IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 February 2024, accepted 4 March 2024, date of publication 5 March 2024, date of current version 13 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374131

RESEARCH ARTICLE

A Method for Surface Defect Detection Based on Multiscale Feature Fusion and Pyramid Attention

YING TANG, HONGYUAN WANG, QUNYING ZHOU, AND BOYAN SUN
School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213000, China

Corresponding author: Hongyuan Wang (hywang@cczu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976028, and in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX22_3068.

ABSTRACT The two-stage defect detection model needs to pay attention to the results of the segmentation network and the classification network, and the results of the segmentation network will have an impact on the classification network. Previous models ignored shallow features in the segmentation network and used relatively simple classification networks that could not make good use of the features of the segmentation network. This paper proposes a surface defect detection algorithm based on multi-scale feature fusion and pyramid attention(MFFPA). First, a multi-scale feature fusion module is added to the segmentation network to fuse shallow features and extract more comprehensive feature information; then a pyramid attention module is added to the classification network to increase the receptive field of the model and enhance the discriminative ability of the model. The method proposed in this article was verified on four datasets, and the experimental results show that the added module can effectively improve the accuracy of the model.

INDEX TERMS Channel attention, convolutional neural networks, defect detection, multi scale feature fusion.

I. INTRODUCTION

Product defect detection is an indispensable process in industrial production, during production monitoring, may occur with the degraded images, some recent image processing methods [1], [2], [3], [4], [5] are considered as the pre-processing steps to handle them. In addition, previous defect detection required manual screening, which was costly and inefficient, making it difficult to cover large-scale quality inspection needs. In recent years, with the continuous development of computer vision technology, algorithms based on machine learning and deep learning have begun to be applied in the field of industrial defect detection [6].

As shown in Fig.1, according to different data labels, deep learning models in defect detection can be divided into fully supervised learning models, unsupervised learning models, hybrid supervised learning models, and weakly supervised learning models [7]. The defect samples used in fully supervised learning model training all have pixel-level annotations [8], [9]. Unsupervised learning models only use defect-free samples for training, but the accuracy of the model is lower compared to fully supervised learning models [10], [11]. Weakly supervised learning models use image-level labeled data for classification or segmentation, which can effectively utilize the data to improve the accuracy of the model [12], [13]. Both unsupervised learning models and weakly supervised learning models reduce the cost of data labeling, but the model accuracy is obviously insufficient compared with fully supervised learning models. Therefore, in the field of defect detection, some researchers have begun to use hybrid supervised learning methods, This method adds a small amount of pixel level sample data on the basis of weakly supervised learning, effectively improving the accuracy of the model [14].Compared with weakly supervised learning, hybrid supervised learning is more flexible and can achieve better results by labeling a small amount of data at the pixel level.

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma.

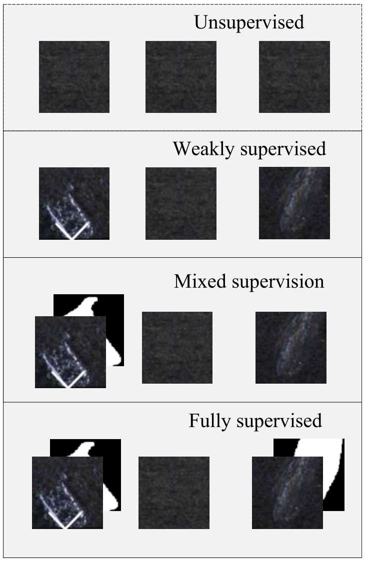

FIGURE 1. As shown above, data labelling in defect detection has been classified into four cases.

Previous hybrid supervised models composed an overall model by building associated sub-models, which played a guiding and strengthening role between different tasks. For example, the MixSup model proposed by Jakob et al. [15], the prediction map generated by the segmentation network of this model is concatenated with the feature map of the classification network through pooling layers, providing guidance for the final classification result. However, this model ignores the shallow features of the segmentation network and uses a relatively simple classification network. The state-of-the-art methods is MaMiNet proposed by Luo et al. [16], This method achieved better results by adding external attention, but also increased the inference time of the model.

In response to the problems mentioned above, this paper proposes a surface detection model based on multi-scale feature fusion and pyramid attention based on the MixSup model. This model can effectively enhance the feature extraction capability of the model and greatly improve the classification accuracy of the model. And by utilizing shallow features and reducing the number of channels for deep features, the computational complexity of the model is reduced. The main contributions of this paper are as follows:

1) This paper proposes a multi-scale feature fusion module with local sensing ability, this module can effectively fuse the shallow features of the model and improve the feature extraction capability of the model.

2) This article proposes an improved pyramid attention module, which allows the model to obtain multi-scale information, focus on more important channel features.

3) The model proposed in this article had a faster inference time than the previous best method, and the performance of the model is also competitive.

II. RELATED WORK

A. DEFECT DETECTION

As early as 2012, Masci et al. have used convolutional neural network to classify defects in steel [17]. But Masci et al. used a shallow network and later in 2017, Kim et al. used a deeper convolutional neural network, VGG16, for defect detection [18]. In 2018 Wang et al. used a convolutional neural network based on a classification approach to achieve high accuracy in cloth defect detection [19]. In 2019, Liu et al. used a lightweight MobileNet-SSD network for defect detection and achieved faster detection speed [20]. In 2020, Huang et al. [21] introduced multi-scale features using multiple parallel null convolutional layers.

Since fully supervised learning requires a large amount of labeled data, some researchers began to use Few-shot learning for defect detection. In 2023, Bao et al. proposed Triplet-Graph Reasoning Network (TGRNet) [22], achieved universal defect detection of metals with few samples. Feng et al. [23] used space-squeeze attention (SSA) module to aggregate multiscale context information of defect features. Xie et al. proposed a new Few-Shot Anomaly Detection method called GraphCore [24], which uses a small amount of normal samples to achieve fast training of new products and competitive accuracy performance.

Unsupervised learning does not require defective samples for training, and is also favored by many researchers. In 2021 Marco et al. used a normalised streaming approach on the MVTEC dataset to achieve the best results for unsupervised anomaly detection [25]. In 2024, Batzner et al. constructed a lightweight teacher-student model [26],achieved detection speed of 2ms. Hyun et al. employs contrastive representation learning to collect and distribute features in a way that produces a target-oriented and easily separable representation.This article uses a hybrid supervised learning method to reduce the data annotation cost caused by full supervision. Using only a small amount of pixel-level annotation data can greatly improve the AP of the model.

B. ATTENTION MECHANISM

In 2015, Xu et al. [27] proposed a visual attention theory, which introduced the attention mechanism into the field of computer vision for the first time. Later, Hu et al. [28] proposed a Squeeze-and-excitation networks(SE) to calculate the weight of each channel, and Hu et al. [29] used spatial attention to assign weights to the pixel points of each feature map. Inspired by these studies, a series of studies such as CBAM [30], SCSE [31], and CoordAttention [32] fused channel attention with spatial attention to achieve better results. The above models have been simplified in some studies, Gcnet [33] proposed a simpler spatial attention module, and ECA-Net [34]introduced one-dimensional convolution to reduce the number of parameters of the model.In order to effectively obtain and utilize the spatial information of feature maps at different scales, Zhang et al. proposed an efficient pyramid squeeze attention net(EPSA) [35].

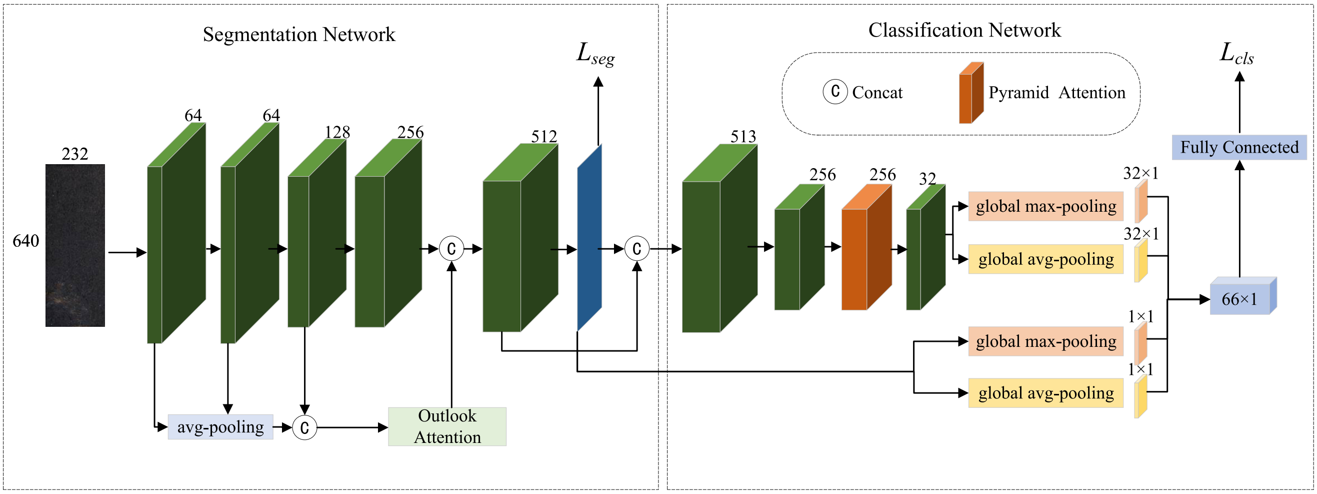

FIGURE 2. Network structure diagram.

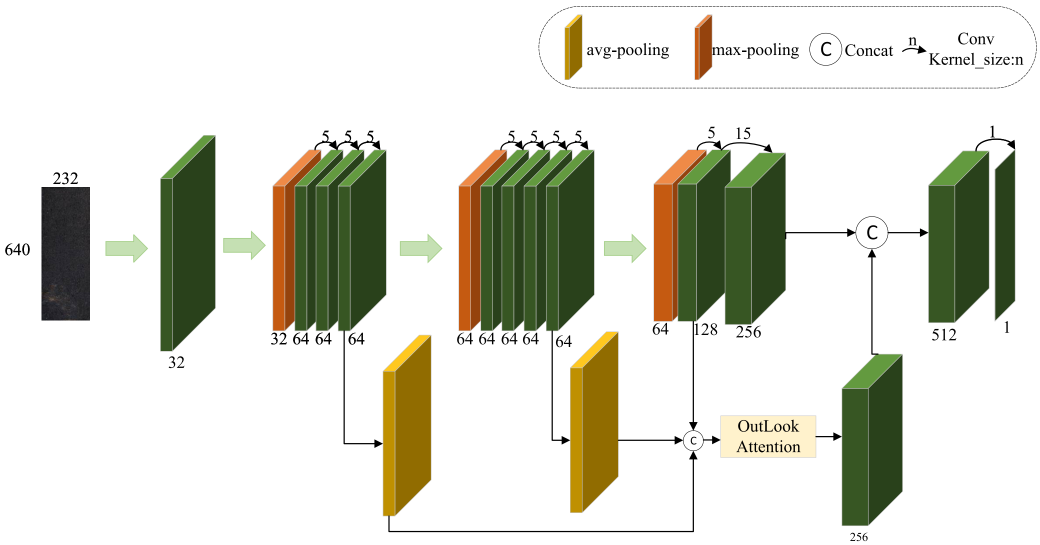

FIGURE 3. Segmentation network structure diagram.

C. MULTI-SCALE FEATURE FUSION

Multi-scale feature fusion is a common target detection technology. Its main function is to integrate features of different depths and different levels in order to better utilize multi-scale features to reduce the semantic gaps between different layers.In 2017, Lin et al. proposed the classic feature pyramid network(FPN) [36],which fusion feature from deep layer to shallow layer.In 2020, Tan et al. proposed BiFPN [37],which fusion feature bidirectionally.Then in 2023, Quan et al. proposed a Centralized Feature Pyramid module [38]to optimize global information,make full use of the same scale of information.Wang et al. proposed a gather-and-distribute module [39], which use different fusion methods for low-stage features and high-stage features.

III. METHOD

As shown in Fig.2, the defect detection model proposed in this paper consists of segmentation network and classification network. After global average pooling and global maximum pooling, the features of the final output of the segmentation network are spliced with the final output of the classification network, which plays a guiding role in the final results of the model.

A. SEGMENTATION NETWORK

The structure diagram of the segmentation network is shown in Fig 3. The input image passed through three stages, and each stage is composed of a 2×2 maximum pool layer and several 5×5 convolution layers. Select the feature map of the last layer of each stage to obtain the feature map of 64 channels, 64 channels and 128 channels respectively, and then use the average pool to sample the feature map down. By concatenating the downsampled feature maps, 256 channel feature maps are obtained and further fed into the outlookattention module [40],obtain the relationship between feature points and surrounding feature points, and further extract local features. The 128 channel feature map output from the last stage is convoluted by 15 × 15 to obtain 256 feature maps. The large convolution kernel can effectively increase the receptive field of the model and bring better segmentation effect. At the end of dividing the network, the feature map output by the fusion module is Concatenated with the feature map output by 15 × 15 convolution, and a single channel feature map is obtained by convolution. The feature map is used to calculate the segmentation loss and the final classification loss.

B. OUTLOOKATTENTION

In order to enhance the local perception ability of the model, outlookattention module is added to the multi-scale feature fusion module.As shown in Fig 5.outlookattention module is divided into two branches. The branch at the top of the picture is the weight production module. The feature map generates weights as shown in equation 1. X denoted the input feature map. At the down of the pictureis,the local window features of the input feature map are obtained by linear layer and unfold operations as shown in equation 2. Finally, as shown in equation 3 the weight is multiplied and accumulated with the feature after Softmax operation to get the final output.

$$A = \mathrm{Reshape}(\mathrm{fc}(x)) \tag{1}$$

$$V = \mathrm{fc}(x)$$

$$V_{\Delta i,j} = \left\{ V_{i+m-\frac{K}{2}}, V_{j+n-\frac{K}{2}} \right\}, 0 \leq m, n < K \tag{2}$$

$$Y = \sum_{0 \leq m,n<K} matmul \left(Softmax(A), V_{\Delta i,j_{i+m-\frac{K}{2}},j+n-\frac{K}{2}} \right) \tag{3}$$

C. CLASSIFICATION NETWORK

The structure diagram of the classification network is shown in Fig 5. The classification network first sends the feature map of 513 channels from the segmentation network to max pooling layer and convolutional layer, reduced the size of the feature maps and the number of channels. Then, the multiscale information of the feature map is obtained by convolution of different kernel sizes. After, the feature maps are concatenated together to calculate the attention weight, and the features of different scales are weighted after softmax operation. Finally, the feature map is reduced to 32 channels through convolution operation. The feature map of 32 channels is used to calculate the final classification loss through global average pooling and global maximum pooling.

D. COORDINATE ATTENTION

Different from the SE Weight module used in EPSA, this paper used Coordinate Attention to calculate attention weights. AS shown in Fig 4, the input features are pooled in two directions, which can encode the spatial information into the attention map. Then, similar to SE Weight module, the attention weight matrix is obtained by convolution. Finally, the weights of the two directions are obtained by splitting operation.

E. LOSS FUNCTIONS AND EVALUATION INDICATORS

The loss function used in this paper is shown in equation 4:

$$L = \lambda * \gamma * L_{seg} + (1 - \lambda) * \theta * L_{cls} \tag{4}$$

where L_{seg} denotes the loss of segmentation network and L_{cls} denotes the loss of classification network. λ is the weight of the balancing factor responsible for balancing the losses of the two networks, γ is an indicator of the presence or absence of pixel-level labeling, and θ is an additional classification loss weight.

In industrial production quality control, products are categorized into defective and non-defective, and the classification result of the image determines whether the product is discarded or not. Therefore, in all experiments, this paper focuses on the classification result of each image. Based on the above considerations, this paper uses AP and AUC as evaluation metrics.The AP evaluation metric is averaged over the Precision corresponding to each threshold, calculated as shown in equation 5:

$$AP = \int_0^1 p(r)\, dr \tag{5}$$

where $p(r)$Indicates the accuracy of the model, which is calculated as shown in equation 6.TP denotes the number of samples that are correctly classified as positive examples, and FP denotes the number of samples that are incorrectly classified as positive examples.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

The AUC value is the area under the ROC curve, when different thresholds are taken, multiple sets of coordinates are obtained,the coordinates are calculated as shown in equation 7, TN indicates the number of samples that are correctly classified as negative cases, and FN indicates the number of samples that are incorrectly classified as negative cases. This evaluation index can effectively see the ability of the model to recognize positive samples.

$$x : FP/(FP + TN)$$

$$y : TP/(TP + FN) \tag{7}$$

Also for further comparison, this paper adds the model's inference time (FPS) as an evaluation index to assess the detection speed by the time of inference of one picture, which is calculated as shown in equation 8, Where T is the time for the model to inference a picture.

$$FPS = \frac{100}{\sum_{100}^{1} T} \tag{8}$$

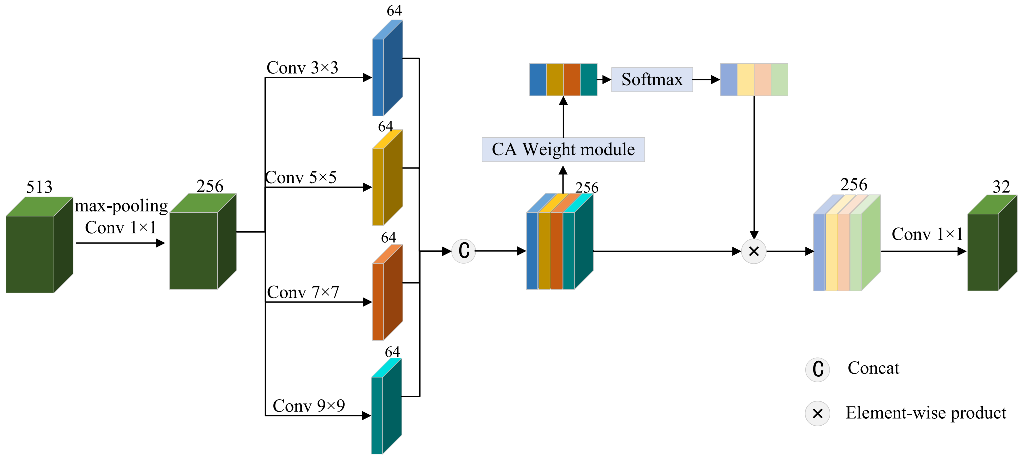

FIGURE 4. Classification network structure diagram.

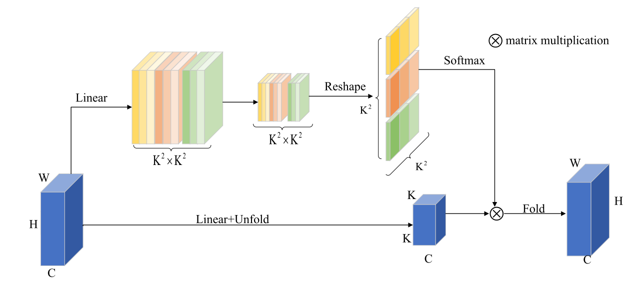

FIGURE 5. OutLookAttention structure diagram.

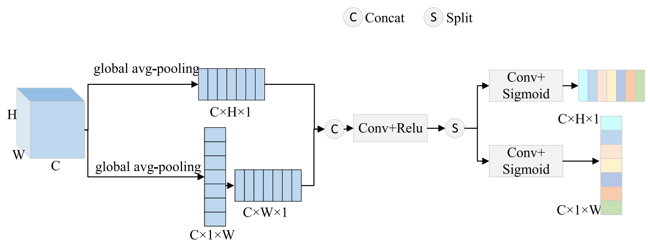

FIGURE 6. Coordinate Attention structure diagram.

IV. EXPERIMENTS

A. DATASET

The experiments in this paper use four datasets that are currently dominant in defect detection: the KolektorSDD (KSDD) dataset [41], the DAGM dataset [42], the KolektorSDD2 dataset (KSDD2) [15] and the Severstal Steel defect dataset (STEEL) [43].

The KSDD dataset was provided by the Kolektor Group doo defect production program. It contains a total of 399 images, 52 of which have visible defects and the remaining 347 images are normal images, each of which has a size of approximately 500*1240 pixels.

The DAGM dataset was provided by the International Pattern Recognition Association. A total of 3450 images are included, and the size of each image is 1600*256 pixels.

The KSDD2 dataset is provided by the Kolektor Group doo defective production program. A total of 3335 images are included, of which 356 images have visible defects and the remaining 2979 images are normal images, each of which has a size of approximately 230*640 pixels. The training set consists of 246 images with defects and 2085 images without defects and the test set consists of 110 images with defects and 894 images without defects;

The STEEL dataset is a steel surface defects dataset provided by Severstal, there are a total of 18074 grayscale images with 4 classifications, and the image size is 1600*256 pixels. There are a total of 12568 images in the training set, containing 7095 defective images and 5473 normal images. Only a subset of this dataset is used in this paper.

B. EXPERIMENTAL SETUP

The experimental setup of this paper is as follows:

(1) Regarding the number of pixel-level annotations N in the KSDD dataset, the settings in this paper are [0, 5, 10, 15, 20, 33]. the Batchsize size is 1, the learning rate is initialized to 0.01, and the number of iterations is 50 epochs;

(2) Regarding the number of pixel-level annotations N in the DAGM dataset, the setting in this paper is [0, 5, 15, 45, 1000]. the Batchsize size is 1, the learning rate is initialized to 0.05, and the number of iterations is 70 epochs;

(3) Regarding the number of pixel-level annotations N in the KSDD2 dataset, the setting in this paper is [0, 16, 53, 126, 246]. the Batchsize size is 1, the learning rate is initialized to 0.01, and the number of iterations is 50 epochs;

(4) Positive samples N of STEEL dataset is set as [0,10,50,150,300,750], Batchsize size is 10, Learning rate is initialized as 0.1, and the number of iterations is 90 epochs.

This paper focuses on three sets of experiments:

(1) Test the AP of the model on the KSDD dataset, DAGM dataset, KSDD2 dataset, and STEEL dataset

(2) Using the KSDD2 dataset to verify the effectiveness of the multi-scale feature fusion module and pyramid attention module;

(3) The effects of different weight modules in the pyramid attention module on the AP and AUC of the model were tested on the KSDD2 dataset.

The experiments in this paper are based on the Ubuntu16.04 system, and the code running environment

TABLE 1. Experimental results of the KSDD dataset.

Methods	F-AnoGAN	SDA	Uninf student	Mix Sup	MaMi	MFFPA
N=0	39.4		57.1	93.4	98.5	94.72
N=5				99.1	99.5	98.95
N=10				99.4	99.7	99.02
N=15				99.2	100	99.11
N=20				99.9	100	100
N=33		99.9		100	100	100

TABLE 2. Experimental results of the DAGM dataset.

Methods	F-AnoGAN	Uninf student	Mix Sup	MaMi	MFFPA
N=0	19.5	66.8	74	80.9	82.9
N=5			94.6	100	99.2
N=15			100		100
N=45			100		100
N=1000			100		100

is Python 3.8, Pytorh 1.8.0, and torchvision 0.9.0. The GPU used in the experiment is RTX2080Ti (video memory: 11GB). The code running environment for FPS calculation is Python 3.8, Pytorh 2.1.2, and torchvision 0.10.0. The GPU used in the experiment is GTX1080Ti (video memory: 11GB).

C. COMPARISON EXPERIMENT

In order to verify the effectiveness of the proposed improvements in this paper, this paper compares with the detection algorithms with excellent results in recent years on the KSDD2 dataset and the STEEL dataset, and the unsupervised methods compared to this article are F-AnoGAN proposed in 2019 [44],Uninf student proposed in 2020 [45], SGSF proposed in 2022 [46]. The fully-supervised method are SDA proposed in 2020 [41],PSIC-Net proposed in 2021 [47]. The mixed-supervised method are TNN proposed in2020 [14],Mix sup proposed in 2021 [15],DSR proposed in 2022 [48], MaMi proposed in 2023 [16].

As shown in Table1, Comparison experiments on the KSDD dataset show that in the weakly supervised case, the AP of this paper is 94.72%, which is 3.78% lower than the current best method;in the fully supervised case, the AP of this paper is 100%, which is the same as the previous best method.

As shown in Table 2, Comparative experiments on the DAGM dataset show that in the weakly supervised case, the AP of this paper is 82.9%, which is a 2% improvement over the previous best method, and in the fully supervised case, the AP of this paper is 100%,which is the same as the previous best method.

As shown in Table3, experiments on the STEEL dataset show that in the weakly supervised case, the AP of this paper is 95.51%, which is a 3.91% improvement over the previous best method.

As shown in Table6, comparing experiments on the KSDD2 dataset, the AP of this paper is 88.01% in the weakly-supervised case, which is a 0.81% improvement over the previous best method, and in the fully-supervised case, the AP of this paper is 95.6%, which is 0.6% lower than the current best method.

TABLE 3. Experimental results of the STEEL dataset.

Methods	F-AnoGAN	Uninf student	TNN	Mix Sup	MFFPA
N=0	50.9	54.9		91.6	95.51
N=10				92.7	96.32
N=50				93.2	97.10
N=150				96.9	98.57
N=300				98.1	98.67
N=750			98.74	98.8	98.94

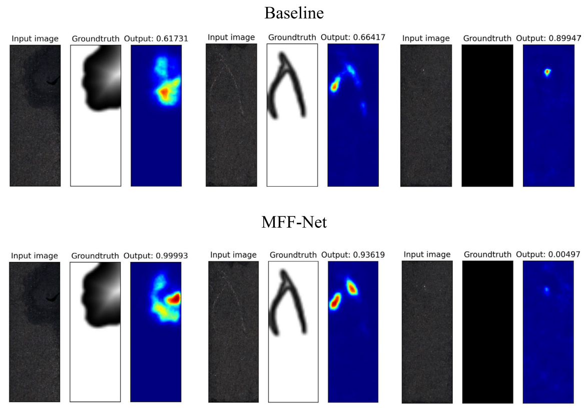

FIGURE 7. Visualisation of results diagram.

D. VISUALIZATION RESULTS

As shown in Fig.7, at the top of the image are the scores for defect detection, When defects are detected and classified correctly, the use of multi-scale feature fusion module and pyramid attention module can make the scope of attention of the model wider, and the model has better discrimination ability in the defective parts. Moreover, when no defects are detected, the model before improvement will have classification errors. After improvement, the model can accurately judge and has higher identification ability. From the visualization results, it can be seen that the proposed multi-scale feature fusion module and pyramid attention module can effectively enhance the feature extraction ability and discrimination ability of the model.

E. ABLATION STUDIES

In order to investigate the effect of the multi scale fusion module(MFF) and pyramid attention module(PA) on the model, several groups of comparative experiments were carried out in this paper. Table 5 and Table 6 shows the results of the ablation experiments on the KSDD2 dataset in this paper.

When the number of pixel level annotations n=0, adding MFF module can improve AP by 6.94%; In all cases, AP increased by 3.44%; When the number of labels is n=0, the AP increases by 11.15% by adding PA module. In all

TABLE 4. Experimental results of the KSDD2 dataset.

Methods	F-AnoGAN	SDA	Uninf student	Mix Sup	PSIC	SGSF	DSR	MaMi	MFFPA
N=0	55.0		65.3	74.4		86.3	87.2	80	88.01
N=16				82.3			91.4	89.7	91.49
N=53				89.4			94.6	92.3	93.05
N=126				92.2				94.1	94.16
N=246		68.7		95.4	93.3		95.2	96.2	95.60

TABLE 5. Ablation experimental results of the KSDD2 dataset(evaluation metrics:AP).

pixel-level annotation	Baseline	MFF	PA	MFF+PA
N=0	74.51	81.45	85.66	88.01
N=16	82.31	87.84	90.01	91.49
N=53	89.45	91.85	92.33	93.05
N=126	92.30	93.84	93.42	94.16
N=246	94.02	94.82	94.72	95.60
Average	86.52	89.96	91.23	92.46

TABLE 6. Ablation experimental results of the KSDD2 dataset(evaluation metrics:AUC)).

pixel-level annotation	Baseline	MFF	PA	MF+PA
N=0	89.49	90.72	95.07	93.87
N=16	92.13	94.62	97.34	95.59
N=53	95.99	96.39	97.78	97.15
N=126	96.54	97.48	98.13	97.46
N=246	97.92	98.49	98.60	98.63
Average	94.41	95.54	97.38	96.54

cases, AP increased by 4.71%. When adding MFF module and PA module at the same time, AP was 88.01% in the case of only image level annotation, which was 13.5% higher than baseline. In all cases, the AP of this experiment increased by 5.94%. The experimental structure proves that adding MFF module and PA module at the same time can effectively improve the AP of the model.

When the evaluation metric is AUC and pixel level annotations N=0, adding MFF module can improve AUC by 1.23%; Adding PA module, AUC increased by 5.58%; When used at the same time, AUC increased by 4.38%. In all cases, using MFF module and PA module, this method also has 2.13% improvement. Experimental results show that the proposed model can effectively detect defects, and also show that the introduction of shallow features will affect the AUC of the model.

As shown in Table7 and Table8, the results of the ablation experiments on the use of Attention weight module. When using AP as an evaluation metric, using CA has approximately 1% improvement compared to SE.When using AUC as an evaluation metric, as seen from the Table8, Using SE or CA as attention weights, there is not much difference in AUC between the two. The experimental results show

TABLE 7. Ablation experimental results of the Attention weight module(evaluation metrics:AP).

	N=0	N=16	N=53	N=126	N=246
Baseline	74.51	82.31	89.45	92.30	94.02
SE	84.86	86.53	90.01	92.92	94.45
CA	85.66	90.01	92.33	93.42	94.72

TABLE 8. Ablation experimental results of the Attention weight module(evaluation metrics:AUC).

	N=0	N=16	N=53	N=126	N=246
Baseline	89.49	92.13	95.99	96.54	97.92
SE	95.21	95.18	97.21	97.86	98.57
CA	95.07	97.34	97.78	98.13	98.60

TABLE 9. Model parametric quantities and computational analysis.

	MFF	PA	Params	FLOPs	FPS
Baseline			15.64M	47.99G	35.15
	✓		8.58M	31.79G	46.03
		✓	10.87M	32.35G	46.24
	✓	✓	11.29M	33.32G	41.15
MaMiNe			16.82M	50.72G	32.86

that using channel attention in classification networks can effectively improve the model's AUC.

F. ANALYSIS OF MODEL PARAMETERS AND FLOPS

As shown in Table9, compared with the baseline, it can be seen that by introducing shallow features and reducing the number of channels in the last layer of the segmentation network, the parameter count of the model is reduced by 27.8%, and the FLOPs is reduced by 30.57%. Compared with MaMiNet, the parameter count of the model is only 67.1% of MaMiNet, and the FLOPs is only 65.69% of MaMiNet, The model is also higher than the previous best model on FPS, fully demonstrating the efficiency of the model.

V. CONCLUSION

The existing two-stage surface defect detection model ignores the shallow characteristics of the segmented network, and the classification network cannot effectively utilize the characteristics of the segmented network transmission. This paper proposes a two-stage detection model based on multi-scale feature fusion and pyramid attention. Ablation experiments show that the multi-scale feature fusion module proposed in this paper can effectively use shallow features and improve the accuracy of the model. In addition, the pyramid attention module proposed in this paper can obtain more comprehensive feature information and effectively improve the discrimination ability of the model. The experimental results on the KSDD2 dataset show that the model proposed in this paper achieves excellent results with less computational overhead.

In future work, we will further study how to effectively integrate multi-scale features. The multi-scale feature fusion module used in this paper is relatively rough, without considering the relationship between features of adjacent layers, and the outlookattention also affects the inference speed of the model. In the next step, we will study how to better integrate multi-scale features and seek better ways to replace outlookattention.

REFERENCES

[1] E. Chen, S. Chen, T. Ye, and Y. Liu, "Degradation-adaptive neural network for jointly single image dehazing and desnowing," *Frontiers Comput. Sci.*, vol. 18, no. 2, pp. 1–3, Apr. 2024.

[2] C. Li, E. Hu, X. Zhang, H. Zhou, H. Xiong, and Y. Liu, "Visibility restoration for real-world hazy images via improved physical model and Gaussian total variation," *Frontiers Comput. Sci.*, vol. 18, no. 1, pp. 1–3, Feb. 2024.

[3] Y. Liu, Z. Yan, J. Tan, and Y. Li, "Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1643–1657, Apr. 2023.

[4] J. Zhang, Y. Cao, Z.-J. Zha, and D. Tao, "Nighttime dehazing with a synthetic benchmark," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 2355–2363.

[5] Y. Liu, Z. Yan, S. Chen, T. Ye, W. Ren, and E. Chen, "NightHazeFormer: Single nighttime haze removal using prior query transformer," in *Proc. 31st ACM Int. Conf. Multimedia*, Oct. 2023, pp. 4119–4128.

[6] X. Zihao, W. Hongyuan, Q. Pengyu, D. Weidong, Z. Ji, and C. Fuhua, "Printed surface defect detection model based on positive samples," *Comput., Mater. Continua*, vol. 72, no. 3, pp. 5925–5938, 2022.

[7] X. Fang, Q. Luo, B. Zhou, C. Li, and L. Tian, "Research progress of automated visual surface defect detection for industrial metal planar materials," *Sensors*, vol. 20, no. 18, p. 5136, Sep. 2020.

[8] Z. Li, X. Tian, X. Liu, Y. Liu, and X. Shi, "A two-stage industrial defect detection framework based on improved-YOLOv5 and optimized-inception-ResnetV2 models," *Appl. Sci.*, vol. 12, no. 2, p. 834, Jan. 2022.

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

[10] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9584–9592.

[11] M. Jonak, S. Jezek, and R. Burget, "Evaluation of nested U-Net models performance on MVTec AD dataset," in *Proc. 14th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2022, pp. 70–75.

[12] Q. Li, A. Arnab, and P. H. Torr, "Weakly-and semi-supervised panoptic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 102–118.

[13] C. Ge, J. Wang, J. Wang, Q. Qi, H. Sun, and J. Liao, "Towards automatic visual inspection: A weakly supervised learning method for industrial applicable object detection," *Comput. Ind.*, vol. 121, Oct. 2020, Art. no. 103232.

[14] J. Božic, D. Tabernik, and D. Skocaj, "End-to-end training of a two-stage neural network for defect detection," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5619–5626.

[15] J. Božič, D. Tabernik, and D. Skočaj, "Mixed supervision for surface-defect detection: From weakly to fully supervised learning," *Comput. Ind.*, vol. 129, Aug. 2021, Art. no. 103459.

[16] X. Luo, S. Li, Y. Wang, T. Zhan, X. Shi, and B. Liu, "MaMiNet: Memory-attended multi-inference network for surface-defect detection," *Comput. Ind.*, vol. 145, Feb. 2023, Art. no. 103834.

[17] J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout, "Steel defect classification with max-pooling convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–6.

[18] S. Kim, W. Kim, Y.-K. Noh, and F. C. Park, "Transfer learning for automated optical inspection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2517–2524.

[19] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *Int. J. Adv. Manuf. Technol.*, vol. 94, nos. 9–12, pp. 3465–3471, Feb. 2018.

[20] Z. Liu, K. Liu, J. Zhong, Z. Han, and W. Zhang, "A high-precision positioning approach for catenary support components with multiscale difference," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 700–711, Mar. 2020.

[21] Y. Huang, C. Qiu, X. Wang, S. Wang, and K. Yuan, "A compact convolutional neural network for surface defect inspection," *Sensors*, vol. 20, no. 7, p. 1974, Apr. 2020.

[22] Y. Bao, K. Song, J. Liu, Y. Wang, Y. Yan, H. Yu, and X. Li, "Triplet-graph reasoning network for few-shot metal generic surface defect segmentation," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[23] H. Feng, K. Song, W. Cui, Y. Zhang, and Y. Yan, "Cross position aggregation network for few-shot strip steel surface defect segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[24] G. Xie, J. Wang, J. Liu, Y. Jin, and F. Zheng, "Pushing the limits of fewshot anomaly detection in industry vision: Graphcore," in *Proc. 11th Int. Conf. Learn. Represent.*, 2022, pp. 1–18.

[25] M. Rudolph, B. Wandt, and B. Rosenhahn, "Same same but DifferNet: Semi-supervised defect detection with normalizing flows," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1906–1915.

[26] K. Batzner, L. Heckler, and R. König, "EfficientAD: Accurate visual anomaly detection at millisecond-level latencies," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2024, pp. 128–138.

[27] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 2048–2057.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[29] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.

[30] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[31] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel 'squeeze and excitation' blocks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 540–549, Feb. 2018.

[32] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.

[33] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1971–1980.

[34] J. Wang, Y. Chen, R. Chakraborty, and S. X. Yu, "Orthogonal convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11502–11512.

[35] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," in *Proc. Asian Conf. Comput. Vis.*, 2022, pp. 1161–1177.

[36] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[37] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.

[38] Y. Quan, D. Zhang, L. Zhang, and J. Tang, "Centralized feature pyramid for object detection," *IEEE Trans. Image Process.*, vol. 32, pp. 4341–4354, 2023.
[39] C. Wang, W. He, Y. Nie, J. Guo, C. Liu, Y. Wang, and K. Han, "Gold-YOLO: Efficient object detector via gather-and-distribute mechanism," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 1–19.
[40] L. Yuan, Q. Hou, Z. Jiang, J. Feng, and S. Yan, "VOLO: Vision outlooker for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 6575–6586, May 2023.
[41] D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, Mar. 2020.
[42] M. Wieler, T. Hahn, and F. A. Hamprecht, "Weakly supervised learning for industrial optical inspection," Dataset, 2007.
[43] I I. O. A. Grishin and V. Boris. (2019). *Severstal: Steel Defect Detection*. [Online]. Available: https://kaggle.com/competitions/severstal-steel-defect-detection
[44] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.
[45] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student–teacher anomaly detection with discriminative latent embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4182–4191.
[46] P. Xing, Y. Sun, and Z. Li, "Self-supervised guided segmentation framework for unsupervised anomaly detection," 2022, *arXiv:2209.12440*.
[47] L. Lei, S. Sun, Y. Zhang, H. Liu, and W. Xu, "PSIC-net: Pixel-wise segmentation and image-wise classification network for surface defects," *Machines*, vol. 9, no. 10, p. 221, Sep. 2021.
[48] V. Zavrtanik, M. Kristan, and D. Skočaj, "Dsr—A dual subspace re-projection network for surface anomaly detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 539–554.

YING TANG received the B.E. degree from the North University of China, in 2021. He is currently pursuing the master's degree in engineering with Changzhou University. His main research interests include computer vision and defect detection.

HONGYUAN WANG received the Ph.D. degree in computer science from Nanjing University of Science and Technology. He is currently a Professor and a Ph.D. Supervisor. His main research interests include computer vision, pattern recognition, and intelligent systems. He is a Senior Member of CCF.

QUNYING ZHOU received the B.S. degree in mathematics from Huaiyin Normal University, in 2021. She is currently pursuing the M.E. degree with Changzhou University. Her research interests include computer vision and defect detection.

BOYAN SUN received the Master of Engineering degree from Changzhou University, in 2023. His main research interests include computer vision and defect detection.