A Dynamic Weights-Based Wavelet Attention Neural Network for Defect Detection

Jinhai Liu, *Senior Member, IEEE*, He Zhao, Zhaolin Chen, *Member, IEEE*, Qiannan Wang, Xiangkai Shen, *Graduate Student Member, IEEE*, and Huaguang Zhang, *Fellow, IEEE*

***Abstract*— Automatic defect detection plays an important role in industrial production. Deep learning-based defect detection methods have achieved promising results. However, there are still two challenges in the current defect detection methods: 1) high-precision detection of weak defects is limited and 2) it is difficult for current defect detection methods to achieve satisfactory results dealing with strong background noise. This article proposes a dynamic weights-based wavelet attention neural network (DWWA-Net) to address these issues, which can enhance the feature representation of defects and simultaneously denoise the image, thereby improving the detection accuracy of weak defects and defects under strong background noise. First, wavelet neural networks and dynamic wavelet convolution networks (DWCNets) are presented, which can effectively filter background noise and improve model convergence. Second, a multiview attention module is designed, which can direct the network attention toward potential targets, thereby guaranteeing the accuracy for detecting weak defects. Finally, a feature feedback module is proposed, which can enhance the feature information of defects to further improve the weak defect detection accuracy. The DWWA-Net can be used for defect detection in multiple industrial fields. Experiment results illustrate that the proposed method outperforms the state-of-the-art methods (mean precision: GC10-DET: 6.0%; NEU: 4.3%). The code is made in https://github.com/781458112/DWWA.**

***Index Terms*— Defect detection, dynamic weights, feature feedback module, multiview attention module, wavelet convolution networks.**

NOMENCLATURE

Parameters

η	Scaling parameter of layer normalization.
λ	Dynamic weight.
μ	Shift parameter of layer normalization.
σ	Activation function.
φ	Scaling factor.
ξ	Feature fusion coefficient.
g	Fully connected network.
k	Translation coefficient of wavelet neural network.
L_c	Classification loss function.
L_r	Regression loss function.
P	Global pooling.
q	Scaling of wavelet neural network.
w	Coefficient of fully connected network.
$\psi_{q,k}$	Base wavelet family.
$a_{q,k}$	Approximation coefficients.
$d_{q,k}$	Detail coefficients.

I. INTRODUCTION

AFFECTED by the lack of standardization production process and harsh industrial environment, products exhibit various types of defects, which reduce their performance and safety [1], [2], [3]. Therefore, detecting those defects reliably plays an important role in industrial production [4], [5], [6]. However, defect detection is facing great challenges due to: 1) inconspicuous features and small size of weak defects make it difficult for reliable detection, as shown in Fig. 1(a, b). 2) The harsh collection environment introduces strong background noise, which reduces the accuracy of defect detection, as shown in Fig. 1(c). If these defects are not detected and repaired in time, they are likely to reduce the service life of the product and cause accidents. For example, weak defects in pipelines may accelerate corrosion in a short period causing pipeline leakage, resulting in environmental pollution and explosion accidents.

Recently, lots of works have emerged in improving the accuracy of defect detection under weak defects and strong background noise. These studies can be categorized into two categories: 1) model optimization methods and 2) data preprocessing methods.

For model optimization methods, researchers have improved defect detection performance by improving the model structure. In [7], a robust kernel function with Bayesian optimization is used to improve the classification accuracy of weak defects. In [8], a combination of the terahertz technique and deconvolution method is designed to effectively detect small defects. Nevertheless, these methods rely heavily on human expertise and manual feature extraction, limiting their

Manuscript received 25 October 2022; revised 10 May 2023 and 14 June 2023; accepted 2 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U21A20481 and Grant 61973071, in part by the Liaoning Revitalization Talents Program under Grant XLYC2002046, and in part by the Fundamental Research Funds for the Central Universities of China under Grant N2104020. *(Corresponding author: Jinhai Liu.)*

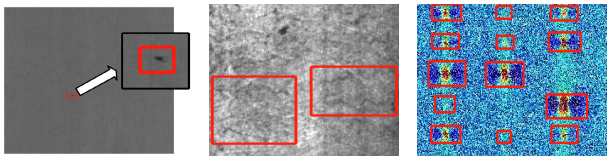
Jinhai Liu is with the State Key Laboratory of Synthetical Automation for Process Industries and the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: liujinhai@mail.neu.edu.cn).

He Zhao, Qiannan Wang, Xiangkai Shen, and Huaguang Zhang are with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: z2110304@163.com; 2210363@stu.neu.edu.cn; 2010302@stu.neu.edu.cn; hgzhang@ieee.org).

Zhaolin Chen is with the Monash Biomedical Imaging, Monash University, Clayton, VIC 3800, Australia (e-mail: Zhaolin.Chen@monash.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3292512.

Digital Object Identifier 10.1109/TNNLS.2023.3292512



(a) (b) (c)

Fig. 1. Defects of different industrial fields. (a) and (b) Weak defects. (c) Defects under strong background noise.

accuracy. To address the above issues, some researchers have turned to deep learning to detect weak defects. In [9], an adaptive image segmentation network is designed, which can capture more contextual information from the defect feature map and improve the detection accuracy. In [10], a novel enhanced multiscale feature fusion method is proposed, which can employ atrous spatial pyramid pooling-balanced-feature pyramid network to make full use of contextual information and thus achieves outstanding results in small defect detection. In [11], a coarse-to-fine framework is proposed, which can effectively save computation and improve the detection accuracy of small defects. In [12], a feature enhancement and loop-shaped fusion convolutional neural network is designed, which enhances shallow features and fuses features with a loop-shaped feature pyramid structure. Therefore, the detection accuracy of small defects can be effectively guaranteed.

Most of the aforementioned methods improve the feature extraction ability of the network by enhancing features, thereby improving the detection accuracy of weak defects. However, the deep-learning approach described above may lose key features of the weak defects during feature extraction. This is mainly due to the following two reasons. 1) Some weak defects may account for only a small fraction of the original data, which may be compressed to an unrecognizable level after downsampling, resulting in the key defect features being drowned out. 2) These methods use fixed weights to superimpose features from different layers, which do not give greater weights to the key feature layers and thus lead to the key features of weak defects may be ignored.

For data preprocessing methods, researchers improve the accuracy of defect detection by enhancing the quality of the data. In [13], a novel wavelet threshold denoising method based on discrete wavelet transform (DWT) is designed to perform image denoising. In [14], a wavelet-inspired invertible network is proposed to learn a nonlinear redundant transform with perfect reconstruction property to facilitate noise removal. In [15], a novel contrast enhancement conditional generative adversarial network is proposed to enhance images. However, these methods require the design of a separate structure to preprocess the input data, which does not take into account the detection network. Therefore, the high-quality images generated by these methods may lose image details that are useful for the detection network, which reduces the defect detection performance.

To address the above issues, a dynamic weights-based wavelet attention neural network (DWWA-Net) is proposed in this article, which can extract and retain key features of weak defects. Meanwhile, DWWA-Net improves data quality by adaptively filtering out background noise based on the input data. Specifically, a dynamic wavelet convolution network (DWCNet) is designed, which can be trained with the detection network to adaptively filter out background noise and retain key features of defects. In addition, a multiview attention module and a dynamic feedback module are proposed to effectively extract the key features of weak defects. Multiview attention modules can enhance the features of weak defects and capture the long-distance dependence between different pixels. Dynamic feedback modules can assign higher weights to relevant features with weak defects and lower weights to irrelevant features. Overall, the contributions of this article can be summarized as follows.

1) A DWCNet is designed, in which wavelet transform and convolution are dynamically connected to filter background noise and improve model convergence.
2) A multiview attention module is designed, which efficiently directs the network's attention toward potential targets, thereby improving the detection results of weak defects in complex backgrounds.
3) A dynamic feedback module is proposed to screen out key features of weak defects and enhance them through network training.
4) Experimental results demonstrate that the proposed DWWA-Net outperforms the state-of-the-art method under multiple datasets, demonstrating an improved generalizability.

The rest of this article is organized as follows. Section II introduces the method. Section III gives the experiments and the comparison of the results in detail. Section IV provides a conclusion.

II. Proposed Method

A. Architecture of DWWA-Net

In this section, an overview of the DWWA-Net is presented. Fig. 2 shows the architecture of the DWWA-Net, which is divided into three phases.

Phase 1 (Feature Extraction): First, a DWCNet is designed to adaptively filter out background noise and extract shallow features. Then, a multiview attention module is designed to extract effective feature information. The extracted features can be obtained. Details are described in Section II-B.

Phase 2 (Feature Fusion): The extracted features are passed through a feature pyramid network. The fused features can be acquired.

Phase 3 (Dynamic Feedback): The fused features are fed into the feature weight network to obtain the weighted weights. The weighted features are fed into the attention network to obtain enhanced features. The enhanced features are fed back to feature extraction and feature fusion to further enhance the features. Finally, the fused features are summed with enhanced features to obtain the final features. Details are described in Section II-C.

Phase 4 (Detection Module): The final features are used to generate an anchor box. Then, the detection module calculates the loss between the anchor box and the ground truth box.

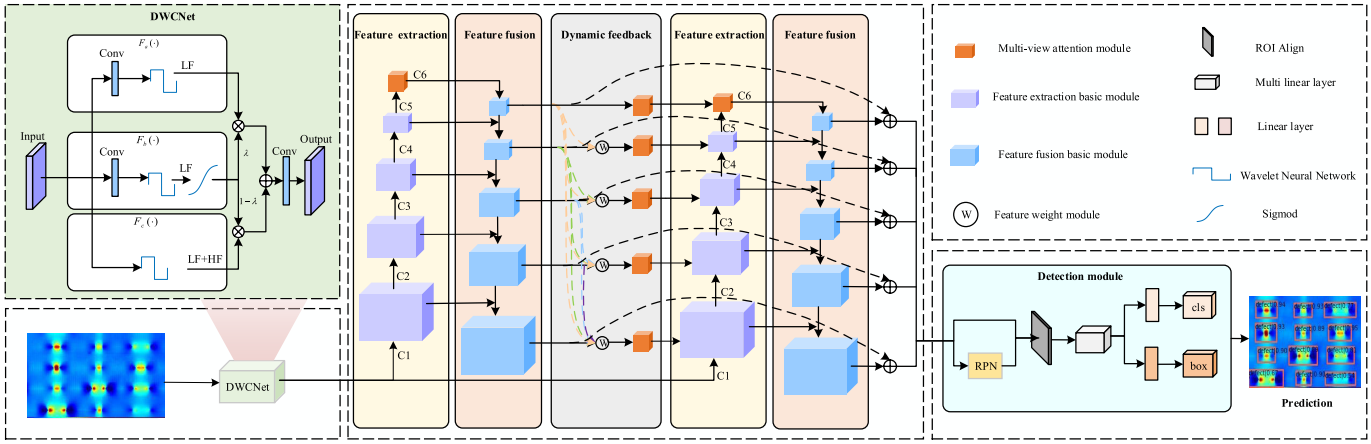


Fig. 2. Architecture of DWWA-Net, first, the image can suppress noises through the DWCNet. Then feature extraction module can extract features from pictures. In particular, the introduction of the attention module can help the feature extraction module highlight important features. In the feature fusion module, a dynamic feedback module is proposed to fuse features at different levels. The fused feature is input into the detection module to calculate the classification and regression loss. Finally, the parameters of the network are updated by minimizing losses.

Finally, the parameters of the three stages are updated by minimizing losses.

B. Wavelet Neural Network and DWCNet

Measurement error, environmental interference, equipment failure, and other factors will lead to the existence of industrial data noise, which can reduce the signal-to-noise ratio of data and make it difficult for defects to be detected effectively. Methods of denoising data using Gaussian filtering and its variants may destroy the underlying structure of the signal. Specifically, deep learning employing Gaussian filtering to noise reduction on industrial data needs to downsample the industry to compress the features. Downsampling operations such as pooling may lose key features of weak defects. To maintain the basic structure of the signals and filter the noise, a wavelet neural network is designed, which can be trained with DWWA-Net to obtain higher quality denoising features. In addition, DWCNet is designed, which can adaptively select the parameters for the wavelet convolution module and the convolution module. Therefore, the convergence of the model is improved. The specific structure is shown in DWCNet of Fig. 2. The specific design details are as follows.

1) Wavelet Neural Network: First, a wavelet neural network needs to be designed, which updates the kernel function of the wavelets through training. For this purpose, we transform the wavelet neural network into a convolutional form that can backpropagate. The specific details are as follows.

Image feature information can be extracted by DWT, specifically

$$wt(q,k) = \left\langle x(t), \psi_{q,k}(t) \right\rangle = \frac{1}{\sqrt{2^q}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-k2^q}{2^q}\right)dt \tag{1}$$

where $x(t)$ represents the input signals, q represents the scale, k represents the translation coefficient, $\langle\cdot\rangle$ represents the inner product operation, and $\psi_{q,k}(t)$ represents the base wavelet family.

Then, multiresolution analysis is used to decompose the original picture signal into high-frequency (HF) components and low-frequency (LF) components. Through multiresolution analysis, the signals $x(t)$ can be expressed as follows:

$$x(t) = \sum_{q=-\infty}^{Q}\sum_{k=-\infty}^{\infty} d_{q,k}\psi_{q,k}(t) + \sum_{k=-\infty}^{\infty} a_{q,k}\phi_{q,k}(t) \tag{2}$$

where $a_{q,k}$ is the approximation coefficients, $a_{q,k} = \langle x(t), \phi_{q,k}(t)\rangle$, $d_{q,k}$ is the detail coefficients, and $d_{q,k} = \langle x(t), \psi_{q,k}(t)\rangle$.

Finally, to convert the wavelet transform into a convolution operation, two-scale-equation is used to express the internal relationship between the scale function $\phi(t)$ and the wavelet function $\psi(t)$, which is shown as follows:

$$\phi(t) = \sqrt{2}\ h(n)\phi(2t-n) \tag{3}$$

$$\psi(t) = \sqrt{2}\ g(n)\phi(2t-n) \tag{4}$$

where $h(n)$ and $g(n)$ are a pair of low- and high-pass filter banks. Through the above analysis, in DWT, $a_{j,k}$ and $d_{j,k}$ can be obtained through convolution operation with low-pass filter $h(n)$ and high-pass filter $g(n)$. Therefore, the image can be decomposed into LF components X_{ll} and HF components X_{lh}, X_{hl}, X_{hh}, as shown in Fig. 3.

As shown in Fig. 3, X_{lh}, X_{hl}, and X_{hh} contain noise information, and X_{ll} contains the basic structure of the image. HF information can be removed to filter the noise in the signals, so a wavelet neural network can replace average pooling and maximum pooling.

2) Network Structure of DWCNet: The DWCNet includes three parts, namely convolution module $F_a(\cdot)$, weight distribution module $F_b(\cdot)$, and wavelet convolution module $F_c(\cdot)$. The specific structure is shown in Fig. 2.

$F_a(\cdot)$ is used to extract feature information of the image, where the wavelet transform is used for filtering and downsampling. The $F_b(\cdot)$ is used to obtain a dynamic weight factor λ, where the network assigns corresponding weights through training. $F_c(\cdot)$ extracts the HF components and LF components of the image. Therefore, the $F_a(\cdot)$ module extracts the local

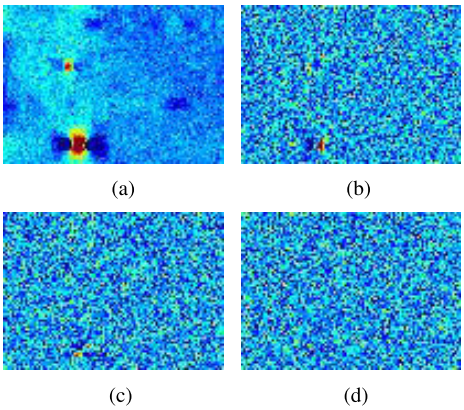


Fig. 3. Image after wavelet neural network, (a) X_{ll} is the LF components after wavelet neural network, (b) X_{lh}, (c) X_{hl}, and (d) X_{hh} are the HF components after wavelet neural network.

features of the data by convolution operation, while the $F_c(\cdot)$ module captures the global features of the data by using wavelet transform. By combining local and global features to better capture the feature information of defects in industrial data, the detection accuracy of weak defects can be improved.

At the beginning of the network training, the $F_c(\cdot)$ can select the appropriate wavelet basis function to extract features more effectively than using $F_a(\cdot)$ with random initialization for feature extraction. Therefore, at the beginning of the network training, we can give higher weights to the $F_c(\cdot)$ by the $F_b(\cdot)$. Based on the above analysis, the weight of λ is set to 0, which can be achieved through the weight distribution module $F_b(\cdot)$. Then, with the iterative training of the network, the network will learn the weight parameters by $F_b(\cdot)$, such that the network will choose the weight adaptively.

C. Multiview Attention Module

In order to achieve reliable detection of weak defects in complex industrial backgrounds, the multiview attention module is proposed, which efficiently directs the network's attention toward potential targets, thereby improving the identification of weak defects. Unlike literature [16], [17], this article designs a multiview attention module that considers irregular and weak defects in industrial data. Specifically, three parallel feature extractions are designed to enhance the focus on defects of different sizes through different receptive fields, which can capture the details of defects from multiple perspectives to obtain the key features of weak defects more comprehensively. Deformable convolution enhances the network's ability to recognize irregularly shaped defects and unknown defects. Besides, DWCNet in Section II-B is added to filter the noise and assist network training. The specific attention structure is shown in Fig. 4, and the specific steps are as follows.

1) First, three different deformable convolutions [18] are used to extract the features of the feature maps, where the kernel sizes of the deformable convolution are 1, 3, and 5. The use of deformable convolution can improve the network's ability to detect deformable defects, and different convolution kernel sizes provide different receptive fields.

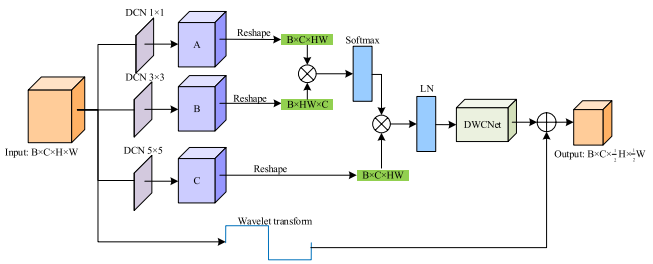


Fig. 4. Architecture of multiview attention module.

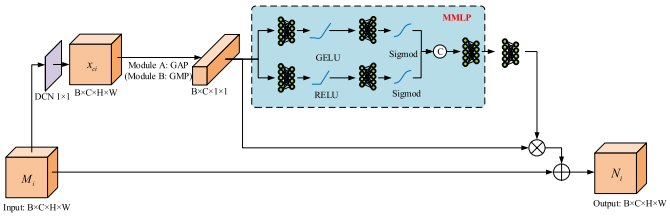


Fig. 5. Architecture of attention modules A and B.

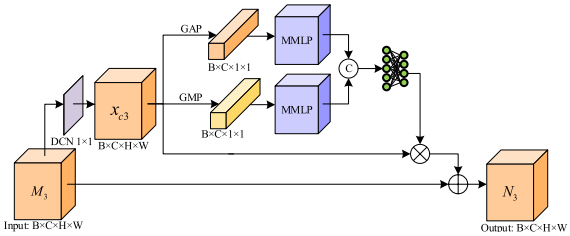


Fig. 6. Architecture of attention module C.

2) Second, three attention modules A, B, and C are designed to enhance the features in 1) and obtain new feature maps N_1, N_2, and N_3. The specific structures are shown in Figs. 5 and 6.
The architectures of A, B, and C are designed by referring to squeeze-and-excitation (SE) networks [19]. SE network can transform the feature map into a feature vector through multiple linear layers. Different from SE network, in this work, the feature maps are convolved to obtain new feature maps, and the information is compressed through multiple multilayer perceptron (MMLP). Therefore, compared with SE Network, residual structure and multiview are proposed to improve the stability and feature diversity of the network. The specific equation of attention modules A and B is as follows:

$$\begin{cases} x_{ci} = W_{1\times1} \otimes M_i + b \\ x_f^i = \text{Sigmoid}(g(\sigma(g(P(x_c), w)), w)) \\ N_i = M_i + g(\text{cat}(x_f^1, x_f^2), w) \otimes x_{ci} \end{cases} \tag{5}$$

where M_i represents the feature map of the input, P represents global pooling, the specific equation is $(1/(H \times W)) \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j)$, $g(\cdot)$ represents a fully connected network, and w is the weight of $g(\cdot)$. $\sigma(\cdot)$ represents a activation function and N_i represents the output features.

In module C, the feature maps are compressed into feature vectors through two different global pooling methods as well as the MMLP, and the results are finally stacked. The specific equation of attention module C is as follows:

$$\begin{cases} x_{c3} = W_{1\times1} \otimes M_3 + b \\ x_f^i = \text{Sigmoid}(g(\sigma(g(P(x_c), w)), w)) \\ N_3 = M_i + g\left(\text{cat}\left(x_f^1, x_f^2, x_f^3, x_f^4\right), w\right) \otimes x_{c3}. \end{cases} \tag{6}$$

3) Feature maps N_1 and N_2 are reshaped into $R^{B\times C\times HW}$, $R^{B\times HW\times C}$ and then multiplied. The results are passed through the softmax layer and multiplied by N_3. The attention networks of the feature extraction module and feature fusion module are designed differently. In the feature extraction module, DWCNet is added to filter the noise of the image and downsample the image. In the feature fusion module, DWCNet is removed to ensure that the sizes of feature maps are consistent. The specific equation for the above steps is shown as follows:

$$\begin{cases} L = \dfrac{e^{(N_1\otimes N_2)_i}}{\sum_{j=1}^{C} e^{(N_1\otimes N_2)_j}} \otimes N_3 \\ K = \dfrac{L-\mu^l}{\sqrt{(\eta^l)^2+\varepsilon}} \end{cases} \tag{7}$$

where L is the feature map after multiplying N_1, N_2, and N_3, μ is the shift parameter, and η is the scaling parameter.

D. Dynamic Feedback Module

In order to further enhance the feature information extracted by the feature extraction module and improve the detection accuracy of weak defects, a dynamic feedback module is designed. Previously, there are methods aiming to enhance the feature information extracted by the feature extraction module, such as YolTrack [20]. However, these methods do not highlight the extracted features and provide feedback to the feature extraction module. In other words, they only fuse features and only marginally improved model performance. Different from the above methods, the dynamic feedback module can dynamically connect feature maps at different levels and enhance features using the attention network of Section II-C. In addition, these features can be fed back to the feature extraction module, resulting in further enhancements. The design steps are as follows.

1) As shown in the feature fusion module of Fig. 2, the extracted features are fused as follows:

$$y_i = U(x_{i+1}) + (W_{1\times1} \otimes x_i + b) \tag{8}$$

where $U(\cdot)$ represents upsampling, specifically linear interpolation, x_i represents the feature map of the ith level.

2) The fused feature y_i is dynamically connected. Here we take the feature map y_1 as an example to better illustrate the process of a dynamic connection. In the feature fusion process of y_1, the feature maps y_5 and y_4 need to be dynamically fused first, and the specific fusion structure is shown in Fig. 7.

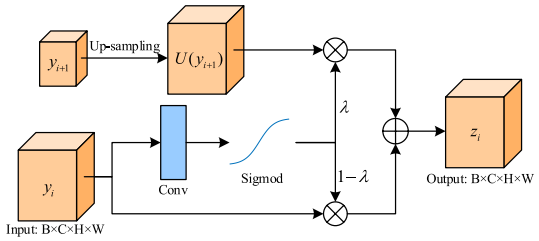


Fig. 7. Architecture of feature weight module.

The feature map y_4 generates weight λ through convolution operation, which can be used in the dynamic connection between y_4 and y_5. The specific equation is

$$\begin{cases} \lambda = W_{1\times1} \otimes y_4 + b \\ z_4 = U(y_5) \cdot (1-\lambda) + y_4 \cdot \lambda. \end{cases} \tag{9}$$

Therefore, a new feature map z_4 can be obtained. Similar to the above operation, z_4 can be dynamically added with y_3 to obtain z_3, and finally, z_1 is obtained.

3) Through the multiview attention module of (Section II-C), the fused features z_i enhance important features of defects and suppress irrelevant details. Therefore, the network can extract clear and richer feature information.
4) The enhanced features are input to the feature extraction module to obtain feedback features m_i. The fused features z_i and the feedback features m_i are added according to the weights to obtain the final features v_i, which can be input into the head network to generate anchors for regression. The specific equation is

$$\begin{cases} \xi = W_{1\times1} \otimes z_i + b \\ v_i = z_i \cdot \xi + m_i \cdot (1-\xi). \end{cases} \tag{10}$$

The details of the proposed method are shown in Algorithm 1.

III. Experimental Results and Analysis

A. Dataset

We employ five datasets to evaluate our method, namely NEU-DET [21], GC10-DET [22], MFL dataset, MFL noise data, and VOC dataset.

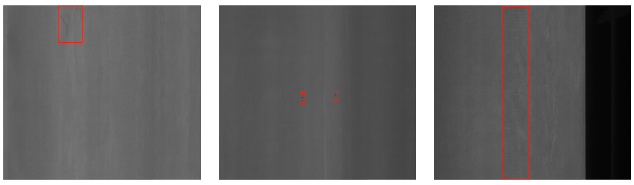
1) *GC10-DET:* GC10-DET is a dataset of surface defects collected in real industry. The dataset includes 3570 grayscale images with ten types of defects, including crescent gap, welding line, water spot, silk spot, inclusion, oil spot, crease, punching, waist folding, and rolled pit. Examples of defect images are shown in Fig. 8.
2) *NEU-DET:* NEU-DET is the Northeastern University (NEU) surface defect dataset. The collected defects are on the surface of the hot-rolled steel strip. There are six types of defects in NEU-DET, including inclusion, crazing, patches, pitted surface, scratches, and rolled-in scale. The dataset includes 1800 gray-scale images, i.e.,

Algorithm 1 The Algorithm Flow of DWWA-Net

Input: The input images G, Feature extraction network F_i, global pooling P, detection module T, regression loss L_r, classification loss L_c, ground truth g_i

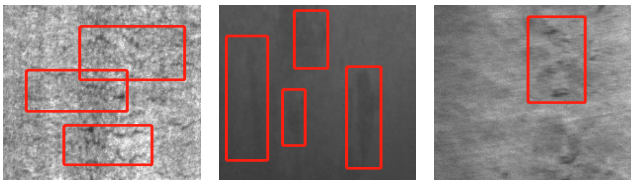
Output: The trained model of DWWA-Net

for each G_i in G **do**
  $D = F_b(G_i) \cdot F_a(G_i) + (1 - F_b(G_i)) \cdot F_c(G_i)$
  $x_i = \odot_{i=1\ldots4} F_i^{L_i}(D_{\langle H_i, W_i, C_i \rangle})$
  $x_{c1} = (W_{1\times1} \otimes M_1 + b)$, $x_{c2} = (W_{3\times3} \otimes M_2 + b)$, $x_{c3} = (W_{5\times5} \otimes M_3 + b)$
  for n in range(2) **do**
    $x_f^i = Sigmoid(g(\sigma(g(P(x_c), w)), w))$
    $N_i = M_i + g(cat(x_f^1, x_f^2), w) \otimes x_c$
  end for
  $N_3 = M_i + g(cat(x_f^1, x_f^2, x_f^3, x_f^4), w) \otimes x_c$
  $L = \frac{e^{(N_1 \otimes N_2)_i}}{\sum_{j=1}^{C} e^{(N_1 \otimes N_2)_j}} \otimes N_3$
  $K = \frac{L - \mu^l}{\sqrt{(\eta^l)^2 + \varepsilon}}$
  $x_5 = F_b(K) \cdot F_a(K) + (1 - F_b(K)) \cdot F_c(K)$
  $y_i = U(x_{i+1}) + (W_{1\times1} \otimes x_i + b)$
  $\lambda = W_{1\times1} \otimes y_i + b$, $z_i = U(y_{i+1}) \cdot \lambda + y_i \cdot (1 - \lambda)$
  $\xi = W_{1\times1} \otimes z_i + b$, $v_i = z_i \cdot \xi + m_i \cdot (1 - \xi)$
  $(p_i^r, p_i^c) = T(v_1, \ldots, v_i)$
  $Loss = L_r(p_i^r, g_i^r) + L_c(p_i^c, g_i^c)$
  $argmin\ Loss$ with SGD;
end for
return the model of trained DWWA-Net



(a) (b) (c)

Fig. 8. Examples of defect images in GC10-DET. (a) Silk spot. (b) Inclusion. (c) Waist folding.



(a) (b) (c)

Fig. 9. Examples of defect images in NEU-DET. (a) Crazing. (b) Inclusion. (c) Rolled in scale.

300 samples in each class of surface defects, which is shown in Fig. 9.

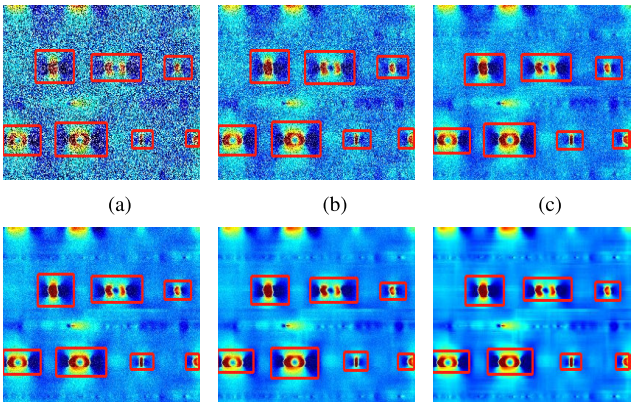
3) *MFL Dataset:* The MFL data is from the pipeline loop platform, which includes both artificial and natural corrosion defects. We divide MFL data into two types: 1) conventional MFL data and 2) MFL data under strong background noise. The MFL data under strong background noise is generated by adding Gaussian noise. The training data is from a 12-in diameter pipeline. The length of the pipelines is 100 m, along with 12.7 mm of wall thickness. The number of test data is from a 16-in diameter pipeline. The length of the pipelines is 200 m, along with 12.7 mm of wall thickness. The number of defects is 200. Different peak signal-to-noise ratios (PSNRs) are used to represent the magnitude of the noise. The PSNR is defined as follows:

$$\mathrm{MSE} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [I(i,j) - K(i,j)]^2 \quad (11)$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \quad (12)$$

where $I(i, j)$ represents the pixel value in i row and j column of the original image, and $K(i, j)$ represents the pixel value in i row and j column of the noise image. H and W represent the height and width of the image, MAX_I^2 represents the maximum value of the image pixel. According to the above equations, the PSNR value of dataset (3) is 35 dB, five levels of Gaussian white noise are added to the original signals, and the PSNR of the signals after adding noise is set as 30, 25, 20, 15, and 10 dB, respectively, as shown in Fig. 10. The pipeline loop test platforms and pipeline defects are shown in Figs. 11 and 12.

4) *VOC Dataset:* Besides the defect dataset mentioned above, the VOC dataset is also used to evaluate the effectiveness of our method. This dataset consists of about 5k trainval images and 5k test images over 20 object categories. In this article, other state-of-the-art object detection algorithms are used to compare with our method.



(a) (b) (c) (d) (e) (f)

Fig. 10. Display of MFL defects at different PSNRs. (a) PSNR10 dB. (b) PSNR15 dB. (c) PSNR20 dB. (d) PSNR25 dB. (e) PSNR30 dB. (f) PSNR35 dB.

Fig. 11. Pipeline loop test platform.

(a) (b) (c)

Fig. 12. Picture of pipeline defects. (a) Artificial defects. (b) and (c) Natural corrosion defects.

B. Experimental Setup

1) Implementation Details: The DWWA-Net hyperparameter settings are as follows. The optimizer is the stochastic gradient descent (SGD) algorithm. The optimizer momentum is 0.9. The initial learning rate is 0.005, where its optimization method is MultiStepLR. The size of the network input is resized to 1333 × 888. The batch size is 4. The hardware environment of the experimental platform: 2 AMD EPYC 7601 2.7 GHz 32-core processors, 1 GeForce RTX 3090 graphics card, 128 GB RAM, LINUX64 Ubuntu18.04 software environment, and the PyTorch framework.

2) Comparison Methods: The 11 following methods are compared. 1) Adaptive faster R-CNN [23], which intelligently selects the ideal preset anchors from 3-D printed lattice structures. 2) TOL-Framework [24], which cascades a regression module inside the basic framework and adds an external postprocess network. 3) YOLO-MF [25], which accelerates the algorithm and median flow (MF) algorithm to count the number of cracks. 4) YOLOF [26], which proposes dilated encoder and uniform matching. 5) Dynamic R-CNN [27], which adjusts the label assignment criteria and the shape of the regression loss function adaptively. 6) Multiscale-SSD [28], which implements a dilated convolution and attention residual module. 7) DCC-CenterNet [29], which proposes a steel surface defect detector to achieve the best speed-accuracy tradeoff. 8) DDN [21], which proposes a multilevel feature fusion network combines that combines multiple hierarchical features into a single feature. 9) YOLO V7 [30], which focuses on proposing a lightweight and accurate detection framework. 10) ResNeSt [31], which introduces a multibranch architecture that leverages channelized attention in different network branches. 11) ConvNet [32], which modernizes a standard ResNet toward the design of a vision transformer.

C. Evaluation Metric

We evaluate the performance of the proposed method by the precision and the recall rate. The precision reflects the accuracy of defect detection, and the recall reflects the detection rate of defects. The high recall rate and precision reflect good detection performance. The equations are

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \cdot 100\% \tag{13}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \cdot 100\% \tag{14}$$

$$\text{mAP} = \frac{1}{n} \times \sum_{i=1}^{n} P_i \tag{15}$$

$$\text{F-score} = (1+\beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \tag{16}$$

where true positive (TP) accounts for correctly classified defects and false negative (FN) accounts for missed classifications. True negative (TN) accounts for correctly classified nondefects and false positive (FP) accounts for false classifications. S is the number of all samples and β is utilized to measure the importance of accuracy and recall rate. For defect detection, the recall rate is more important, and thus β is set to 2 in this article.

D. Analysis of Results

1) GC10-DET: Table I shows the comparison results of defect detection. With the backbone ResNet 50, the mAP of DWDA-Net is 6.0% higher than the highest detection algorithm above (YOLO V7). Specifically, when the defect features are obvious, such as punching and crescent gap, the detection performance of each method is excellent. However, when the defects are small or the features are not obvious, such as silk spot and inclusion, it is difficult for other detection methods to effectively identify such defects. By contrast, the attention network and dynamic feedback network of DWWA-Net can better focus on potential targets, thereby improving the detection success rates of such defects. In addition, the mAP of small defects is also used to evaluate the performance of DWDA-Net. Compared with the above-mentioned method with the highest accuracy (YOLO V7), DWWA-Net further improves the mAP by 9.6%, which proves the effectiveness of the method in this article. Finally, it can be seen from Table I that DWWA-Net has great detection rates on various types of defects, which can meet the actual industrial requirements.

2) NEU-DET: Table II shows the results of defect detection experiments. As Table II shows, DWDA-Net can also achieve excellent detection results in the NEU dataset. Specifically, DWDA-Net can improve mAP by 4.3% compared to the detection algorithm with the highest accuracy (YOLO V7). In particular, DWWA-Net has a more significant detection effect on weak defects (crazing and rolled in scale) of the NEU-DET. This is because weak defects have subtle features, which makes it difficult to achieve effective detection. However, the multiview attention module and dynamic feedback module strengthen the feature expression of weak defects, thus guaranteeing the detection accuracy of weak defects. Finally, when the defect target is small, DWWA-Net can improve the mAP by 14% compared with the method with the highest accuracy (YOLOF) mentioned above, which verifies that DWWA-Net can successfully detect weak defects.

3) MFL Data: As Table III shows, MFL data are used to evaluate the effectiveness of DWWA-Net. Compared with other detection algorithms, DWWA-Net achieves an average improvement of 6.4% and 3.9% in detection precision and recall. Besides, the F2 score is introduced to verify the effectiveness of this method. mAF2 has been improved by an average of 7.5%, providing better performance of DWWA-Net.

As Fig. 13 shows, when the collected MFL data contains noise, the detection accuracy of different detection methods

TABLE I
DEFECT DETECTION RESULTS ON GC10-DET DATA

Method	mAP	mAP_s	Punching	Welding line	Crescent gap	Water spot	Oil spot	Silk spot	Inclusion	Rolled pit	Crease	Waist folding
Adaptive Faster R-CNN [23]	60.2	23.2	95.4	69.7	93.0	71.4	55.3	55.8	20.1	33.4	29.7	78.1
TOL-Framework [24]	63.0	28.2	98.3	85.2	90.5	73.8	59.1	59.1	36.5	41.7	13.5	72.6
YOLO-MF [25]	57.5	16.8	94.0	77.8	90.4	74.5	48.4	48.8	17.6	22.9	15.8	85.2
YOLOF [26]	61.4	17.9	97.2	57.6	89.4	76.1	63.9	56.6	31.7	33.3	25.6	82.4
Dynamic R-CNN [27]	57.5	13.4	95.2	44.8	91.9	76.9	61.7	58.5	35.4	20.7	22.3	67.9
Multiscale-SSD [28]	61.3	17.4	95.2	90.2	88.8	66.6	49.8	44.1	18.8	47.4	39.0	73.5
DCC-CenterNet [29]	61.9	29.4	84.1	85.5	96.2	77.3	50.8	54.7	30.2	13.9	49.9	76.6
DDN [21]	61.7	31.2	97.2	56.2	95.3	69.4	59.5	52.6	34.7	40.1	37.2	75.2
YOLO V7 [30]	66.1	33.8	97.6	85.2	90.6	70.7	68.5	56.1	35.7	37.2	43.6	75.9
ResNeSt [31]	64.4	29.8	98.3	75.7	94.2	75.4	59.3	61.0	34.7	33.4	31.0	81.0
ConvNet [32]	65.8	32.0	97.5	79.9	93.7	78.2	63.4	51.8	37.3	30.6	42.4	83.6
DWWA-Net	**72.1**	**43.4**	**98.9**	**91.4**	**93.4**	**82.1**	**63.9**	**67.2**	**46.7**	**42.5**	**43.7**	**90.8**

TABLE II
DEFECT DETECTION RESULTS ON NEU-DET DATA

Method	mAP	mAP_s	Crazing	Inclusion	Patches	Pitted Surface	Rolled in scale	Scratches
Adaptive Faster R-CNN [23]	70.5	52.5	41.4	72.3	85.5	77.8	55.1	91.0
TOL-Framework [24]	76.0	61.4	42.7	78.4	89.2	81.7	69.4	94.4
YOLO-MF [25]	73.2	53.1	41.2	74.3	89.1	77.2	64.8	92.6
YOLOF [26]	73.9	63.7	37.6	75.9	86.0	82.3	66.7	94.9
Dynamic R-CNN [27]	73.5	60.5	41.8	74.2	89.2	80.4	61.7	93.6
Multiscale-SSD [28]	73.3	45.6	44.6	74.8	87.9	85.4	69.0	78.2
DCC-CenterNet [29]	72.9	56.8	40.7	78.1	79.6	80.5	65.8	92.9
DDN [21]	72.0	50.3	35.1	78.4	91.4	86.1	57.7	83.5
YOLO V7 [30]	76.1	62.3	44.3	79.0	90.8	84.5	66.5	91.3
ResNeSt [31]	73.9	54.7	39.4	76.9	90.1	83.7	62.6	91.0
ConvNet [32]	74.6	55.8	36.4	78.2	90.4	83.0	66.6	93.1
DWWA-Net	**80.4**	**77.7**	**54.1**	**81.0**	**90.8**	**85.3**	**75.4**	**95.5**

TABLE III
DEFECT DETECTION RESULTS ON MFL DATA

Method	mAP_s	mAP	mAR	$mAF2$
Adaptive Faster R-CNN [23]	78.4	82.6	84.5	84.1
TOL-Framework [24]	78.9	83.0	85.4	84.9
YOLO-MF [25]	82.8	84.9	90.5	89.3
YOLOF [26]	78.8	82.3	91.6	89.6
Dynamic R-CNN [27]	79.5	84.0	84.3	84.2
Multiscale-SSD [28]	81.5	84.1	92.4	90.6
DCC-CenterNet [29]	79.8	84.5	91.4	89.9
DDN [21]	79.5	83.8	88.5	87.5
YOLO V7 [30]	83.2	85.6	88.2	87.7
ResNeSt [31]	83.4	85.9	89.1	88.4
ConvNet [32]	80.8	84.5	88.7	87.8
DWWA-Net	**90.3**	**92.4**	**95.8**	**95.1**

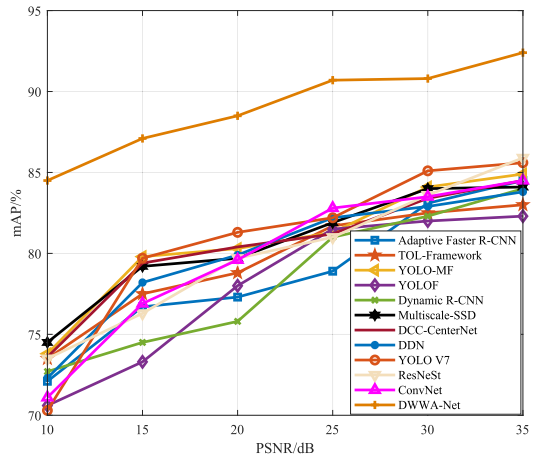


Fig. 13. mAP values of defects under different PSNRs.

will decrease. However, compared with other detection methods, DWWA-Net still has great detection results. In particular, at a PSNR of 10 dB, DWWA-Net can improve mAP by 10.1% compared to the method with the highest detection accuracy (Multiscale-SSD). At a PSNR of 35 dB, DWWA-Net can improve mAP by 6.8% compared to the method with the highest detection accuracy (YOLO V7). Through the comparison, it can be found that DWWA-Net has a better performance with noise and can detect defects in low image quality.

4) VOC Dataset: To verify the effectiveness of DWWA-Net on the object detection dataset, the VOC dataset is used. During training, random flips are used to augment the data. As can be seen from Table IV, compared with the above detection algorithm with the best detection effect (MFFAMM), DWWA-Net can still improve mAP by 2.3%, indicating the effectiveness of our method.

E. Ablation Study

1) Analysis of Each Module of the DWWA-Net: First, we use the MFL data at 10PSNR to evaluate the performance of each part of our method. As shown in Table V, DWCNet, feedback network, and attention can improve the detection effect of defects, and the proposed attention network has a crucial impact on the performance of the network. This is because in the MFL data, the proportion of small targets is

TABLE IV
DEFECT DETECTION RESULTS ON VOC DATASET

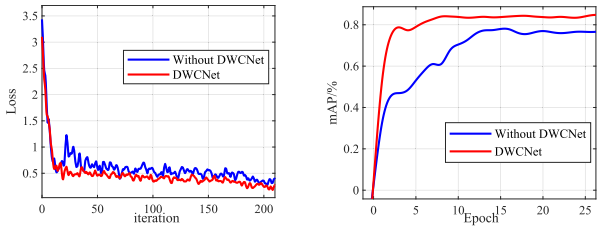
Method	mAP
Faster R-CNN [33]	76.4
ScratchDet300 [34]	80.4
Dual-YOLO V3 [35]	80.3
DSOD [36]	77.7
GRP-DSOD [37]	78.7
DES [38]	79.7
DSSD [39]	78.6
RefineDet [40]	80.0
Multiscale-SSD300 [28]	77.1
MFFAMM [41]	80.7
DWWA-Net	**83.0**

TABLE V
PERFORMANCE OF OUR METHOD UNDER DIFFERENT SETTINGS

DWCNet	wavelet neural network	Multi-view attention module	Dynamic feedback module	mAP	mAR
✓				74.8	78.6
✓	✓			76.9	81.6
✓	✓	✓		81.6	84.6
✓	✓	✓	✓	84.5	89.7

TABLE VI
PERFORMANCE OF OUR METHOD UNDER DIFFERENT SETTINGS

wavelet neural network	average pooling	max pooling	mAP	mAR
	✓		82.3	85.1
		✓	82.6	85.5
✓			84.5	89.7



(a) (b)

Fig. 14. Experimental comparison of convergence in the method. (a) Loss values obtained by training under MFL noisy data. (b) mAP values obtained by training under MFL noisy data.

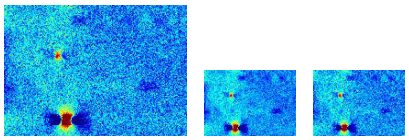
large, and the attention network can have a better recognition effect for these potential targets.

Second, we conduct ablation experiments to analyze the denoising ability of the wavelet neural network and the pooling. The specific results are shown in Table VI.

As can be seen from Table VI, the wavelet neural network can filter the noise in the signal and retain the key features of weak defects, to improve the precision and recall rate of defect detection.

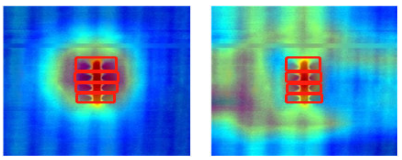
2) Convergence Analysis of the Model: To verify that DWCNet can effectively improve the convergence of the model, we conducted a comparison test of mAP and loss values, and the specific experimental results are shown in Fig. 14.

As shown in Fig. 14, the loss value of DWCNet can be lower and the initial mAP value can be higher when the network is first trained. In addition, the loss value of DWCNet decreases rapidly and the mAP value increases sharply during the first few epochs of training, which proves that choosing the appropriate wavelet basis function can improve the convergence of the model. Finally, DWCNet can have a higher upper limit at the end of network training. Based on the above analysis, the method designed in this article can effectively improve the convergence of the model.



(a) (b) (c)

Fig. 15. Downsampling visualization. (a) Original image, whose PSNR value is 18.1 dB. (b) Image after wavelet neural network, whose PSNR value is 20.6 dB. (c) Maximum pooling whose PSNR value is 19.5 dB.



(a) (b)

Fig. 16. Attention visualization results. (a) Attention result map of this article. (b) Attention result map without the multiview attention module.

3) Visualization Analysis: We visualized the downsampling operation on the image. As shown in Fig. 15, the wavelet neural network filters noise more effectively.

In addition, we perform a visual analysis of the attention designed in this article, shown in Fig. 16. The attention used in this article effectively directs the network's attention toward the underlying target information and improves the detectability of signals under complex backgrounds.

IV. CONCLUSION

This article proposes a DWWA-Net, which can solve the problem of the poor defect detection rate caused by the complexity of defects in the real world. There are three phases in this method: 1) wavelet neural network with dynamic weights; 2) multiview attention module; and 3) feedback module. In the first phase, the influence of background noise is reduced and the convergence of the network is enhanced. In the second phase, the network is designed to pay attention to the target signals and improves the detection rates. In the last phase, a dynamic feedback network is designed to improve the feature extraction capability.

Multiple real-world datasets are used to comprehensively evaluate the performance proposed method. Based on the evaluation results, we have shown that DWWA-Net can achieve significantly improved detection results when compared with other methods. Our comprehensive evaluation demonstrates that the proposed method is effective and well-suited for use in various industrial applications. In the future, we plan to address the following issues.

1) Although the proposed method achieves excellent performance on several industrial datasets, it still requires

a certain amount of sample data. In some cases, the insufficient number of samples can make it difficult to ensure satisfactory detection accuracy. Hence, we hope to overcome the high dependence on manual marker annotation to achieve high-accuracy defect detection with few samples.

2) In the training of the DWWA-Net network, the process of hyperparameter tuning is time-consuming. Therefore, efficient hyperparameter tuning methods should be considered [42].
3) The collected data may be incomplete due to sensor anomalies in industrial processes. In future research, we hope to draw on latent cause analysis methods [43], [44] to effectively extract knowledge from incomplete data for defect detection.

REFERENCES

[1] S. C. Tan, J. Watada, Z. Ibrahim, and M. Khalid, "Evolutionary fuzzy ARTMAP neural networks for classification of semiconductor defects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 933–950, May 2015.

[2] H. Zhang, L. Wang, J. Wang, F. Zuo, J. Wang, and J. Liu, "A pipeline defect inversion method with erratic MFL signals based on cascading abstract features," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[3] Z. Geng, C. Shi, and Y. Han, "Intelligent small sample defect detection of water walls in power plants using novel deep learning integrating deep convolutional GAN," *IEEE Trans. Ind. Informat.*, vol. 19, no. 6, pp. 7489–7497, Jun. 2023.

[4] X. Ren, W. Lin, X. Yang, X. Yu, and H. Gao, "Data augmentation in defect detection of sanitary ceramics in small and non-i.I.d datasets," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 9, 2022, doi: 10.1109/TNNLS.2022.3152245.

[5] H. Zhao, J. Liu, J. Tang, X. Shen, S. Lu, and Q. Wang, "A MFL mechanism-based self-supervised method for defect detection with limited labeled samples," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[6] J. Liu, X. Shen, J. Wang, L. Jiang, and H. Zhang, "An intelligent defect detection approach based on cascade attention network under complex magnetic flux leakage signals," *IEEE Trans. Ind. Electron.*, vol. 70, no. 7, pp. 7417–7427, Jul. 2023.

[7] A. AlBahar, I. Kim, and X. Yue, "A robust asymmetric kernel function for Bayesian optimization, with application to image defect detection in manufacturing systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 3222–3233, Oct. 2022.

[8] H. Mei, H. Jiang, F. Yin, L. Li, and L. Wang, "Detection of small defects in composite insulators using terahertz technique and deconvolution method," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8146–8155, Oct. 2020.

[9] T. Liu, Z. He, Z. Lin, G. Cao, W. Su, and S. Xie, "An adaptive image segmentation network for surface defect detection," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 29, 2022, doi: 10.1109/TNNLS.2022.3230426.

[10] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[11] D. Li, Y. Li, Q. Xie, Y. Wu, Z. Yu, and J. Wang, "Tiny defect detection in high-resolution aero-engine blade images via a coarse-to-fine framework," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[12] W. Liang and Y. Sun, "ELCNN: A deep neural network for small object defect detection of magnetic tile," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[13] H. Li, J. Shi, L. Li, X. Tuo, K. Qu, and W. Rong, "Novel wavelet threshold denoising method to highlight the first break of noisy microseismic recordings," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5910110.

[14] Y. Chang, M. Chen, L. Yan, X. Zhao, Y. Li, and S. Zhong, "Toward universal stripe removal via wavelet-based deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2880–2897, Apr. 2020.

[15] R. Guo, H. Liu, G. Xie, and Y. Zhang, "Weld defect detection from imbalanced radiographic images based on contrast enhancement conditional generative adversarial network and transfer learning," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10844–10853, May 2021.

[16] P. Wu, Z. Wang, B. Zheng, H. Li, F. E. Alsaadi, and N. Zeng, "AGGN: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106457.

[17] Z. Zheng, G. An, D. Wu, and Q. Ruan, "Global and local knowledge-aware attention network for action recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 334–347, Jan. 2021.

[18] S. Dong, Q. Yang, Y. Fu, M. Tian, and C. Zhuo, "RCoNet: Deformable mutual information maximization and high-order uncertainty-aware learning for robust COVID-19 detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3401–3411, Aug. 2021.

[19] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[20] X. Chang, H. Pan, W. Sun, and H. Gao, "YolTrack: Multitask learning based real-time multiobject tracking and segmentation for autonomous vehicles," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5323–5333, Dec. 2021.

[21] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.

[22] X. Lv, F. Duan, J.-J. Jiang, X. Fu, and L. Gan, "Deep metallic surface defect detection: The new benchmark and detection network," *Sensors*, vol. 20, no. 6, p. 1562, Mar. 2020.

[23] Y. Zhang, Z. Zhang, K. Fu, and X. Luo, "Adaptive defect detection for 3-D printed lattice structures based on improved faster R-CNN," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.

[24] J. Zhong, Z. Liu, C. Yang, H. Wang, S. Gao, and A. Núñez, "Adversarial reconstruction based on tighter oriented localization for catenary insulator defect detection in high-speed railways," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1109–1120, Feb. 2022.

[25] D. Ma, H. Fang, N. Wang, C. Zhang, J. Dong, and H. Hu, "Automatic detection and counting system for pavement cracks based on PCGAN and YOLO-MF," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22166–22178, Nov. 2022.

[26] Q. Chen, Y. Wang, T. Yang, X. Zhang, J. Cheng, and J. Sun, "You only look one-level feature," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 13034–13043.

[27] H. Zhang, H. Chang, B. Ma, N. Wang, and X. Chen, "Dynamic R-CNN: Towards high quality object detection via dynamic training," 2020, *arXiv:2004.06002*.

[28] L. Yang, Z. Wang, and S. Gao, "Pipeline magnetic flux leakage image detection algorithm based on multiscale SSD network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 501–509, Jan. 2020.

[29] R. Tian and M. Jia, "DCC-CenterNet: A rapid detection method for steel surface defects," *Measurement*, vol. 187, Jan. 2022, Art. no. 110211.

[30] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

[31] H. Zhang et al., "ResNest: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2022, pp. 2735–2745.

[32] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.

[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[34] R. Zhu et al., "ScratchDet: Training single-shot object detectors from scratch," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2263–2272.

[35] J. Zhong, J. Chen, and A. Mian, "DualConv: Dual convolutional kernels for lightweight deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 1, 2022, doi: 10.1109/TNNLS.2022.3151138.

[36] Z. Shen, Z. Liu, J. Li, Y. Jiang, Y. Chen, and X. Xue, "DSOD: Learning deeply supervised object detectors from scratch," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1937–1945.

[37] Z. Shen et al., "Learning object detectors from scratch with gated recurrent feature pyramids," 2017, *arXiv:1712.00886*.

[38] Z. Zhang, S. Qiao, C. Xie, W. Shen, B. Wang, and A. L. Yuille, "Single-shot object detection with enriched semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5813–5821.

[39] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*.

[40] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-shot refinement neural network for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4203–4212.

[41] Z. Qu, T. Han, and T. Yi, "MFFAMM: A small object detection with multi-scale feature fusion and attention mechanism module," *Appl. Sci.*, vol. 12, no. 18, p. 8940, Sep. 2022.

[42] X. Luo, Y. Yuan, S. Chen, N. Zeng, and Z. Wang, "Position-transitional particle swarm optimization-incorporated latent factor analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3958–3970, Aug. 2022.

[43] Y. Yuan, X. Luo, M. Shang, and Z. Wang, "A Kalman-filter-incorporated latent factor analysis model for temporally dynamic sparse data," *IEEE Trans. Cybern.*, early access, Jul. 25, 2022, doi: 10.1109/TCYB.2022.3185117.

[44] Y. Yuan, Q. He, X. Luo, and M. Shang, "A multilayered-and-randomized latent factor model for high-dimensional and sparse matrices," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 784–794, Jun. 2022.

Jinhai Liu (Senior Member, IEEE) received the B.S. degree in automation from the Harbin Institute of Technology, Harbin, China, in 2002, and the M.S. degree in power electronics and power transmission and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2005 and 2009, respectively.

He is currently a Professor and Doctoral Supervisor with Northeastern University. His current research interests include industrial artificial intelligence, data-driven fault diagnosis, and safety technology of long pipelines.

He Zhao received the bachelor's degree from the Shandong University of Science and Technology, Qingdao, China, in 2019, and the M.S. degree in control engineering from Northeastern University, Shenyang, China, in 2021, where he is currently pursuing the Ph.D. degree in control science and engineering with the School of Information Science and Engineering.

His current research interests include fault diagnosis, nondestructive testing, and deep learning.

Zhaolin Chen (Member, IEEE) received the B.Eng. degree in automation from the Harbin Institute of Technology, Harbin, China, in 2002, and the Ph.D. degree in MRI image reconstruction from the Department of Electrical and Computer Systems Engineering, Monash University, Clayton VIC, Australia, in 2007.

His current research interests include MRI and PET data acquisition physics, image reconstruction and data analysis using machine learning, and image analysis for neuroscience and clinical applications.

Dr. Chen is a Program Committee Member for ISMRM and a member of the Medical Image Computing and Computer-Assisted Interventions (MICCAI). He was a recipient of the Douglas Lampard Research Prize and Medal from Monash University and Cum Laude for Best Papers from the International Society of Magnetic Resonance in Medicine (ISMRM). He served as an Area Chair for IEEE ISBI Conference.

Qiannan Wang received the bachelor's degree from the Shandong University of Science and Technology, Qingdao, China, in 2019, and the M.S. degree in control engineering from Northeastern University, Shenyang, China, in 2022, where she is currently pursuing the Ph.D. degree in control science and engineering with the School of Information Science and Engineering.

Her current research interests include data-driven fault diagnosis, industrial big data analysis, and long pipeline safety technology.

Xiangkai Shen (Graduate Student Member, IEEE) received the B.S. degree in automation from the Shandong University of Science and Technology, Qingdao, China, in 2018, and the M.S. degree in control engineering from Northeastern University, Shenyang, China, in 2020.

He is currently working toward the Ph.D. degree in control science and engineering from Northeastern University. His research interests include fault diagnosis, industrial big data analysis, and deep learning.

Huaguang Zhang (Fellow, IEEE) received the B.S. and M.S. degrees in control engineering from the Northeast Dianli University of China, Jilin City, China, in 1982 and 1985, respectively, and the Ph.D. degree in thermal power engineering and automation from Southeast University, Nanjing, China, in 1991.

In 1992, he joined the Department of Automatic Control, Northeastern University, Shenyang, China, as a Post-Doctoral Fellow for two years. Since 1994, he has been a Professor and the Head of the Institute of Electric Automation, School of Information Science and Engineering, Northeastern University. He has authored or coauthored more than 280 journal and conference articles and six monographs and co-invented 90 patents. His current research interests include fuzzy control, stochastic system control, neural networks-based control, and nonlinear control and their applications.