Journal of Intelligent Manufacturing (2021) 32:1833–1843
https://doi.org/10.1007/s10845-020-01670-2

A steel surface defect inspection approach towards smart industrial monitoring

Ruiyang Hao[1] · Bingyu Lu[1] · Ying Cheng[2] · Xiu Li[3] · Biqing Huang[1]

Received: 30 May 2020 / Accepted: 14 September 2020 / Published online: 20 September 2020

Abstract

With the advance in Industry 4.0, smart industrial monitoring has been proposed to timely discover faults and defects in industrial processes. Steel is widely used in manufacturing equipment, and steel surface defect inspection is of great significance to the normal operation of steel equipment in manufacturing workshops. In steel defect inspection systems, industrial inspection robots generate images via scanning steel surface, and processors perform surface defect inspection algorithms on images. We focus on applying advanced object detection techniques to surface defect inspection algorithm for sheet steel. In the proposed steel surface defect inspection model, a deformable convolution enhanced backbone network firstly extracts complex features from multi-shape steel surface defects. Then the feature fusion network with balanced feature pyramid generates high-quality multi-resolution feature maps for the inspection of multi-size defects. Finally, detector network achieves the localization and classification of steel surface defects. The proposed model is evaluated on a typical steel surface defect dataset. Our model achieves 0.805 mAP, 0.144 higher than baseline models, and our model shows high efficiency in inference. Experiments are performed to reveal the effect of employed approaches, and results also show our model achieves a balance between inspection performance and inference efficiency.

Keywords Defect inspection · Object detection · Smart industrial monitoring · Steel surface · Deep learning

Introduction

With the advance in Industrial Internet of Things and artificial intelligence, the fourth industrial revolution, i.e. Industry 4.0, is rapidly developing (Cohen et al. 2019). As a typical application of Industry 4.0, the smart industrial monitoring system adopts ubiquitous sensors and processors to monitor equipment and timely discover faults, where artificial intelligence algorithms are employed to automatically inspect defects (Pimenov et al. 2018). Towards smart industrial monitoring, we propose a steel surface defect inspection model based on advanced object detection approaches, and the main application scenarios of our research focus on defect inspection of steel equipment when it has already been formed and deployed in manufacturing workshops.

For most equipment in manufacturing workshops, steel is used as a common and crucial material. And the good condition of steel is of great significance to the normal operation of manufacturing equipment. However, due to indicators such as external force, equipment fatigue and low quality of steel products, various defects such as crazing, inclusion, patches and scratches may arise on the steel surface. These steel surface defects affect the bearing capacity and residual life of manufacturing equipment (Leinenbach et al. 2012), especially for equipment with sheet steel. Some defects may lead to structure rupture and abnormal stress

✉ Xiu Li
li.xiu@sz.tsinghua.edu.cn

✉ Biqing Huang
hbq@tsinghua.edu.cn

Ruiyang Hao
hry20@mails.tsinghua.edu.cn

Bingyu Lu
lu-by19@mails.tsinghua.edu.cn

Ying Cheng
ycheng@buaa.edu.cn

[1] Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

[2] School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

[3] Division of Information Science, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

state in steel equipment, and some defects may affect the corrosion and wear resistance of steel equipment. Therefore, the timely discovery of these defects can provide early warning to workers. Considering inspected defects and equipment operation condition, workers can better control the equipment to avoid serious production accidents such as structure rupture. And the equipment with inspected defects can be timely maintained before defects become more serious. Specifically, we focus on surface defect inspection for sheet steel.

Surface defect inspection for sheet steel is mainly performed by manual inspection which is unreliable (Ghorai et al. 2013). In order to avoid artificial problems and improve efficiency, it is desirable to automatically inspect steel surface defects. The application and promotion of industrial detection robot (Vithanage et al. 2019) bring convenience to the defect detection of steel in workshop environments. The automatic inspection robot can scan the steel surface as it moves over steel. The scanned images are fed into processors for on-line steel surface inspection. The inspected defects are reported for further equipment control and maintenance as mentioned in the previous paragraph. Therefore, the steel surface inspection algorithms are expected to achieve defect localization and classification, and a balance between inference efficiency and inspection performance is also pursued for high-quality defect inspection.

With the development of computing power and the growing of massive data, deep learning methods have been widely employed in numerous industries (Badmos et al. 2020; Tabernik et al. 2020). Based on deep learning, object detection techniques, which can realize the localization and classification for specific objects, have developed into a topic research field. Object detection techniques have been applied in numerous industries (Hu et al. 2019; Li et al. 2019), but they have not been fully used in steel surface defect inspection. We focus on applying advanced object detection techniques to surface defect inspection for sheet steel. Aiming at six kinds of steel surface defects, a deformable convolution enhanced backbone network firstly extracts complex features of multi-shape steel surface defects. Then feature fusion network with balanced feature pyramid further improves multi-resolution features for the detection of multi-size defects. Finally the detector network achieves the localization and classification of steel surface defects.

The rest of the paper is organized as follows. "Related work" introduces traditional steel defect inspection methods and advanced object detection methods. In "Methodology" section, the proposed models are explained in detail as well as the training strategies. "Experimental results and discussion" is devoted to show ablation experiments and analyze several related problems. Summary of this paper and future work are presented in "Conclusions and future work".

Related work

In manufacturing workshops, the main causes of steel surface defects include low quality of steel products, unexpected external force, equipment fatigue, etc. Some defects arise in the process of steel production (Yun et al. 2014; Nioi et al. 2019). In steel manufacturing industry, the processes such as secondary refining, continuous casting, rolling, and cooling may generate diversified defects (Zhang et al. 2020), and such defects may affect the corrosion and wear resistance of steel equipment (He et al. 2019). Some defects arise because of unexpected external force in workshops (de Vooys and van der Weijde 2011), especially when external force is excessive or incorrectly applied, and such defects may affect the strength of steel equipment. Besides, some defects on steel equipment arise during use because of fatigue (Zhang et al. 2020), which may affect the stress state in equipment structure. Therefore, the timely discovery of these defects is of great significance to the normal operation of manufacturing equipment. The surface defect inspection system for sheet steel utilizes moving industrial inspection robots with cameras to collect images of steel surface, and then the defect inspection algorithm is employed to localize and classify defects. Therefore, the defect inspection algorithm plays a vital role in the system.

General steel surface defect inspection algorithms mainly focus on two aspects: regional proposal and defect classification. Region proposal aims to find the regions that are more likely to contain defects, while defect classification is to extract features from the regions and distinguish their categories. Traditional region proposal methods include Ostu threshold segmentation, watershed algorithm (Zhang et al. 2017), adaptive threshold segmentation (Liu et al. 2017), selective search algorithm (Uijlings et al. 2013), associated region search (Khan 2018), and so on. However, mentioned segmentation algorithms are suitable for dealing with saliency defects, while search-based algorithms are often slow and computationally expensive. Defect classification is the core module of defect inspection algorithm. It usually extracts features of defects, and then classify defects by classical machine learning algorithms. Handcraft features are often used such as Gabor filter feature (Yun et al. 2009), Fourier transform feature (Paulraj et al. 2010), local binary pattern feature (Song and Yan 2013), wavelet transform feature (Ghorai et al. 2013), and manually customized spatial statistical feature (Chu et al. 2017). These feature extraction schemes often have limited ability and need manual experience design. Classical machine learning algorithms applied to steel surface defects classification include support vector machine (SVM) (Ghorai et al. 2013), nearest neighbor classifier (Luo et al. 2019), decision tree algorithm (Sun et al. 2019), artificial neural network (Bustillo et al. 2018), etc. These classification algorithms are usually used to deal with low-dimensional features, and are not suitable for complex high-dimensional features of multi-shape defects.

With the rapid development of deep learning, object detection techniques have attracted extensive attention from both academia and industry in recent years. Generally, deep convolution neural network (CNN) can automatically extract complex high-dimensional features from images, and detector network structure can generate Region of Interests (RoIs) and distinguish categories. Therefore, object detection has a broad application prospect for the steel surface defect detection task. Girshick et al. (2014) first proposed a region-based convolutional neural network (R-CNN) for object detection, where selective search is used to generate RoIs, and SVM as well as linear regression is employed for classification and localization. Then inspired by the spatial pyramid pooling network (He et al. 2014), Girshick (2015) proposed Fast R-CNN, where the entire image is fed into CNN, and a ROI pooling layer is designed to obtain the feature corresponding to each RoI. Ren et al. (2015) proposed Faster R-CNN which can generate RoIs by region proposal network (RPN) instead of selective search. Many subsequent studies focus on improving the performance of object detection. Lin et al. (2017) introduced the feature pyramid network (FPN) into Faster R-CNN to improve the detection accuracy of multi-size object. He et al. (2017) replaced the RoI pooling layer with RoI align layer to avoid the quantization error. Dai et al. (2017) proposed the deformable convolution which expands the receptive field of the convolution operator by adding offset to the convolution sampling point. Pang et al. (2019) proposed Libra R-CNN where balanced feature pyramid is designed to address the imbalance of extracted feature. Although object detection has become an attractive topic in machine vision and artificial intelligence, but it has not been fully applied in the detection of steel surface defects.

There are only a few object detection approaches applied in steel surface defect detection, and the existing models have not reached a balance between inspection performance and inference efficiency. He et al. (2020) opened a steel surface defect detection dataset NEU-DET, and proposed a steel surface defect inspection approach. This model fuses multi-resolution features into a single-resolution feature map to replace the traditional specific feature map. However, single-resolution feature map may not accurately detect multi-size defects, and the efficiency of this method can be improved via more advanced approaches. Lv et al. (2020) proposed an active learning approach for steel surface defect inspection, and a YOLO-v2 based model is performed on the NEU-DET dataset. This model is concise and efficient, but it shows poor inspection performance.

As reviewed in previous paragraphs, the defect inspection algorithm plays a vital role in the steel surface defect inspection system. And there are some shortcomings in the traditional region proposal methods and defect classification methods. While object detection approaches based on deep learning can provide solutions to this issue, but have not been fully applied in steel surface defect detection. And existing models with object detection approaches have not reached a balance between inspection performance and inference efficiency. In this paper, we propose a steel surface defect inspection network with advanced object detection approaches. In our model, a deformable convolution enhanced backbone network is designed to extract the complex features of multi-shape defects. A feature fusion network with balanced feature pyramid is proposed to generate multi-resolution features for the inspection of multi-size defects. And detector network is employed for the localization and classification of defects. The proposed model is evaluated on the NEU-DET dataset, and experimental results are fully analyzed.

Methodology

In this section, the structure and the training of defect inspection network (DIN) are fully illustrated.

Defect inspection network

The framework of the DIN is illustrated in Fig. 1. For a padded image of steel surface, multi-resolution feature maps are extracted by a CNN backbone. Because of the diversity in shapes of steel surface defects, deformable convolution operator is introduced to backbone network to make receptive field of convolution more suitable to defect shapes. Then the feature fusion network with balanced feature pyramid fuses the multi-resolution feature maps to enhance the semantic information of multi-resolution feature maps, which would help the detection of multi-size defects. And RPN is employed to generate RoIs which are more likely to contain steel surface defects. Next, the features corresponding to each RoI, which are extracted from appropriate feature maps, are mapped to fixed size. At last, the Fast R-CNN module is employed for defect classification and localization.

Deformable convolution enhanced backbone network

Feature extraction is the first step of the whole model, and high-quality features extracted from image pixels are of great importance to the subsequent modules. The backbone network is used to extract complex high-dimension features of steel surface defects. The mainstream backbone networks include visual geometry group model (Simonyan and Zisserman 2015), GoogleNet (Szegedy et al. 2015) and residual network (ResNet) (He et al. 2016). Since the ResNet model has reached a good balance between the parameter scale and the feature extraction performance (He et al. 2020, 2016), it is selected to design the backbone convolutional network.

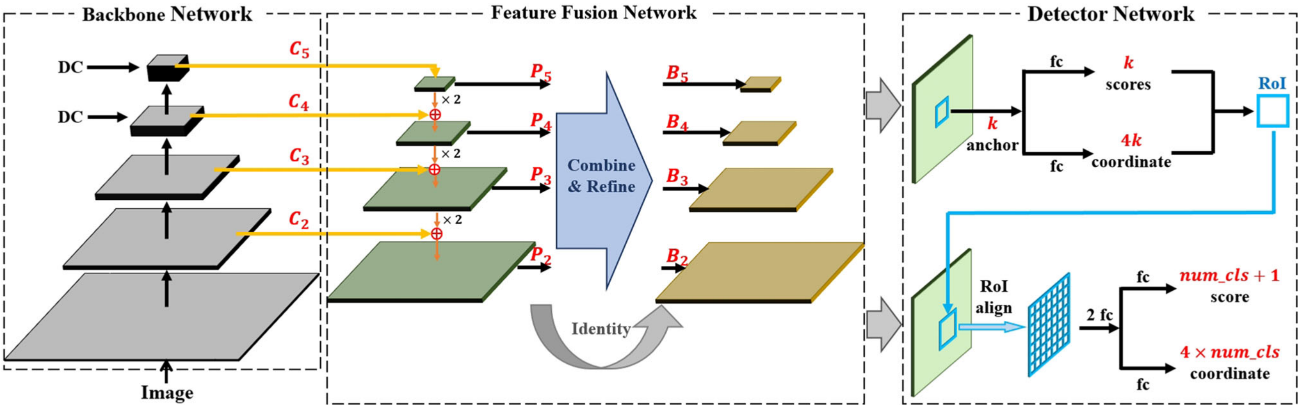

Fig. 1 The architecture of steel surface defect detection network

Table 1 Backbone network architecture

Stage	Output resolution / initial resolution	Layer config
1	1/2	7×7, 64
		3×3, pooling
2	1/4	$\begin{bmatrix} 1\times 1, 64 \\ 3\times 3, 64 \\ 1\times 1, 256 \end{bmatrix} \times 3$
3	1/8	$\begin{bmatrix} 1\times 1, 128 \\ 3\times 3, 128 \\ 1\times 1, 512 \end{bmatrix} \times 4$
4	1/16	$\begin{bmatrix} 1\times 1, 256 \\ 3\times 3, 256, dc \\ 1\times 1, 1024 \end{bmatrix} \times 6$
5	1/32	$\begin{bmatrix} 1\times 1, 512 \\ 3\times 3, 512, dc \\ 1\times 1, 2048 \end{bmatrix} \times 3$

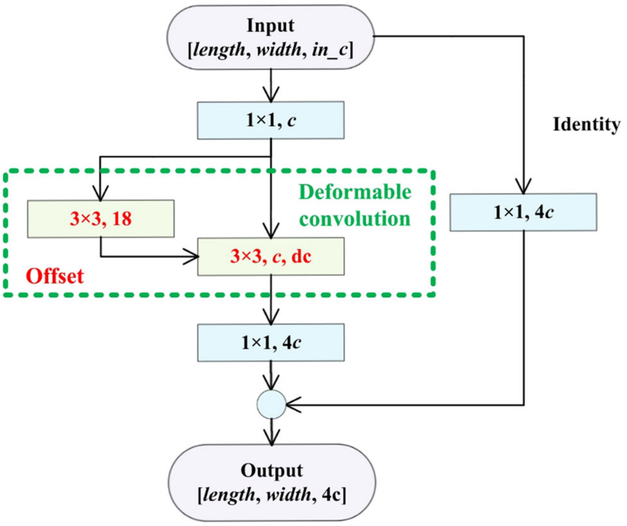

Fig. 2 Deformable convolution block

Besides, since the defect shapes are diversified and irregular. Defects' shapes are different from the receptive field shape of convolution structure which is regularly set as square. Formable convolution (DC) can enhance the transformation modeling of CNNs by augmenting the spatial sampling locations with offsets (Dai et al. 2017), which can better adapt to the defect shapes. Therefore, the DC operator is introduced to the ResNet in this study. The details of the DC enhanced backbone network are described in detail below.

The backbone network architecture based on ResNet-50 with DC enhanced is shown in Table 1. Cropped images are fed into the backbone network. And the last feature maps in stage 2, 3, 4 and 5 regarded as C_2, C_3, C_4, C_5 are output to the feature fusion network. In Table 1, $k \times k, c$ means a $k \times k$ kernel with c channels, while $k \times k, c, dc$ means a $k \times k$ deformable convolution kernel with c channels. The repeating structure in each stage in Table 1 is called deformable convolution block, which is formally expressed in Eq. 1 and illustrated in Fig. 2. In each deformable convolution block, 1 × 1 convolution layers control the feature dimension in the start and the end of blocks to reduce the size of convolution parameters. 3 × 3 deformable convolution layer is the main convolution layer in each block, which is marked in the green box in Fig. 2. More details about deformable convolution are supplemented below.

$$\begin{bmatrix} 1\times 1, c \\ 3\times 3, c, dc \\ 1\times 1, 4c \end{bmatrix} \tag{1}$$

In the traditional convolution structure, the offset from sampling location to center of the convolution kernel is fixed. The set of fixed offsets is denoted as Ω, which defines the size and dilation of square receptive field. For example, Ω for a 3 × 3 kernel with dilation 1 is supplemented in Eq. 2.

$$\begin{aligned} \Omega = &\{(-1,-1),(-1,0),(-1,1),(0,-1), \\ &(0,0),(0,1),(1,-1),(1,0),(1,1)\} \end{aligned} \tag{2}$$

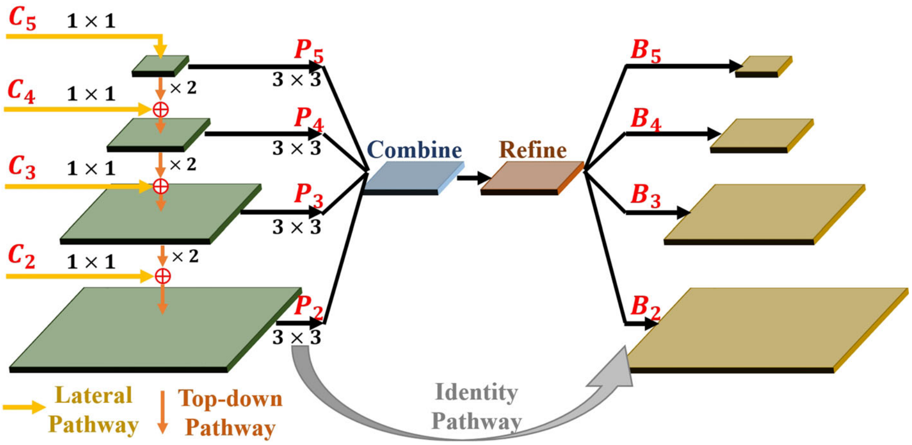

Fig. 3 Feature fusion network

However, the deformable convolution module augments extra offsets to the basic sampling location, which can make the receptive field more adaptive to defect shapes. For each location p_0 in output feature map y, deformable convolution module is expressed in Eq. 3.

$$y(p_0) = \sum_{p_n \in \Omega} w(p_n)x(p_0 + p_n + \Delta p_n) \tag{3}$$

In Eq. 3, x represents the input feature map and w represents the sampling weight parameter. p_n indicates the fixed offset from the basic sampling location to the center of the convolution kernel. Δp_n indicates the offset from the true sampling location to the basic sampling location, which is computed by an extra traditional convolution layer (shown in Fig. 2). Since the computed offset Δp_n is typically fractional, the sapling procedure $x(p_0 + p_n + \Delta p_n)$ is completed by the bilinear interpolation method.

Feature fusion network The size of steel surface defects is diversified, so the single-resolution feature map output by backbone network may show poor performance in detecting multi-size defects. As indicated in Lin et al. (2017), low-resolution feature maps are suitable to detect large object, while high-resolution feature map is suitable to detect small objects, so the application of multi-resolution feature maps can help the inspection of multi-size defects. Based on feature pyramid network (Lin et al. 2017) and balanced feature pyramid structure (Pang et al. 2019), feature fusion network is designed to improve the semantic information of multi-resolution feature maps. As shown in Fig. 3, the multi-resolution feature maps output by backbone network are fed into the lateral pathway for dimension reduction. Then these maps are merged in a top-down pathway, where low-resolution feature maps are up-sampled and then merged with nearby high-resolution feature maps. Afterward, fused feature maps integrates and refines to obtain balanced feature pyramid. The implementation details are discussed below.

As shown in Fig. 3, to fuse multi-resolution feature maps C_2, C_3, C_4, C_5 which have different dimensions, the lateral pathway reduces these dimensions to 256 via a 1×1 convolution layer. In the top-down pathway, the up-sampling procedure adopts nearest neighbor upsampling for simplicity, and corresponding feature maps are merges via element-wise addition. Besides, fused maps after the top-down pathway undergo a 3×3 convolution layer to reduce the aliasing effect of upsampling, and the output fused maps are called P_2, P_3, P_4, P_5. To integrate multi-resolution features and preserve their semantic information, multi-resolution features P_2, P_3, P_4, P_5 are firstly resized to an intermediate size (the same size as P_4) by interpolation and max-pooling respectively. Then resized feature maps are averaged to obtain a balanced feature map. Next, the balanced feature map is refined to be more discriminative via a non-local module. In the end, the refined feature map is rescaled to the initial resolutions and further appended to P_2, P_3, P_4, P_5 via element-wise addition. The obtained multi-resolution feature maps B_2, B_3, B_4, B_5 are the output of feature fusion network.

Detector network Based on high-quality feature maps from feature fusion network, detector network would ultimately achieve the localization and classification of steel surface defects. As shown in Fig. 1, the detector network firstly employs RPN to select RoIs which are more likely to contain steel surface defects. Next, the features corresponding to each RoI, which are extracted from the appropriate feature map, are mapped to a fixed-size feature by a RoI align module. The subsequent part is a Fast R-CNN module, where the mapped feature is fed into fully-connected layers for defect classification and more precise localization.

RPN takes multi-resolution features B_2, B_3, B_4, B_5 as input and outputs RoIs. For each feature map, RPN uses a 3×3 spatial sliding window on the feature map to obtain local feature. Then the feature is fed into two sibling fully-connected layers including box-regression layer and box-classification layer. At each sliding window location,

Table 2 Feature level assignment for RoIs

Feature maps	Pre-set anchor area	RoI area range
B_2	32^2	$0 < S < 46^2$
B_3	64^2	$46^2 <= S < 92^2$
B_4	128^2	$92^2 <= S < 184^2$
B_5	256^2	$S >= 184^2$

multiple region proposals called anchors are simultaneously predicted. The Anchor with a certain size and aspect ratio is at the center of each sliding window (Ren et al. 2015). The sliding window and fully-connected layers are respectively implemented with 3 × 3 and 1 × 1 convolution layers. It's worth noting that the classification layer in RPN only performs defect/non-defect classification, and it cannot predict the defect category. RoIs are selected from anchors according to classification scores and the Intersection over Union (IoU) with ground-truth boxes as in Faster R-CNN (Ren et al. 2015).

To fully use RoIs with multi-resolution feature maps, RoIs in different sizes are assigned to appropriate pyramid levels, which is summarized in Table 2. Then RoI align module (He et al. 2017) is employed to map the features corresponding to each RoI into a 7 × 7 feature. Afterward, this feature is flattened and fed into two fully-connected layers with ReLU following. At last, two sibling fully-connected layers including regression layer and classification layer are employed for further localization and final category classification. In the outputs of detector network, the defect category is expressed by confidence score, and the defect localization is expressed by bounding-box coordinates. The procedures after RoI align are also known as the Fast R-CNN module.

Training

Multi-task loss function RPN and Fast R-CNN modules in DIN respectively produce RPN loss and Fast R-CNN loss in training. And both modules produce classification loss and regression loss. For box-classification task in the RPN module, positive labels and negative labels are firstly assigned to each anchor, then the log loss between the ground-truth label and the predicted probability is calculated as RPN classification loss. For box-classification task in the Fast R-CNN module, ground-truth categories are firstly assigned to each RoI, then the log loss between the ground-truth category and the predicted confidence score is calculated as Fast R-CNN classification loss. For box-regression task in both modules, 4 coordinates (box's center coordinates and its width and height) of box-regression results and ground-truth boxes are firstly parameterized. Then the smooth L1 loss between regression coordinates and ground-truth box coordinates is calculated as regression loss (Girshick 2015).

Approximate joint training Since the RPN module and the Fast R-CNN module share multi-resolution feature maps, if these two modules are trained independently, parameters in shared networks (backbone network and feature fusion network) will be repeatedly modified. Approximate joint training scheme is employed for this issue. In each training iteration, RoIs generated by RPN are treated as fixed pre-trained RoIs when training the Fast R-CNN module in the forward propagation. In the backward propagation, RPN loss and Fast R-CNN loss are backward propagated as usual before shared networks, and are combined to backward propagated in shared networks. Although this training scheme is approximate since it ignores the derivative with respect to RoIs in Fast R-CNN, it is easy to implement and can produce good results (Ren et al. 2015).

Experimental results and discussion

In this section, employed dataset NEU-DET is firstly introduced. Implementation details and evaluation metric are supplemented. Experiments are performed to evaluate the proposed DIN model, and the experimental results are analyzed and discussed in detail.

Dataset

Steel surface defect dataset is of great importance for the data-driven inspection model. Using high-quality and large-scale defect datasets can improve the generalization performance of the model. NEU-DET dataset is a defect detection dataset opened by He et al. (2020). Six types of defects including crazing, inclusions, patches, pitted surface, rolled-in scale and scratches are gathered. For each type of defects, 300 images with defects were collected. These defects are representative on the surface of manufacturing equipment. Among them, some defects such as crazing may lead to structure rupture, some defects such as inclusions may affect the stress state in structure of steel equipment and even develop into more serious defects, and some defects such as patches and pitted surface may affect the corrosion and wear resistance of steel equipment. Therefore, the timely defect inspection is pursed as discussed in Introduction. It's worth noting that there may be several defects in one image, which meets the application needs in industry. Several typical images with defects are shown in Fig. 4.

In order to inspect defects with various ratios and diversified sizes, it is of necessity to analyze the ratio distribution and size distribution of all the defects. As shown in Fig. 5, the ratios of the long side to the short side of most defects are

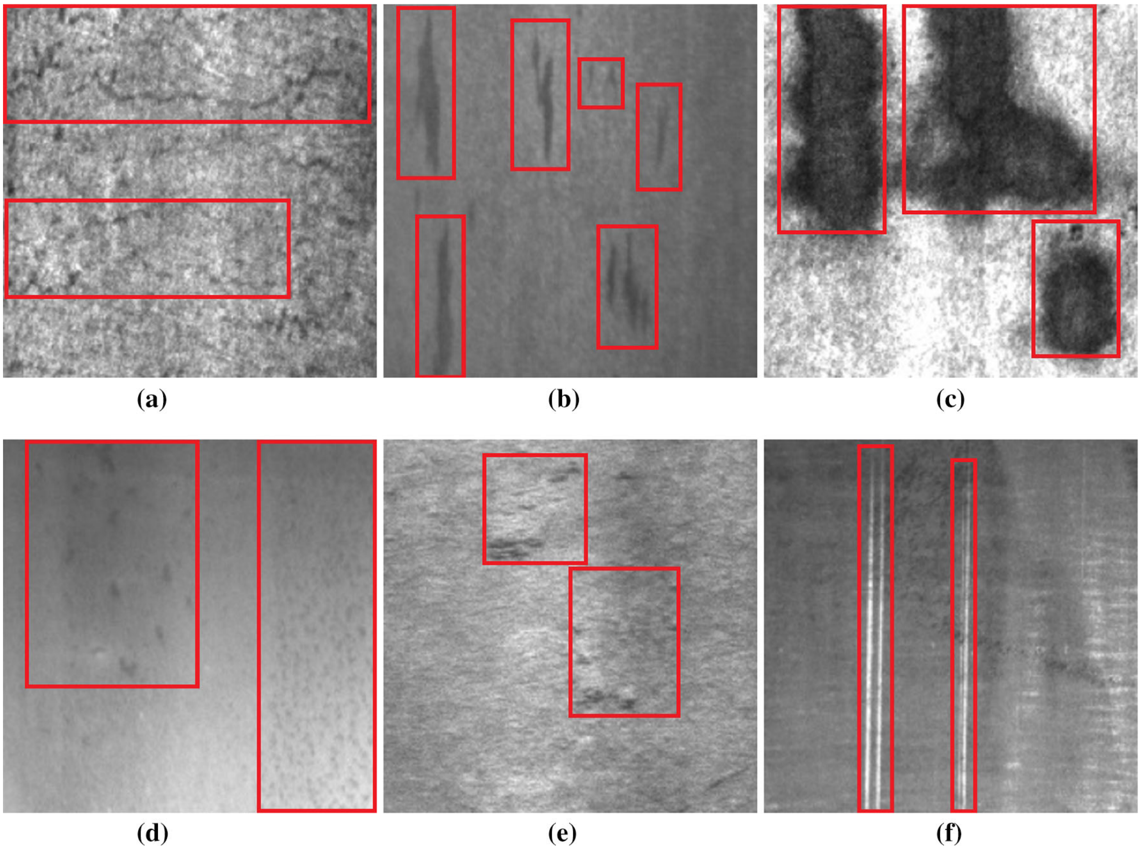

Fig. 4 Image examples of steel surface defects. Red boxes are the groundtruth boxes with a defect class label and position coordinates. Each defect image belongs to the following class: **a** Crazing. **b** Inclusions. **c** Patches. **d** Pitted surface. **e** Rolled-in scale. **f** Scratches

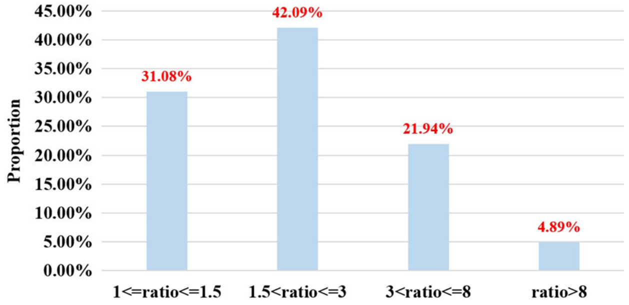

Fig. 5 The distribution of the ratio of the long side to the short side of all the defects

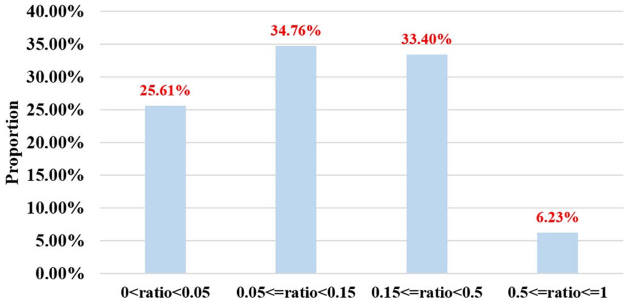

Fig. 6 The distribution of the ratio of the defect area to the image area

around 1, 3 and 5, and the proportion of extreme ratio is very low. As shown in Fig. 6, the ratios of the defect area to the image area show that the steel surface defects have diversified sizes and most defects are small-size defects. These dual diversities in defect shapes and defect sizes also correspond to the necessity of deformable convolution and feature fusion network in DIN as explained in Section "Defect inspection network". And these two distribution features of defects are used to set anchor parameters in DIN.

Implementation details

The backbone network is initialized with an ImageNet-pretrained model, and other networks are initialized randomly. End-to-end training is performed for 15 epochs with an initial learning rate of 0.01, and decrease it by 0.1 after 9 epochs. Stochastic gradient descent is employed as an optimizer with a momentum of 0.9 and a weight decay of 0.0001. We take four images per mini-batch iteration, and the mini-batch size is 256 RPN training and 512 for Fast R-CNN training. Considering the distribution of defect size and shape analyzed in

Table 3 Inspection results on NEU-DET

Models	mAP	Crazing	Inclusions	Patches	Pitted surface	Rolled-in scale	Scratches
Faster R-CNN	0.661	0.314	0.724	0.777	0.715	0.589	0.844
Faster R-CNN with FPN	0.781	0.536	0.840	0.956	0.884	0.625	0.845
Faster R-CNN with BFP	0.791	0.578	0.849	0.935	0.890	0.640	0.852
DIN (our model)	0.805	0.614	0.856	0.930	0.903	0.646	0.883

Section "Defect inspection network", anchors have area of 32^2, 64^2, 128^2, 256^2 pixels on B_2, B_3, B_4, B_5 respectively, and anchors have multiple aspect ratios 1:5, 1:2, 1:1, 2:1, 5:1 at each level. The input images are padded to 224 × 224 pixels for DIN. The NEU-DET dataset is split into training set and test set with 1440 and 360 images respectively. The experiments below are implemented on an NVIDIA GeForce RTX 2080 Ti (11 GB) GPU and an Intel 3.20 GHz i7-6700 CPU.

Metrics

The metric of inspection performance is firstly introduced. The criterion of defect A is defined below: (1) the IoU between predicted bounding box the ground-truth defect A is more than 0.5; (2) the confidence score for category A is higher than threshold th_{conf}. Parameter th_{conf} is the threshold of confidence score to determine the category in bounding box. The precision and recall of defect A are defined in Eqs. 4 and 5, where TP_A, FP_A and FN_A represent the number of true positive, false positive, and false negative of defect A, respectively.

$$Precision_A = \frac{TP_A}{TP_A + FP_A} \tag{4}$$

$$Recall_A = \frac{TP_A}{TP_A + FN_A} \tag{5}$$

If we increase the parameter th_{conf} in defect criterion, which means that the defect criterion is more strict, the precision will increase while the recall will decrease. Therefore, influenced by parameter th_{conf}, precision is negatively correlated with recall. Referring to the evaluation metric for COCO object detection task (Everingham et al. 2010), average precision (AP) is defined as the area under the precision-recall curve (PR curve), which can comprehensively reflect inspection effect including both precision and recall. AP of 6 types of defects are averaged as mean average precision (mAP). Besides, the metric of inspection efficiency is defined by frame per second (FPS), which means the number of images that the model can infer in a second.

Main results

The results of DIN are shown in Table 3. The baseline model (Faster R-CNN) and models with specific components are also constructed and trained for comparison and analysis. Our model achieves 0.805 mAP, 0.144 higher than the baseline model. Besides, our model shows high efficiency in inference. On a single GPU, our model can infer 43.5 images per second.

A comparison among several model results is analyzed below. Since the multi-resolution fused feature maps can better represent the features of multi-size steel surface defects, feature pyramid network (FPN) effectively improves mAP from 0.661 to 0.781. And the improvements of AP in patches, crazing and inclusions are significant. Based on FPN, balanced feature pyramid (BFP) can further enhance the multi-resolution fused features via refining balanced semantic feature. Compared with Faster R-CNN with FPN, the implementation of BFP further improves mAP from 0.781 to 0.791. The deformable convolution (DC) can attach extra offsets to the traditional spatial sampling locations, which can make convolution more adaptive to defect shapes. The addition of deformable convolution into the backbone network improves mAP from 0.791 to 0.805. By comparing results of four models, AP of crazing, inclusions, pitted surface and rolled-in scale are continuously improved by adding mechanisms including FPN, BFP and DC. These comparative experiments fully show the superior inspection performance of our model. Comparing with other researches, for the same defects, APs in our model are significantly higher than APs reported in Lv et al. (2020). And our model shows higher efficiency in inference. Our model can infer 43.5 images per second on a single GPU, much higher than 20 reported in He et al. (2020).

PR curves can also show the inspection performance of our model. For example, the PR curves of inclusions and patches output by our model are shown in Fig. 7, although the neg-

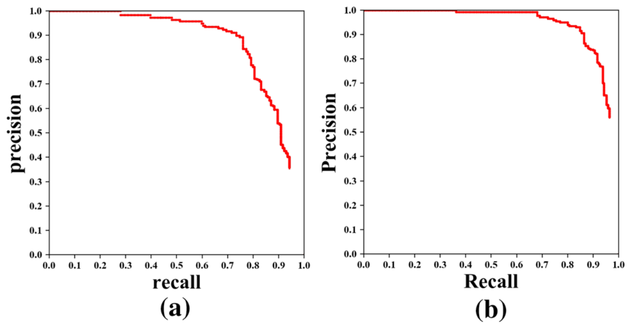

Fig. 7 PR curves. AP is defined as the area under the PR curve. **a** Inclusions. **b** Patches

Table 4 Results of different sizes of balanced feature map

Size of balanced feature map	mAP	Crazing	Inclusions	Patches	Pitted surface	Rolled-in scale	Scratches
56 × 56 (the same as P_2)	0.791	0.556	0.851	0.928	0.898	0.661	0.853
28 × 28 (the same as P_3)	0.795	0.594	0.854	0.927	0.897	0.640	0.862
14 × 14 (the same as P_4)	0.805	0.614	0.856	0.930	0.903	0.646	0.883
7 × 7 (the same as P_5)	0.787	0.569	0.855	0.941	0.863	0.646	0.847

Table 5 Results of different DC-added ResNet stages

ResNet stages with DC added				mAP	Crazing	Inclusions	Patches	Pitted surface	Rolled-in scale	Scratches
Stage 1	Stage 2	Stage 3	Stage 4							
F	T	T	T	0.804	0.610	0.857	0.926	0.904	0.643	0.886
F	F	T	T	0.805	0.614	0.856	0.930	0.903	0.646	0.883
F	F	F	T	0.797	0.589	0.847	0.932	0.895	0.644	0.875
F	T	T	T	0.791	0.578	0.849	0.935	0.890	0.640	0.852

ative correlation between precision and recall is distinctive, there are some points that can accomplish both high precision and satisfying recall. More model details are analyzed in the discussion below.

Discussion

What is the optimal size of balanced feature map? BFP appends refined balanced feature map to traditional FPN to enhance the semantic information of multi-resolution feature maps. In Section "Defect inspection network", we resized multi-resolution features to the size of P4, and merge them into a balanced feature map. So, what is the optimal size of balanced feature map? The size of balanced feature map is adjusted, and experimental results are shown in Table 4. These contrastive results indicate that the intermediate-scale balanced feature map (the same size as P4) can better improve the semantic information via merging it to multi-resolution feature maps.

How many DC layers are added to backbone network at least? DC is added to ResNet-based backbone network to make convolution more adaptive to multiple defect shapes. However, DC introduces more convolution layers to calculate the offsets of spatial sampling locations. Therefore, DC layers are expected to be as few as possible without affecting the inspection performance. The traditional DC for general object detection task is added to the last stages of backbone network, and we follow this rule in our research. The stages to which DC layers are added are adjusted, and experimental results are shown in Table 5. The contrastive results show that the last 2 stages with DC layers added are sufficient for good effect.

Achieved a balance between inspection performance and inference efficiency? Steel surface defect inspection models should achieve not only good inspection performance, but also high inference efficiency. Both factors are of great importance to the industrial implementation of our model. RPN simultaneously predicts many region proposals, however, only a few region proposals with higher confidence score are selected as RoIs which are further input to Fast R-CNN. Therefore, the RoI number is a critical hyper-parameter which influences inspection performance and inference efficiency. If more RoIs are output to Fast R-CNN, Fast R-CNN may discover the defects which are not well detected in RPN, but inference efficiency is weakened since Fast R-CNN is applied to more RoIs. As shown in Table 6, with the decrease of RoI number, our model shows robust inspection performance and higher inference efficiency. mAP and FPS are respectively employed to evaluate inspection performance and inference efficiency. Even RoI number decreases from 500 to 50, mAP still reaches 94.8% of best performance, while the inference efficiency increased by 17%. And the increase of RoI number from 500 to 1000 does not effectively improve mAP, but suppresses inference efficiency. Therefore, RoI number is set as 500 for the balance between inspection performance and inference efficiency.

Why are some defects not accurately detected? Although our model generally achieves promising results, there exist some defects not accurately inspected. We attempt to explore the reasons for unsatisfying inspection results. (1) The difference between confusing defects and background is not distinct. For example, our model cannot recognize the upper crazing defect in Fig. 8a, and it recognizes the bottom box as a rolled-in scale defect in Fig. 8b. Such defects are so difficult that even humans cannot accurately distinguish them

Table 6 Results of different RoI numbers

RoI number	FPS	mAP	Crazing	Inclusions	Patches	Pitted surface	Rolled-in scale	Scratches
50	50.9	0.763	0.555	0.820	0.900	0.887	0.561	0.833
100	50.0	0.776	0.585	0.833	0.903	0.878	0.594	0.840
300	47.8	0.797	0.609	0.854	0.924	0.893	0.640	0.866
500	43.5	0.805	0.614	0.856	0.930	0.903	0.646	0.883
1000	39.9	0.806	0.615	0.859	0.929	0.903	0.648	0.881

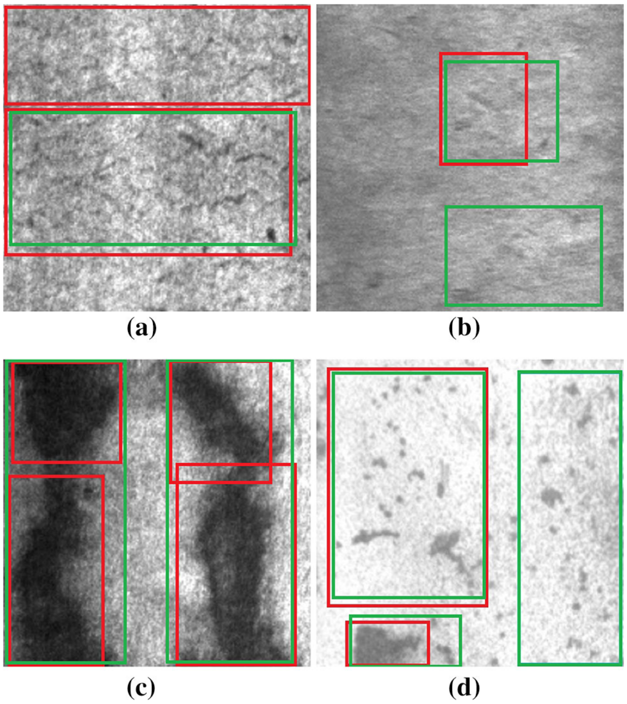

Fig. 8 Failure analysis. Red boxes are ground-truth boxed while green boxes are detected defects. **a** Confusing defect (crazing). **b** Confusing defect (rolled-in scale). **c** Unclear boundary. **d** Incomplete label

from background. (2) The boundaries between some adjacent defects are not distinct. As shown in Fig. 8c, our model might not be able to precisely find unclear boundaries, and may inspect such adjacent defects as a single defect, so the smaller one of adjacent defects cannot be correctly defined according to the general defect criterion. (3) Some distinctive defects can be precisely inspected but they may be unlabeled. As shown in Fig. 8d, the right side of the image clearly contains pitted surface which our model detects, while it is not labeled in the original dataset.

Conclusion and future work

Steel surface defect detection is of great significance to the normal operation of steel equipment in manufacturing workshops. Towards smart industrial monitoring, we apply advanced object detection techniques to steel surface defect inspection. The proposed model achieves 0.805 mAP, 0.144 higher than baseline model, and our model shows high efficiency in inference (43.5 FPS on a single GPU). Comparative experiments are performed to show the effect of employed approaches. The main contributions of our work are summarized as follows: (1) A steel surface defect inspection network (DIN) based on advanced object detection approaches is proposed. And the proposed model achieves a balance between inference efficiency and inspection performance. (2) The diversity in defect shapes brings challenges to the extraction of defect features. Therefore, a deformable convolution enhanced backbone network is designed to improve the quality of extracted features, where deformable convolution can make convolution more adaptive to defect shapes. (3) The diversity in defect sizes also poses challenges to the inspection of multi-size defects. Therefore, a feature fusion network with balanced feature pyramid is proposed to generate high-quality multi-resolution feature maps to help inspect multi-size defects.

Possible future work is also summarized below. (1) Data augmentation techniques are expected to be employed to make up the deficiency of high-quality defect images. (2) How to control and maintain equipment considering both inspected defects and equipment condition is an interesting issue to be further discussed. (3) We expect to evaluate our model in the actual industrial environment.

Acknowledgements This work was supported by a grant from the Institute for Guo Qiang, Tsinghua University (No. 2019GQG0002), and supported by National Natural Science Foundation of China (No. 41876098).

References

Badmos, O., Kopp, A., Bernthaler, T., & Schneider, G. (2020). Image-based defect detection in lithium-ion battery electrode using convolutional neural networks. *Journal of Intelligent Manufacturing*, *31*, 885–897.

Bustillo, A., Pimenov, D. Y., Matuszewski, M., & Mikolajczyk, T. (2018). Using artificial intelligence models for the prediction of surface wear based on surface isotropy levels. *Robotics and Computer-integrated Manufacturing*, *53*, 215–227.

Chu, M., Gong, R., Gao, S., & Zhao, J. (2017). Steel surface defects recognition based on multi-type statistical features and enhanced twin support vector machine. *Chemometrics and Intelligent Laboratory Systems*, *171*, 140–150.

Cohen, Y., Faccio, M., Pilati, F., & Yao, X. (2019). Design and management of digital manufacturing and assembly systems in the industry 4.0 era. *The International Journal of Advanced Manufacturing Technology*, *105*(9), 3565–3577.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y. (2017). Deformable convolutional networks. In: *International Conference on Computer Vision*, pp. 764–773.

de Vooys, A., & van der Weijde, H. (2011). Investigating cracks and crazes on coated steel with simultaneous svet and eis. *Progress in Organic Coatings*, *71*(3), 250–255.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*(2), 303–338.

Ghorai, S., Mukherjee, A., Gangadaran, M., & Dutta, P. K. (2013). Automatic defect detection on hot-rolled flat steel products. *IEEE Transactions on Instrumentation and Measurement*, *62*(3), 612–621.

Girshick, R. (2015). Fast r-cnn. In: International Conference on Computer Vision, pp. 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Conference on Computer Vision and Pattern Recognition*, pp. 580–587.

He, D., Xu, K., & Zhou, P. (2019). Defect detection of hot rolled steels with a new object detection framework called classification priority network. *Computers & Industrial Engineering*, *128*, 290–297.

He, K., Gkioxari, G., Dollar, P., Girshick, R. (2017). Mask r-cnn. In: *International Conference on Computer Vision*, pp. 2980–2988.

He, K., Zhang, X., Ren, S., Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *European Conference on Computer Vision*, pp. 346–361.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In: *Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

He, Y., Song, K., Meng, Q., & Yan, Y. (2020). An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Transactions on Instrumentation and Measurement*, *69*(4), 1493–1504.

Hu, X., Xu, X., Xiao, Y., Chen, H., He, S., Qin, J., et al. (2019). Sinet: a scale-insensitive convolutional neural network for fast vehicle detection. *IEEE Transactions on Intelligent Transportation Systems*, *20*(3), 1010–1019.

Khan, S. (2018). Ant colony optimization (aco) based data hiding in image complex region. *International Journal of Electrical and Computer Engineering*, *8*(1), 379–389.

Leinenbach, C., Koster, M., & Schindler, H. (2012). Fatigue assessment of defect-free and defect-containing brazed steel joints. *Journal of Materials Engineering and Performance*, *21*(5), 739–747.

Li, W., Li, H., Wu, Q., Chen, X., & Ngan, K. N. (2019). Simultaneously detecting and counting dense vehicles from drone images. *IEEE Transactions on Industrial Electronics*, *66*(12), 9651–9662.

Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017). Feature pyramid networks for object detection. In: *Conference on Computer Vision and Pattern Recognition*, pp. 936–944.

Liu, K., Wang, H., Chen, H., Qu, E., Tian, Y., & Sun, H. (2017). Steel surface defect detection using a new haar-weibull-variance model in unsupervised manner. *IEEE Transactions on Instrumentation and Measurement*, *66*(10), 2585–2596.

Luo, Q., Sun, Y., Li, P., Simpson, O., Tian, L., & He, Y. (2019). Generalized completed local binary patterns for time-efficient steel surface defect classification. *IEEE Transactions on Instrumentation and Measurement*, *68*(3), 667–679.

Lv, X., Duan, F., Jiang, J., Fu, X., & Gan, L. (2020). Deep active learning for surface defect detection. *Sensors*, *20*(6), 1650.

Nioi, M., Pinna, C., Celotto, S., Swart, E., Farrugia, D., Husain, Z., et al. (2019). Finite element modelling of surface defect evolution during hot rolling of silicon steel. *Journal of Materials Processing Technology*, *268*, 181–191.

Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., Lin, D. (2019). Libra r-cnn: towards balanced learning for object detection. In: *Conference on Computer Vision and Pattern Recognition*, pp. 821–830.

Paulraj, M.P., Shukry, A.M.M., Yaacob, S., Adom, A.H., Krishnan, R.P. (2010). Structural steel plate damage detection using dft spectral energy and artificial neural network. In: *International Colloquium on Signal Processing and Its Applications*, pp. 1–6.

Pimenov, D. Y., Bustillo, A., & Mikolajczyk, T. (2018). Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth. *Journal of Intelligent Manufacturing*, *29*(5), 1045–1061.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Conference and Workshop on Neural Information Processing Systems*, *2015*, 91–99.

Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations*.

Song, K., & Yan, Y. (2013). A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, *285*, 858–864.

Sun, J., Li, C., Wu, X., Palade, V., & Fang, W. (2019). An effective method of weld defect detection and classification based on machine vision. *IEEE Transactions on Industrial Informatics*, *15*(12), 6322–6333.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. In: *Conference on Computer Vision and Pattern Recognition*, pp. 1–9.

Tabernik, D., Sela, S., Skvarc, J., & Skocaj, D. (2020). Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, *31*, 759–776.

Uijlings, J., Sande, K. E., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *International Journal of Computer Vision*, *104*(2), 154–171.

Vithanage, R. K. W., Harrison, C. S., & De Silva, A. K. M. (2019). Autonomous rolling-stock coupler inspection using industrial robots. *Robotics and Computer-integrated Manufacturing*, *59*, 82–91.

Yun, J. P., Choi, D., Jeon, Y., Park, C., & Kim, S. W. (2014). Defect inspection system for steel wire rods produced by hot rolling process. *The International Journal of Advanced Manufacturing Technology*, *70*(9), 1625–1634.

Yun, J. P., Choi, S., Kim, J., & Kim, S. W. (2009). Automatic detection of cracks in raw steel block using gabor filter optimized by univariate dynamic encoding algorithm for searches (udeas). *Ndt & E International*, *42*(5), 389–397.

Zhang, X., Kano, M., Tani, M., Mori, J., Ise, J., & Harada, K. (2020). Prediction and causal analysis of defects in steel products: Handling nonnegative and highly overdispersed count data. *Control Engineering Practice*, *95*, 104,258.

Zhang, J., Li, H., Yang, B., Wu, B., & Zhu, S. (2020). Fatigue properties and fatigue strength evaluation of railway axle steel: Effect of micro-shot peening and artificial defect. *International Journal of Fatigue*, *132*, 105,379.

Zhang, C., Xie, Y., Liu, D., & Wang, L. (2017). Fast threshold image segmentation based on 2d fuzzy fisher and random local optimized qpso. *IEEE Transactions on Image Processing*, *26*(3), 1355–1362.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.