Anchor-based Detection and Height Estimation Framework for Particle Defects on Cathodic Copper Plate Surface

Chen Sun, Qian Wan, Zhaofu Li, Liang Gao, *Senior Member, IEEE,* Xinyu Li, *Member, IEEE,* and Yiping Gao, *Member, IEEE*

***Abstract*—Particle defects on the cathodic copper plate surface always happen due to the immaturity of electrolytic copper processing. The removal of defects mainly depends on their height exceeding the plate and current removal requires manual measurement and operation, which is time-consuming and laborious. To automate the removal process, machine vision-based defect detection methods need to be developed. However, copper defects are of very small size, which increases the difficulty of feature extraction and prediction. Therefore, this paper proposes a novel Anchor-based Detection and Height Estimation (ADHE) framework, to locate the defect out and estimate the height of the defect in an end-to-end way. Large-scale raw images are transformed into several image blocks as input. Defect features are obtained by Defect Region Extraction Network and then sent into Height-RCNN for defect detection and height prediction. Dataset of cathodic copper plate surface defects has been collected from a real-world manufacturing factory. Experimental results show that the proposed ADHE method can effectively address the small size problem of copper defects and achieve excellent results in detection and height estimation.**

I. Introduction

The production of the cathodic copper plate is important for the development of the whole society, especially for the manufacturing industry. Due to the problem in current electrolytic processing, when the cathode copper is precipitated, defects will come out on the surface of the copper plate as particles. Those copper particles beyond the national standard requirements (the height exceeding the copper plate) are considered unqualified surface defects. It's necessary to detect these copper particles and remove them. However, due to the undeveloped and unintelligent production conditions in cathodic copper factories, detection and removal of surface defects rely heavily on human labor currently. It's dangerous and time-consuming for human workers to be exposed to an industrial environment to remove defects of copper plates. Working for a long time also tends to make their estimation less accurate. Thus, to meet the requirements of intelligent manufacturing, automatic detection of surface defects on cathodic copper plates needs to be developed.

Machine vision has been widely used in many quality inspections of various products because it is a non-contact and harmless approach. The vision-based defect recognition is stated in [1] and the image-based surface defect detection is systematically studied in [2], where many pieces of research about image or vision-based methods for surface defect recognition and detection have been introduced. Defect detection in images of nanofibrous materials has been studied in [3], and an improved method based on convolutional autoencoders is proposed in [4]. A multiscale GAN [5] method is studied for fabric surface defect detection [6]. Recently, weighted double-low-rank decomposition is applied to fabric defect detection [7]. Multiple Hierarchical Features are fused in an end-to-end way for steel surface defect detection [8]. Pyramid feature fusion and global context attention network is proposed recently for surface defect detection [9]. These recent studies suggest that image-based methods for defect detection have shown effectiveness and efficiency. Inspired by these researches, a novel anchor-based framework is studied for detection and height estimation of particle defects on the surface of cathodic copper plate with captured images in this paper.

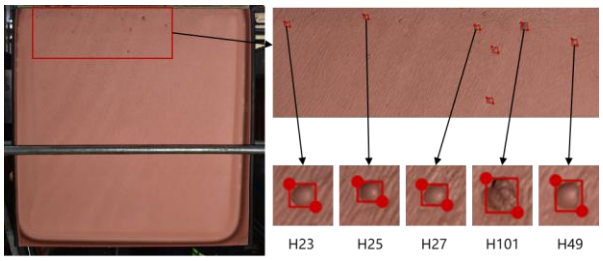

Fig. 1 The example of the defective cathodic copper plate. These circled in red are defects and zoomed-in right, under the followings are the corresponding height of each copper defect. The unit of height is millimeter.

An example of defective cathodic copper plate is shown in Fig. 1. The Difficultly of this dataset is two-folded. On the one hand, it can be seen in Fig. 1 that the sizes of the defects are much smaller than the whole image, which makes it hard to accurately detect them. On the other hand, different from normal defect datasets, defect particles on copper plates are labeled with not only locations and categories, but also their heights, as shown in Fig. 1. So object detection methods like Faster RCNN [10] cannot be directly applied to this data., while conventional image processing algorithms are based on hand-designed features, which is time-consuming and needs much expert experience [1].

To address the two problems in an end-to-end manner, a novel Anchor-based Detection and Height Estimation (ADHE) framework is proposed for cathodic copper plate surface defects. The main novelty and contribution in this article are listed as follows:

1) A novel anchor-based framework is proposed for small particle defect detection on the large cathodic copper plate

Chen Sun, Qian Wan, Zhaofu Li, Yiping Gao, Liang Gao, and Xinyu Li are with the State Key Laboratory of Digital Manufacturing Equipment & Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: sun_chen@hust.edu.cn; wanqian19@hust.edu.cn; lizhaofu0215@hust.edu.cn; gaoyiping@hust.edu.cn; gaoliang@mail.hust.edu.cn; lixinyu@mail.hust.edu.cn, Corresponding author: *Yiping Gao*.)

surface. Especially, a Height-RCNN network is designed to conduct detection and height estimation simultaneously in an end-to-end manner.

2) A dataset of cathodic copper plate surface defects has been collected from a real-world industrial factory. And the extensive experiments on the dataset are conducted in this paper to show the effectiveness of the proposed ADHE method.

The remaining of this article is organized as follows. The related work about surface defect detection and anchor-based object detection is stated in Section II. The proposed ADHE method is introduced in Section III. The experiments are extensively conducted and analyzed in Section IV. Finally, conclusions and future researches are discussed in Section V.

II. Related Work

A. Surface Defect Detection

Surface defect detection is a significant stage before products leave the factory or the current station. And many pieces of research on surface defect recognition and detection have been studied [2], [11], [12]. Object detection-based methods are widely applied for surface defect detection, and the proposed method in this paper can be categorized into this branch.

Image pyramid convolution neural network based on Mask R-CNN [13] has been proposed to detect surface defects [14], which combines image pyramid and deep convolutional neural network to extract pyramid features and performs well on the inspection of the oil leak defect on the freight train. In the recent proposed DefectNet [15], a shared weight binary classification network is applied to determine whether an image contains the defects, and then the detection network is used for detection. The DefectNet shows a good detection speed and effect for surface defect detection. A decoupled two-stage object detection model is proposed for surface defect detection of flexible printed circuit boards [16], in which a multi-hierarchical aggregation module is used for a location feature enhancement module in the defect localization and a locally non-local module is further used to enhance the defect classification, achieving an accurate surface defect detection. To address the surface defect detection of the printed circuit board, an extended feature pyramid network model is proposed to accurately locate and identify small defects and shows well performance [17].

Many excellent pieces of research on surface defect detection of various objects have been reported recently, while an end-to-end height estimation of defects is still required in the work of this paper.

B. Anchor-based Objection Detection

Object detection has attracted a lot of attention these years [18]. And anchor-based objection detection is widely applied in various object detection scenarios due to its outstanding detection precision.

The classical Fast Region-based Convolutional Network (Fast R-CNN) is proposed for object detection in an anchor-based approach with two stages [19]. Faster R-CNN [10] based on Fast R-CNN is then improved with an end-to-end trained Region Proposal Network (RPN) for object detection, achieving a faster detection speed. One-stage anchor-based SSD model [20] spreads out anchor boxes on multi-scale layers to directly predict anchor box offsets and object categories. Recently, RetinaNet [21] is further proposed for dense object detection with focal loss, which is trained in a one-stage way and achieves high detection accuracy and detection speed.

Anchor-based objection detection methods have shown high detection accuracy. Therefore, the proposed ADHE method in this paper also follows the main framework and re-design RCNN module for extra height estimation of defects

III. Proposed ADHE Method

In this paper, an Anchor-based Defect Detection and Height Estimate network (ADHE) is proposed which consists of Image Block, Defect Region Extraction, and Height-RCNN (H-RCNN) three modules. Firstly, large-scale raw images are cropped into several input blocks without overlapping. Then input blocks are sent into the defect region extraction network, which follows a two-stage network structure used in [10] as an overall framework, including feature extraction network, Region Proposals Network (RPN), and Region of Interest (ROI) pooling. A feature extraction network is designed to learn abstract representations from input blocks and generate feature maps of different scales. And then RPN network is applied to produce candidate proposal regions that are likely to contain defects with pre-defined anchor boxes. ROI layer is applied to pool candidate proposal regions into fixed-size vectors with fixed dimensions. Finally, the Height-RCNN

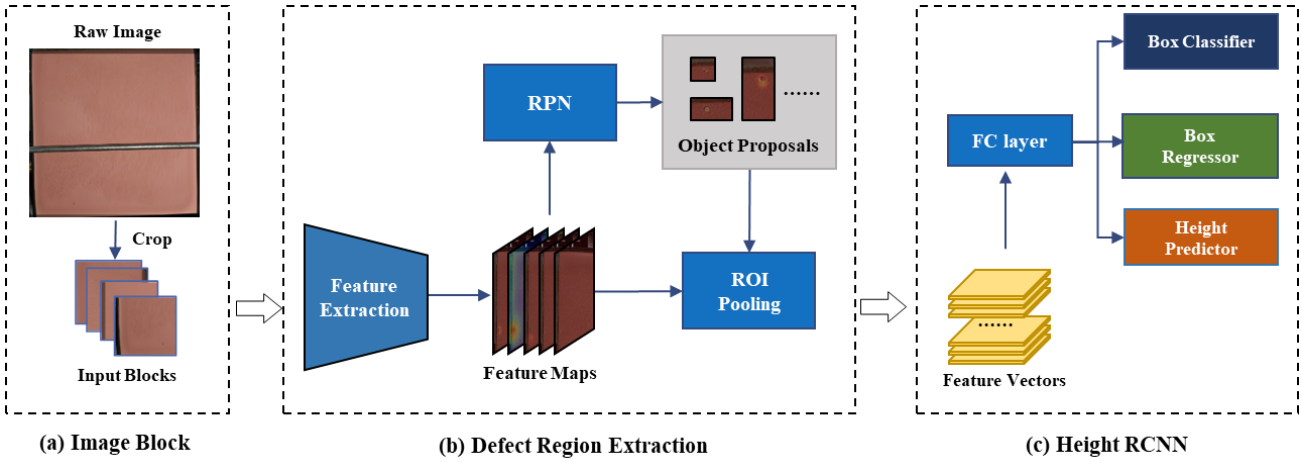

Fig. 2 The framework of the proposed ADHE method.

module, which is an RCNN model with an extra height predictor branch, takes fix-size vectors as input and conducts defect classification, location, and height estimation tasks simultaneously. The framework of the proposed ADHE is presented in Fig. 2. All the modules are trained in an end-to-end approach.

A. Image Block

As shown in Fig. 1, defects on the surface of the cathodic copper plate are much smaller than on the whole plate. The dramatic contrast in size between the foreground defect object and background plate makes it easier for the network to neglect defect features, causing a lower recall rate. Large-scale images cannot be directly fed into detection networks due to hardware limitations, so an image block operation is firstly adopted to mitigate these issues. Raw images are cropped into $n \times m$ blocks, n stands for the number of blocks by row and m stands for the number of blocks by column, and all these blocks share no overlapping areas between each other. This operation does not increase any parameters and adjusts the receptive field for the detection of small defects. In Fig. 3, a 2×2 block operation is shown. More, the operation can increase the number of the training set and is conducive to obtain a better detector. In the section on experiments, the improved performance of image blocking can be figured out.

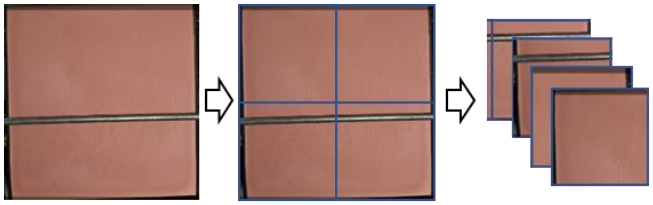

Fig. 3 Image blocking used in this paper.

B. Defect Region Extract Network

To accurately locate the defects and compute defect features in image blocks, an anchor-based method is adopted like in [10], including feature extraction network, RPN network, and ROI Pooling.

First of all, a feature extraction network consisting of a backbone and neck network is used to get multi-scale feature maps, which contains high-resolution feature maps for small features and low resolution but more abstract semantic feature map for large features. ResNet with Feature Pyramid Network (FPN) is a common combination. But for defects on cathodic copper plates, HRNet [22] is a better choice for defects of small size. After feature extraction, a region proposals network is applied to generate candidate proposals. The label of anchors is assigned like as in [10], a positive label is that the anchor with the highest inter-section-over-union (IoU) overlaps with a ground-truth box, or an anchor that has an IoU overlap higher than 0.3 with any ground-truth box. And the training loss of RPN L_{RPN} is defined as follows.

$$\mathrm{L_{RPN}} = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda_1 \frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

Where L_{cls} is cross-entropy loss over two classes (defect or background), and p_i^* denotes the true label of the anchor; L_{reg} is smooth-L1 loss over position regression, t_i^* denotes the transformed coordinates of the bounding box of ground-truth, N_{cls}, N_{reg} respectively denote the normalized term of classification, regression. And λ_1 is the weighted term.

Candidate Proposals, together with feature maps gained from the feature extraction network are then sent into the ROI pooling layer to extract fixed-size feature vectors for the following Height-RCNN model.

C. Height RCNN

The classical RCNN model in [10] takes fixed-size feature vectors from the ROI layer as inputs and generates class labels and object locations as output. However, for particle defects on cathodic copper plates, not only class and location, but also the height of defects need to be predicted. Based on this, a Height RCNN(H-RCNN) model is proposed. As shown in Fig. 2, H-RCNN consists of a shared Fully Connected (FC) layer and three separate branches. Feature vectors processed by ROI layers are input into shared FC layers for channel-wise reduction. A box classifier branch is used to classify whether the region is a defect, a box regressor branch is responsible for outputting the defect location, and a height predictor branch is proposed to estimate the height of the defect to help defect removal in this paper

The loss of the H-RCNN model is defined as follows:

$$\mathrm{L_{H-RCNN}} = \frac{1}{N_{bcls}}\sum_i L_{bcls}(u_i, u_i^*) + \lambda_2 \frac{1}{N_{breg}}\sum_i u_i^* L_{breg}(t_i, t_i^*) + \lambda_3 \mathrm{L_{Height}} \quad (2)$$

$$\mathrm{L_{Height}} = \frac{1}{N_{height}}\sum_i u_i^* L_{reg}(d_i, d_i^*) \quad (3)$$

where u_i^* denotes the true label of ground-truth, L_{bcls} is also log loss over two classes (defect or background), and L_{breg} is also smooth-L1 loss over position regression. N_{bcls}, N_{breg} respectively denote the normalized term of classification, regression. λ_2 and λ_3 are the weighted terms; L_{height} is also smooth-L1 loss [19] over height estimation. H_{height} denotes the normalized term of height estimation.

All the parameters of the whole ADHE model are trained in an end-to-end way, and the total training loss is defined as follows:

$$\mathrm{L} = \mathrm{L_{RPN}} + \mathrm{L_{H-RCNN}} \quad (4)$$

At the end of the training, the ADHE model can be applied for testing.

IV. Experiment Results and Discussion

In this section, the cathodic copper plates defect dataset is introduced, on which the experiment is conducted and analyzed. And the influence of the backbone network, image block numbers, and parameter λ_3 is respectively studied.

A. Experimental Details and Settings

Cathodic Copper Plates Defect Dataset (CCPD) consists of 160 images collected from the real-world industrial copper manufacturing factory. And, the defects on each image are labeled with class, coordinates of the bounding box, and height, as shown in Fig. 1. The resolution of each image is 2516×2428 while the size of defects varies from 147.45 to 6911.71. And the height of defects varies from 2 to 300, as shown in Fig. 5. There are 333 ground-truth bounding boxes in total. The training set and testing set are randomly divided into 7:3.

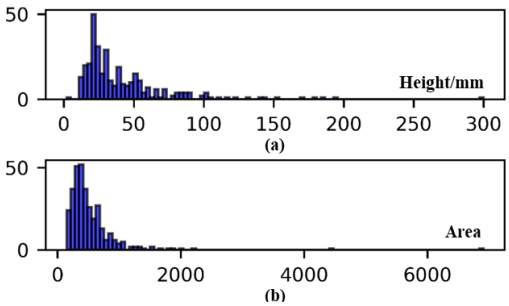

Fig. 5 Information collected from Cathodic Copper Plates Defect Dataset (CCPD). (a) distribution of height labels; (b) distribution of area for bounding boxes

The code of the ADHE method is implemented with PyTorch1.10 and MMDetection toolbox [23]. All the experiments are performed on NVIDIA RTX 3090 GPU (with 24G memory) under CUDA11.1 on Ubuntu 16.04 system. The size of the input image is 800×800. All the backbone networks used in experiments are pre-trained on the ImageNet classification task. And The default setting of the backbone network is HRNet-w40 [22]. For fine-tuning on copper defect dataset, the batch size is set as 4, the number of total training epochs is set as 20, and the learning rate is set as 0.02 with a decay of 0.0001. Others follow default settings in the MMDetection toolbox. For loss functions, the settings of parameters λ_1, λ_2 follow the ones in [10]. The default setting of the parameter λ_3 is 0.1. And the default setting of the image block is 3×3.

Evaluation metric: The effect of detection is evaluated by Average Precision (AP), a comprehensive metric used in object detection. AP calculates areas under the P-R curve, which means both Precision and Recall are evaluated. Since height estimation can be viewed as regression generally, Root Mean Squared Error (RMSE) is used to evaluate the effect of height predictor in H-RCNN, which calculates differences between ground truth height values and predicted ones.

The aforesaid metrics are defined as follows.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{6}$$

$$\text{AP} = \int_0^1 P(R)dR \tag{7}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{8}$$

where TP and FN represent the number of copper defect areas that are correctly detected or not, and FP represents the number of background areas that are misclassified as defect. $\hat{y}_i$ stands for ground-truth value for the height of defects and y_i is predicted one.

Since there are three parts to the ADHE framework, each of them contributes to detection and height estimation tasks differently. The following 3 sections study the influence of the feature extraction network, image block number, and hyperparameter λ_3 separately.

B. The influence of Feature Extraction Network

In this section, the effect of the feature extraction network is discussed. A feature extraction network is vital for learning from small defect objects. In this experiment, 4 different networks, including ResNet50, ResNet101 [24], HRNet-w32, and HRNet-w40 four are studied and compared.

TABLE I. THE EXPERIMENTS OF ADHE WITH DIFFERENT FEATURE EXTRACTION NETWORK ON CCPD DATASE

Backbone	Params	P↑	R↑	AP_{50}↑	RMSE↓
ResNet-50	26.63M	0.372	1.000	0.949	9.135
ResNet-101	45.62M	0.383	1.000	0.955	9.640
HRNet-w32	32.38M	0.379	1.000	0.951	10.016
HRNet-w40	48.63M	**0.393**	**1.000**	**0.956**	**8.572**

"↑" means higher results are better and "↓" means lower results are better

TABLE I. lists AP_{50}, Precision, Recall, RMSE results as well as parameter counts of 4 networks. According to AP_{50}

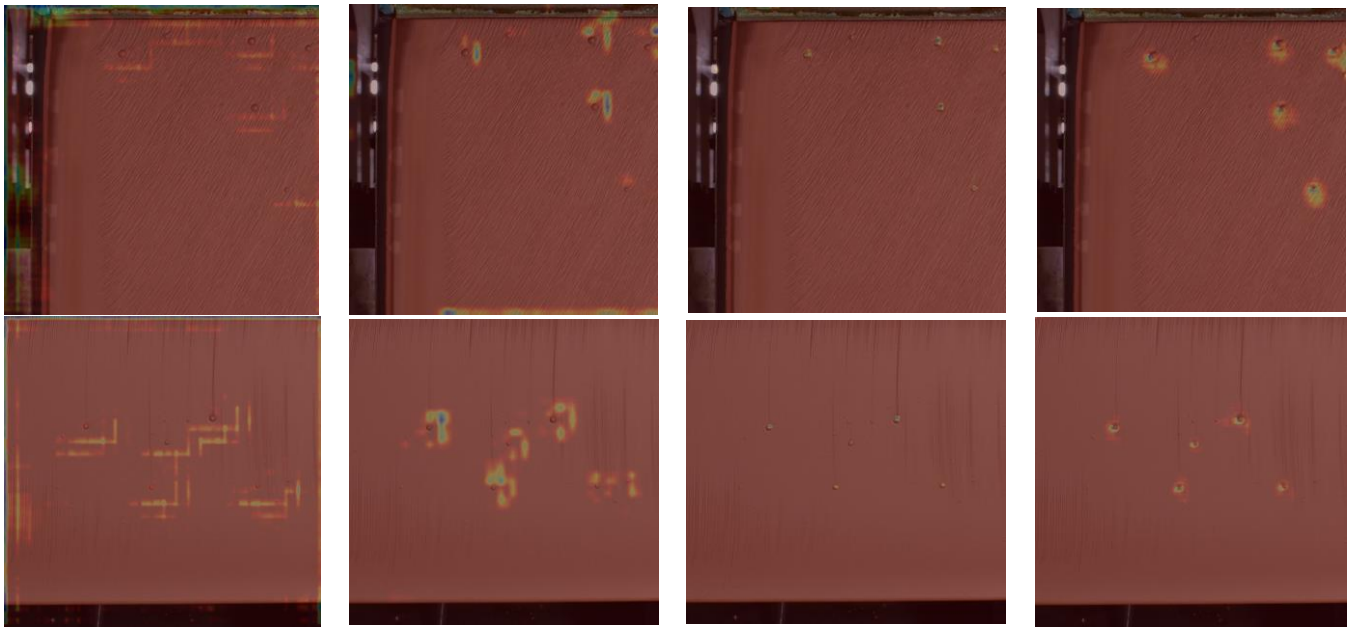
(a) (b) (c) (d)

Fig. 4 Feature Map from ADHE feature extraction network with 4 different backbones (a)ResNet-50 (b)ResNet-101 (c)HRNet-w32 (d)HRNet-w40

results, the detection performance gets better when models get bigger and ADHE with HRNet-w40 achieves the highest average precision value. It shows that the HRNet-w40 backbone is in favor of detecting defects of small size. And based on the well-learned features from the HRNet-w40 backbone, the height estimation also can be well achieved. So, HRNet-w40 also achieves the best result on the height estimation task, which is 1.444 lower than HRNet-w32,1.086 lower than ResNet-101, and 0.563 lower than ResNet-50.

Fig. 4 shows features maps of the ADHE method with 4 different backbones listed in TABLE I. Although cropped, the size of the images is still much bigger than the defects. As shown in Fig. 4, feature maps from ADHE methods can capture the information of defects roughly. Compared between 4 backbones, HRNet-based methods in Fig. 4(c)(d) is generally better than ResNet-based methods in Fig. 4(a)(b). HRNet-based methods locate the position of defects more accurately and less active to useless backgrounds. Moreover, HRNet-w40 learns abundant information not only from defect regions but also from surrounding areas in Fig. 4(d), while HRNet-w32 in Fig. 4(c) only focuses on little areas of the defect region. This difference between HRNet-w40 and HRNet-w32 makes their performance gap on the RMSE metric.

C. *The Influence of Image Block Numbers*

The operation of image blocking helps detection of defects of small size and carries no additional parameters. The smaller size of image blocks, the better the ADHE method can locate defects. In this paper, the influence of numbers of image blocks has been studied and the results are shown in TABLE III. . Default feature extraction network is HRNet-w40.

As shown in TABLE II. , raw images are cropped into 3 different patches. It can be seen that when using the same feature extraction network, the detection and height estimation results increase as the block number adds. And 3x3 of image blocking achieves the best detection result on AP_{50} and height estimation result on RMSE.

TABLE II. THE INFLUENCE OF DIFFERENT RATES OF IMAGE BLOCKING ON DEFECT DETECTION AND HEIGHT ESTIMATION.

Block	P↑	R↑	AP_{50}↑	RMSE↓
1x1	0.438	0.767	0.715	13.127
2x2	0.392	0.971	0.927	12.314
3x3	**0.393**	**1.000**	**0.956**	**8.572**

Since $ADHE_{3x3}$ method with HRNet-w40 backbone achieves the best results on AP_{50} and Height RMSE metric, some examples of defect height estimation visualizations of defect detection are given in TABLE III. and Fig. 6 separately.

TABLE III. THE HEIGHT ESTIMATION OF DEFECTS OUTPUTTED BY THE PROPOSED ADDE METHOD. (/MM)

No.	**1**	**2**	**3**	**4**	**5**	**6**
Ground Truth	26.00	19.00	18.00	38.00	72.00	82.00
Estimation	29.94	23.69	19.60	37.68	72.36	84.67
No.	**7**	**8**	**9**	**10**	**11**	**12**
Ground Truth	50.00	22.00	27.00	39.00	65.00	21.00
Estimation	52.69	27.21	28.47	37.55	64.71	18.98

12 height estimation results from $ADDE_{3x3}$ with their ground truth values are listed in TABLE III. It shows that the height estimator in the ADDE method also achieves good results which are close to ground truth height.

More, some visualization results of defect location are shown in TABLE III. Ground truth bounding boxes are painted in green as a whole area while predicted bounding boxes are drawn as red boxes. As shown in Fig. 6proposed ADHE can achieve a good detection performance while taking an extra defect height estimation task. All the ground truth defects shown in Fig. 6 have been found, even including those defects which can be neglected.

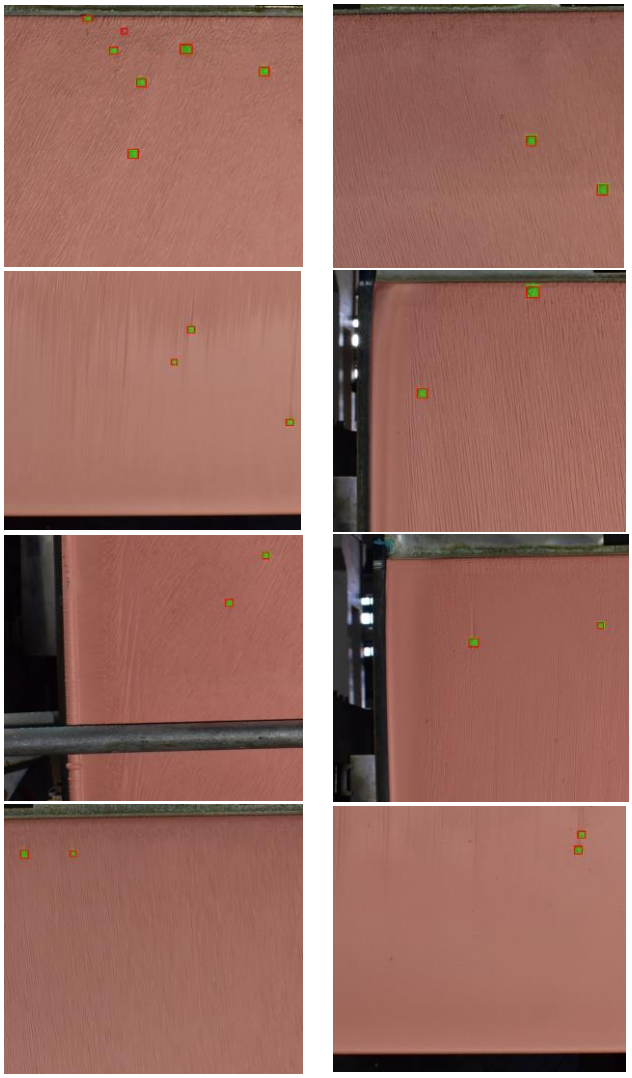

Fig. 6 The visualizations of defect detection results on real-world cathodic copper plates with the proposed ADHE method. The backbone of ADHE is HRNet-w40 and rate of image blocking is 3x3.

D. *The Influence of Hyperparameter* λ_3

The parameter λ_3 is the weighted term to balance the training loss and adjust the weight of loss L_{Height}. In this paper, the influence of parameter λ_3 on defect detection and height estimation is also experimentally studied and analyzed. The smaller the parameter λ_3, the less contribution of the loss L_{Height} is given during training. But the larger the parameter λ_3,

it can affect the training of the detection. Thus, making a balance during training is much important for the proposed ADHE method.

As shown in TABLE IV. , the value of the parameter λ_3 is set from 0.05 to 5.00 to study the influence. According to AP_{50} and RMSE results, the value of 0.5 of the parameter λ_3 achieves the best performance on defect detection while the value of 0.1 height estimation. Since the best result on defect detection and height estimation cannot be achieved at the same time, a trade-off needs to be taken into consideration. The default setting of the parameter λ_3 is 0.1 because it also balances the training of both tasks and obtains a relatively better performance.

TABLE IV. THE INFLUENCE OF PARAMETER λ_3 ON DEFECT DETECTION AND HEIGHT ESTIMATION.

λ_3	P↑	R↑	AP_{50}↑	RMSE↓
0.05	**0.395**	**1.000**	0.944	9.134
0.10	0.393	1.000	0.956	**8.572**
0.50	0.350	1.000	**0.962**	9.610
1.00	0.244	1.000	0.960	10.593
5.00	0.037	0.547	0.314	40.079

V. CONCLUSION AND FUTURE WORK

In this paper, a novel Anchor-based Detection and Height Estimation (ADHE) framework is proposed to detect particle defects on cathodic copper plate surfaces. The proposed method utilizes Image Block operation to crop large-scale raw images into several image blocks. As to small size defects, a defect region network and Height-RCNN network are deployed to conduct feature extraction and detection task. Especially, the proposed ADHE method can locate the defect out and estimates the height of the defect in an end-to-end way. The experiments on the cathodic copper plate defect dataset show the effectivity of the proposed ADHE method.

Though the operation of image blocking can help achieve better performance, the rate of image blocking requires human setting. It is significant to develop an adaptive image blocking method for defect detection and height estimation. And it shows different background and light illumination from images of cathodic copper plates, thus, one of the research directions is to develop a domain adaptation ADHE method. Additionally, the automatic equipment for defect removal of the cathodic copper plate should be further researched to help validate the proposed ADHE method on more real-world images.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 52188102 and the National Key R&D Program of China under Grant 2018AAA0101700.

REFERENCES

[1] Y. Gao, X. Li, X. V. Wang, L. Wang, and L. Gao, "A Review on Recent Advances in Vision-based Defect Recognition towards Industrial Intelligence," *J. Manuf. Syst.*, vol. 62, pp. 753–766, Jan. 2022.

[2] P. M. Bhatt *et al.*, "Image-based surface defect detection using deep learning: A review," *J. Comput. Inf. Sci. Eng.*, vol. 21, no. 4, 2021

[3] D. Carrera, F. Manganini, G. Boracchi, and E. Lanzarone, "Defect detection in SEM images of nanofibrous materials," *IEEE Trans. Ind. Inform.*, vol. 13, no. 2, pp. 551–561, 2016.

[4] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," *ArXiv Prepr. ArXiv180702011*, 2018.

[5] I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[6] J. Liu, C. Wang, H. Su, B. Du, and D. Tao, "Multistage GAN for fabric defect detection," *IEEE Trans. Image Process.*, vol. 29, pp. 3388–3400, 2019.

[7] D. Mo, W. K. Wong, Z. Lai, and J. Zhou, "Weighted double-low-rank decomposition with application to fabric defect detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 1170–1190, 2020.

[8] 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.

[9] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, "PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection," *IEEE Trans. Ind. Inform.*, vol. 16, no. 12, pp. 7448–7458, 2019.

[10] 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, 2017.

[11] Y. Gao, L. Gao, and X. Li, "A Generative Adversarial Network Based Deep Learning Method for Low-Quality Defect Image Reconstruction and Recognition," *IEEE Trans. Ind. Inform.*, vol. 17, no. 5, pp. 3231–3240, 2021.

[12] Y. Gao, L. Gao, X. Li, and X. V. Wang, "A multilevel information fusion-based deep learning method for vision-based defect recognition," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 3980–3991, 2019.

[13] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017.

[14] L. Xiao, B. Wu, and Y. Hu, "Surface defect detection using image pyramid," *IEEE Sens. J.*, vol. 20, no. 13, pp. 7181–7188, 2020.

[15] F. Li and Q. Xi, "DefectNet: toward fast and effective defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[16] J. Luo, Z. Yang, S. Li, and Y. Wu, "FPCB surface defect detection: A decoupled two-stage object detection framework," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[17] C. Li, Z. Qu, S. Wang, K. Bao, and S. Wang, "A Method of Defect Detection for Focal Hard Samples PCB Based on Extended FPN Model," *IEEE Trans. Compon. Packag. Manuf. Technol.*, 2021.

[18] L. Liu *et al.*, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020.

[19] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[20] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *European conference on computer vision*, 2016, pp. 21–37.

[21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[22] J. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, 2020.

[23] K. Chen *et al.*, "MMDetection: Open mmlab detection toolbox and benchmark," *ArXiv Prepr. ArXiv190607155*, 2019.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.