NEU dataset. The defect category with the highest confidence score is predicted as the classification result.

4.2. Performance analysis

In Section 2.2, we present two different techniques to improve defect recognition accuracy of our proposed SDC-SN-baseline model (the detailed configurations of SDC-SN-baseline model are provided in Fig. 1 and Table 1). In this section, we set up ablation experiments to evaluate their effectiveness.

Since the SqueezeNet model is pre-trained on images of common objects (e.g., bird, flower, person, and vehicle), it is required to update its parameters based on training images of specific surface defects (e.g., rolled-in scale, patches, crazing, and scratches). Here we consider two alternatives for parameter fine-tuning including (1) SDC-SN-ELF–setting higher learning rates for shallower layers/modules including Conv-1, Fire module 2 and Fire module 3 to emphasize the fine-tuning of low-level features; (2) SDC-SN-baseline–setting equal learning rates for all different layers in a CNN model. The comparative results (classification accuracy) are illustrated in Table 3. It is observed that using higher learning rates to emphasize the fine-tuning of low-level features is an effective technique to update the pre-trained SqueezeNet model for steel surface defect classification, leading to higher recognition accuracy without incurring additional computational costs. Such improvement is particularly evident for images with severe noise disturbances (e.g., the recognition accuracy significantly increases from 82.9% to 92.8% on the Noise SNR = 20db dataset).

Moreover, we evaluate the effectiveness of adding a MRF module to the SqueezeNet architecture as illustrated in Fig. 2. The experiments are conducted based on the SqueezeNet-based model which employs higher learning rates for low-level feature fine-turning (SDC-SN-ELF). Table 3 illustrates the experimental results (classification accuracy and running time) on images in both the NEU dataset and the diversity-enhanced testing dataset. It is experimentally proven that adding a MRF module in the SqueezeNet-based model (SDC-SN-ELF + MRF) can capture multi-scale information to improve the distinctiveness of high-level features and lead to further improvement of classification accuracy. For instance, the recognition accuracy is significantly increased from 87.6% to 97.2% on the Blur L_{cm}=5 dataset. A reasonable explanation for this phenomenon is that the MRF module can cover a larger image region to learn more distinctive features in blurred images. It is worth mentioning that the architecture of the proposed MRF module is simple (containing 36,876 parameters in total), thus only a small amount of computational overhead is added in our SDC-SN-ELF + MRF module. The light-weight SDC-SN-ELF + MRF model can process over 100 fps (on 200 × 200 images) on a computer equipped with a single NVIDIA TITAN X GPU (12G memory) which is adequate to facilitate online inspection tasks.

4.3. Comparisons with state-of-the-arts

We compare our proposed model (SDC-SN-ELF + MRF) with a number of state-of-the-art steel surface defect recognition methods [1,28,29,32,33]. For the traditional machine-learning-based approaches, we consider three different feature extraction techniques including gray level co-occurrence matrix (GLCM) [32], adaptive extended local ternary pattern (AELTP) [33], and adjacent evaluation completed local binary patterns (AECLBP) [1], three feature classifiers including support vector machine (SVM), Nearest neighbor clustering (NNC) and multiple linear regression (MLR). Source codes or pre-trained models of these feature extractors and classifiers are publicly available. Moreover, we consider two deep-learning-based surface defect classification approaches including the end-to-end CNN model (ETE) proposed

Table 4
Number of training images, model size, running time, and recognition accuracy on the original NUE and the diversity-enhanced datasets of our proposed SDC-SN-ELF + MRF model and other Deep-learning based approaches [28,29].

Method	Number of training images	Running time	Model size	Accuracy on the NEU dataset	Accuracy on the enhanced dataset
ETE [28]	11,520	5.3ms	1.9 MB	95.8%	80.3%
DECAF+MLR [29]	1440	10.3 ms	244 MB	99.7%	91.3%
SDC-SN-ELF+MRF	1440	8.0 ms	3.1 MB	100%	97.5%

by Li et al. [28] and the Decaf model-based approach (DECAF + MLR) proposed by Ren et al. [29], which are re-implemented and trained according to the original papers. To train the build-from-scratch ETE model, we rotate (90°, 180°, 270°) and flip (horizontally) images to expand training dataset. Other defect classifiers are trained using 1440 original images (240 images of each defect category) in the NEU dataset without any data augmentation techniques.

In Table 3, we show quantitative evaluation results on images in both the NEU dataset and a diversity-enhanced testing dataset. We made three important observations. First of all, although classifiers based on hand-crafted features [1,32,33] achieved good results (>95%) on images in the NEU dataset (training and testing using images captured under similar conditions), their performances drop significantly on images with non-uniform illumination, sensor noise, and motion blur. These undesired disturbances commonly exist during the on-site image acquisition process. In comparison, deep-learning-based methods generally achieved higher recognition accuracy and remain more robust against the disturbances mentioned above. Our experimental results illustrate that the learned features significantly outperform hand-crafted ones on heterogeneous and unfamiliar (unseen) datasets [41]. Second, the build-from-scratch ETE model does not perform well even using an augmented training dataset. For instance, ETE model even performs worse than some methods based on hand-crafted features on images in the NEU dataset. The underlying reason is that its parameters are randomly initialized and cannot be adequately fine-tuned using a small defect-specific image dataset. Since steel surface defects occur in low probability and are not visually distinctive during the inspection process, it is difficult and time-consuming to build a large scale defect dataset to train a CNN model from scratch. Therefore, it is recommended to build a classifier based on a pre-trained model (e.g., SqueezeNet [27]) and fine-tune its parameters for defect-specific recognition tasks. Finally, our proposed light-weight SDC-SN-ELF + MRF model achieves higher recognition accuracy compared with the Decaf model-based approach (DECAF + MLR). We set higher learning rates for the shallower layers to emphasize the fine-tuning of the low-level features, better characterizing texture-related defects. Also, we add a MRF module to learn more distinctive features using information in multi-scale image patches. In comparison, DECAF + MLR directly applies the pre-trained Decaf model [30] for feature extraction which are sub-optimal to depict defect-specific characteristics. The number of training images, running time, model size, and recognition accuracy on the original NUE and the diversity-enhanced datasets of three deep learning-based approaches are shown in Table 4. Although ETE model [28] has a smaller size and runs faster than our SDC-SN-ELF + MRF model (ETE - 5.4 ms vs. SDC-SN-ELF + MRF - 8.0 ms), its recognition accuracy is significantly lower (4.2% lower on the NEU dataset and 17.2% lower on the diversity-enhanced dataset). Moreover, ETE model uses more training images (ETE - 11,520 frames vs. SDC-SN-ELF + MRF - 1440 frames). Compared with the DECAF + MLR method, our method achieves 0.3% higher accuracy on the NEU dataset and 6.2% on the diversity-enhanced dataset using significantly fewer parameters (DECAF + MLR - 230MB vs. SDC-SN-ELF + MRF - 3.0MB). The results confirm that our SDC-SN-ELF + MRF model can obtain both good accuracy and fast running speed.

Table 5
The confusion matrix of our SDC-SN-ELF + MRF model on different subsets of the diversity-enhanced testing dataset which contains six defect categories including crazing (Cr), inclusion (In), patches (Pa), pitted surface (PS), rolled-in scale (RS), and scratches (Sc).

			(a) Luminance $\alpha = \pm 0.4$			
	Cr	In	Pa	PS	RS	Sc
Cr	299	0	0	1	0	0
In	0	300	0	0	0	0
Pa	0	0	300	0	0	0
PS	0	1	0	299	0	0
RS	0	0	0	0	300	0
Sc	0	0	0	0	0	300
			(b) Luminance $\alpha = \pm 1$			
	Cr	In	Pa	PS	RS	Sc
Cr	221	0	13	66	0	0
In	0	295	2	3	0	0
Pa	0	0	299	1	0	0
PS	0	3	18	279	0	0
RS	0	3	4	9	284	0
Sc	0	6	0	0	0	294
			(c) Noise SNR=20db			
	Cr	In	Pa	PS	RS	Sc
Cr	300	0	0	0	0	0
In	0	206	0	94	0	0
Pa	0	0	300	0	0	0
PS	0	0	0	300	0	0
RS	0	0	0	0	300	0
Sc	0	0	0	0	0	300
			(d) Noise SNR=35db			
	Cr	In	Pa	PS	RS	Sc
Cr	300	0	0	0	0	0
In	0	300	0	0	0	0
Pa	0	0	300	0	0	0
PS	0	0	0	300	0	0
RS	0	0	0	0	300	0
Sc	0	0	0	0	0	300
			(e) Blur L_{cm}=2			
	Cr	In	Pa	PS	RS	Sc
Cr	300	0	0	0	0	0
In	0	300	0	0	0	0
Pa	0	0	300	0	0	0
PS	0	0	0	300	0	0
RS	0	0	0	0	300	0
Sc	0	0	0	0	0	300
			(f) Blur L_{cm}=5			
	Cr	In	Pa	PS	RS	Sc
Cr	260	0	0	15	25	0
In	0	300	0	0	0	0
Pa	0	0	300	0	0	0
PS	0	2	0	298	0	0
RS	0	0	0	7	293	0
Sc	0	0	0	0	0	300

To systematically investigate the classification results of six different defect categories, we calculate the confusion matrix of our SDC-SN-ELF + MRF model on different subsets of the diversity-enhanced testing dataset in Table 5. In each subset, 1800 defect samples (300 for each defect category) are used for evaluation. In this confusion matrix, the first column indicates the ground-truth defect categories and the numbers in each row record predictions of our model. Note all correct predictions should be recorded in the diagonal cells of the confusion table. Overall, our method can achieve high-accuracy recognition results when images contain moderate disturbances. As expected, the accuracy rates drop on images with severe camera noise, non-uniform illumination, and motion blur. In the Luminance $\alpha = \pm 1$ subset, 66 crazing defects are

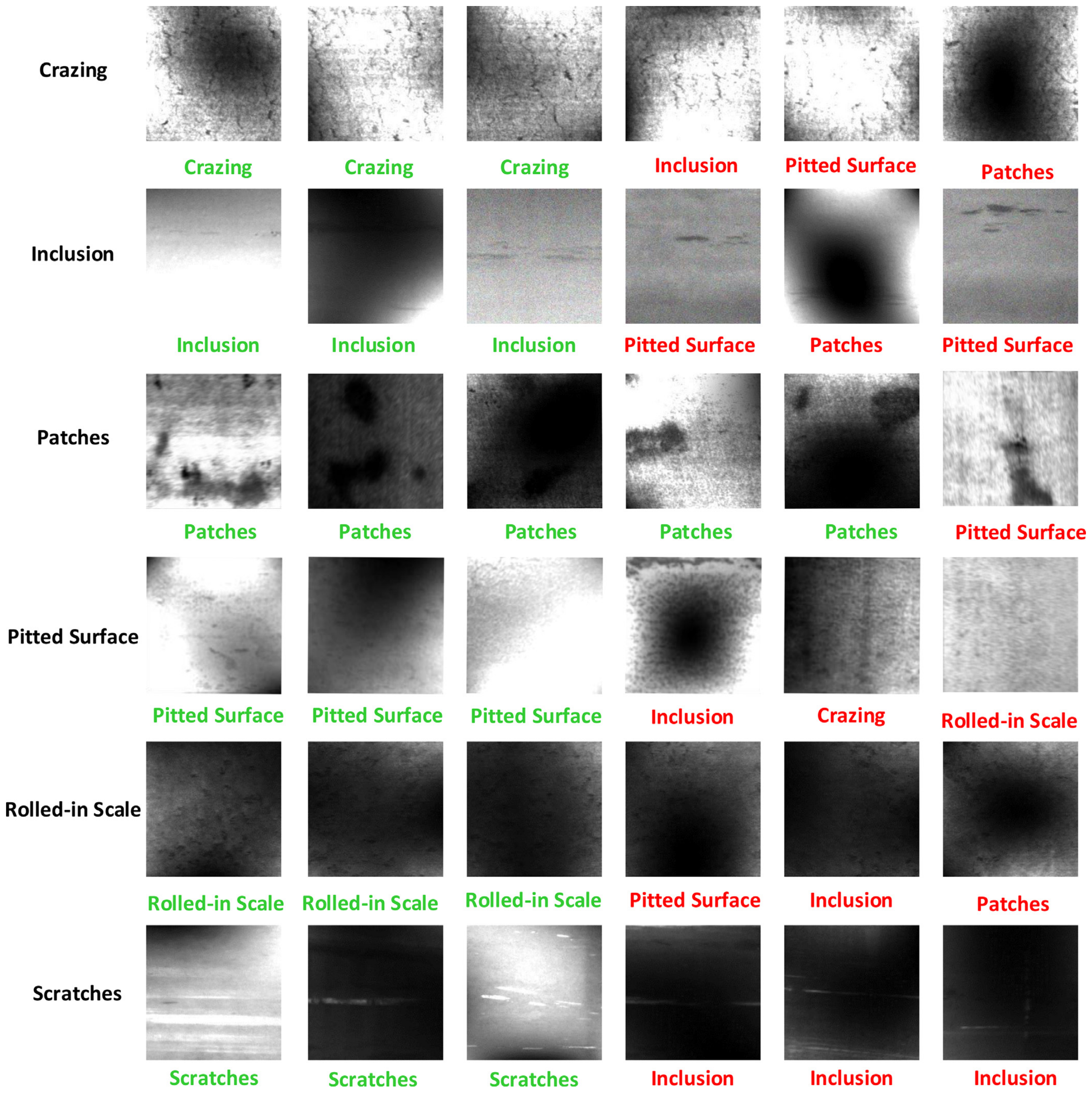

Fig. 5. Some correctly and wrongly classified images in the diversity-enhanced testing dataset of steel surface defects. Results highlighted in *green* and **red** indicate correct and incorrect classification results, respectively.

misclassified as pitted surfaces. In the Noise SNR=20db subset, 94 inclusion defects are misclassified as pitted surfaces. In the Blur L_{cm}=5 subset, 40 crazing samples are misclassified as pitted surface or rolled-in scale defects due to “inter-class” similarity and “intra-class” diversity in defect categories [1]. Some correctly and wrongly classified defect samples are shown in Fig. 5. In the future, we plan to improve the robustness of our approach further to achieve higher recognition accuracy on images containing severe camera noise, non-uniform illumination, and motion blur. It is worth mentioning that the pre-train SqueezeNet model can be easily modified to facilitate classification task with more defect categories. For instance, when applied on a dataset with n defect classes, the channel number of the last convolutional layer *Conv*10 in the SqueezeNet model is set to n accordingly.

5. Conclusion

In this paper, a fast and robust SqueezeNet-based model is proposed for steel surface defects classification which is very important to steel strip production and quality control. Based on the pre-trained SqueezeNet, we recommend to emphasize the learning of low-level features and add a MRF module. As a result, our method only requires a small amount of defect-specific training samples to achieve accurate defect recognition. We also construct a diversity-enhanced testing dataset of steel surface defects to evaluate the robustness of classification models. Extensive experimental results on images containing severe camera noise, non-uniform illumination, and motion blur are provided. It is observed that our proposed method achieves significantly higher recognition accuracy compared with the state-of-the-art steel surface defect classifiers. Moreover, our proposed light-weight model achieves real-time speed, processing more than 100 image patches (200 × 200) per second on a single NVIDIA TITAN X GPU (12G memory), thus it can be utilized for online steel production inspection tasks. In the future, we plan to develop a complete steel surface defect diagnosis framework, which consists of image acquisition, data pre-processing, defect region detection, and defect category classification, to perform fully automatic quality control of steel strip production.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (Nos. 51605428, 51575486, and U1664264).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2019.05.005.

References

[1] Song K, Yan Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Appl Surf Sci 2013;285(21):858–64.

[2] Xu K, Xu Y, Zhou P, Wang L. Application of RNAMlet to surface defect identification of steels. Opt Lasers Eng 2018;105:110–17.

[3] Neogi N, Mohanta DK, Dutta PK. Review of vision-based steel surface inspection systems. EURASIP J Image Video Process 2014;2014(1):50.

[4] Xie Y, Ye Y, Zhang J, Liu L, Liu L. A physics-based defects model and inspection algorithm for automatic visual inspection. Opt Lasers Eng 2014;52:218–23.

[5] Ghorai S, Mukherjee A, Gangadaran M, Dutta PK. Automatic defect detection on hot-rolled flat steel products. IEEE Trans Instrum Meas 2013;62(3):612–21.

[6] Kuo C-FJ, Lai C-Y, Kao C-H, Chiu C-H. Integrating image processing and classification technology into automated polarizing film defect inspection. Opt Lasers Eng 2018;104:204–19.

[7] Kapsalas P, Maravelaki-Kalaitzaki P, Zervakis M, Delegou E, Moropoulou A. Optical inspection for quantification of decay on stone surfaces. NDT E Int 2007;40(1):2–11.

[8] Tikhe C, Chitode J. Metal surface inspection for defect detection and classification using Gabor filter. Int J Innov Res Sci Eng Technol 2014;3(6):13702–9.

[9] Medina R, Gayubo F, González-Rodrigo LM, Olmedo D, Gómez-García-Bermejo J, Zalama E, et al. Automated visual classification of frequent defects in flat steel coils. Int. J. Adv Manuf Technol 2011;57(9–12):1087–97.

[10] Jeon Y-J, Choi D-c, Lee SJ, Yun JP, Kim SW. Defect detection for corner cracks in steel billets using a wavelet reconstruction method. JOSA A 2014;31(2):227–37.

[11] Park C, Choi S, Won S. Vision-based inspection for periodic defects in steel wire rod production. Opt Eng 2010;49(1):017202.

[12] Jia H, Murphey YL, Shi J, Chang T-S. An intelligent real-time vision system for surface defect detection. In: Proceedings of the seventeenth international conference on pattern recognition, ICPR, 3. IEEE; 2004. p. 239–42.

[13] Choi SH, Yun JP, Seo B, Park YS, Kim SW. Real-time defects detection algorithm for high-speed steel bar in coil. In: Proceedings of world academy of science, engineering and technology, 21; 2007. p. 1307–6884.

[14] Yuan X-c, Wu L-s, Peng Q. An improved Otsu method using the weighted object variance for defect detection. Appl Surf Sci 2015;349:472–84.

[15] Amid E, Aghdam SR, Amindavar H. Enhanced performance for support vector machines as multi-class classifiers in steel surface defect detection. World Acad Sci Eng Technol 2012;6(7):1096–100.

[16] Chu M, Gong R, Gao S, Zhao J. Steel surface defects recognition based on multi-type statistical features and enhanced twin support vector machine. Chemomet Intell Lab Syst 2017;171:140–50.

[17] Soukup D, Huber-Mörk R. Convolutional neural networks for steel surface defect detection from photometric stereo images. In: Proceedings of the international symposium on visual computing. Springer; 2014. p. 668–77.

[18] Mirapeix J, García-Allende P, Cobo A, Conde O, López-Higuera J. Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks. NDT E Int 2007;40(4):315–23.

[19] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the international conference on neural information processing systems; 2012. p. 1097–105.

[20] Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: a large-scale hierarchical image database. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2009. p. 248–55.

[21] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis (IJCV) 2015;115(3):211–52. doi:10.1007/s11263-015-0816-y.

[22] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2014. p. 580–7.

[23] Chen S, Wang H, Xu F, Jin Y-Q. Target classification using the deep convolutional networks for SAR images. IEEE Trans Geosci Remote Sens 2016;54(8):4806–17.

[24] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the international conference on neural information processing systems; 2012. p. 1097–105.

[25] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the international conference on learning representation; 2015.

[26] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

[27] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size; 2016. arXiv preprint arXiv:1602.07360.

[28] Li Y, Li G, Jiang M. An end-to-end steel strip surface defects recognition system based on convolutional neural networks. Steel Res Int 2016;88(2):176–87.

[29] Ren R, Hung T, Tan KC. A generic deep-learning-based approach for automated surface inspection. IEEE Trans Cybern 2018;48(3):929–40.

[30] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, et al. DeCAF: a deep convolutional activation feature for generic visual recognition. In: Proceedings of the international conference on machine learning; 2014. p. 647–55.

[31] Pernkopf F, O'Leary P. Image acquisition techniques for automatic visual inspection of metallic surfaces. NDT E Int 2003;36(8):609–17.

[32] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;smc-3(6):610–21.

[33] Mohamed AA, Yampolskiy RV. Adaptive extended local ternary pattern (AELTP) for recognizing avatar faces. In: Proceedings of the international conference on machine learning and applications; 2013.

[34] Szegedy C, Liu W, Jia Y, Sermanet P, Reed SE, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); 2015. p. 1–9.

[35] Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2017. p. 1302–10.

[36] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39(12):2481–95.

[37] Faraji H, MacLean WJ. CCD noise removal in digital images. IEEE Trans Image Process 2006;15(9):2676–85.

[38] Tasdizen T, Jurrus E, Whitaker RT. Non-uniform illumination correction in transmission electron microscopy. In: Proceedings of the MICCAI workshop on microscopic image analysis with applications in biology; 2008. p. 5–6.

[39] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics; 2010. p. 249–56.

[40] Kingma DP, Ba LJ. Adam: A method for stochastic optimization. 2014. arXiv preprint arXiv:1412.6980.

[41] Antipov G, Berrani S-A, Ruchaud N, Dugelay J-L. Learned vs. hand-crafted features for pedestrian gender recognition. In: Proceedings of the twenty third ACM international conference on multimedia. ACM; 2015. p. 1263–6.