Surface Defects Inspection of Cylindrical Metal Workpieces Based on Weakly Supervised Learning

Mu Ye (yemu1138178251@163.com)
Shanghai University of Engineering Science https://orcid.org/0000-0002-1163-6579
Weiwei Zhang
Shanghai University of Engineering Science
Guohua Cui
Hebei University of Engineering
Xiaolan Wang
Huazhong University of Science and Technology: Wuhan

Research Article

Keywords:

Posted Date: June 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-598050/v1

Version of Record: A version of this preprint was published at The International Journal of Advanced Manufacturing Technology on December 2nd, 2021. See the published version at https://doi.org/10.1007/s00170-021-08399-z.

Noname manuscript No.
(will be inserted by the editor)

Surface defects inspection of cylindrical metal workpieces based on weakly supervised learning

Mu Ye · Weiwei Zhang · Guohua Cui · Xiaolan Wang

Received: date / Accepted: date

Abstract In industrial vision system, metal surface is anisotropic under light in all directions and it is inevitable to cause local overexposure due to the natural reflection of active strong light, especially on the cylindrical metal surface. In this paper, injector valve is taken as the representative of cylindrical metal workpieces. Since the variety and complexity of cylindrical metal workpieces defects, and its contrast with the background of workpieces fluctuates, making samples annotating time-consuming and be of high cost. In order to solve the above challenges, this paper proposes an end-to-end weakly supervised learning framework to classify and segment defects. Firstly, a deep integrated residual attention convolutional neural network (IRA-CNN) is designed. IRA-CNN is composed of multiple IRA-Block. Two residual maps are included in IRA-Block to improve its bilateral nonlinearity and the robustness. IRA-block adds integrated attention module (IAM) which includes channel attention submodule and spatial attention submodule. The channel attention submodule adaptively extracts information from the global average pool layer and the global maximum pool layer to obtain the channel attention feature map. IAM can be well integrated into the IRA-CNN makes the neural network suppress the interference of useless background area and highlight the defect area. Finally, the weakly supervised segmentation method relies on Grad-CAM++ to generate saliency map to improve segmentation accuracy. The experimental results show that the accuracy of defect classification reaches 97.7% and the segmentation accuracy is significantly improved compared with the benchmark method in the injector valve dataset which include 6747 images.

Keywords Machine vision · Defect detection · Deep learning · attention network · Convolutional neural network

Mu Ye
School of Mechanical and Automotive Engineering,Shanghai University Of Engineering Science, Shanghai, 201620, China
E-mail: yemu1138178251@163.com

Weiwei Zhang
School of Mechanical and Automotive Engineering,Shanghai University Of Engineering Science, Shanghai, 201620, China
School of Vehicle and Mobility,Tsinghua University,100089,Beijing,China
E-mail: zwwsues@163.com

1 Introduction

Cylindrical metal workpiece needs to be matched with other parts in kinematic pairs. Its surface adhesion directly affects the performance of workpiece and even mechanical system. The injector valve, as the representative of cylindrical metal workpieces, has multiple surface quality hazards in its total lifecycle. The surface quality of the injector valve affects the open/close dynamic performance of Liquid pressure circuit. The surface defects of injector valve will have a serious negative impact on the engine injector and reduce automotive service life. Therefore, in the manufacturing process of injector valve, defect inspection is vital for production. Due to the improper manufacturing process and raw material quality, there are some defects on the surface of injector valve, such as surface Dirt, Scratch, Electrochemical Corrosion (EC) and so on. In recent years, machine vision has gradually replaced the manual detection and applied to automatic defect inspection, which can improve the quality and life of products by automatically identifying product defects.

The surface image of injector valve acquired by the system is shown in Fig.1, This paper is dedicated to de-

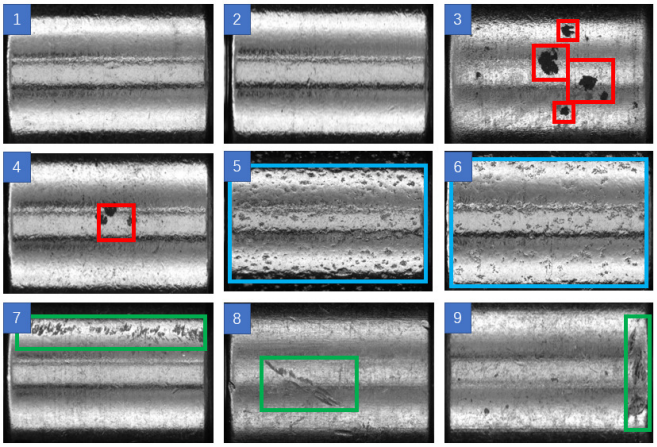

Fig. 1 Three kinds of defects in injector valve: Non-Defect:1,2. Dirt:3,4. EC:5,6. Scratch :7,8,9. It can be seen EC covers the whole valve area, so it does not need to be segmented.

tecting three types of defects with different shape features: Dirt, Scratch, Electrochemical Corrosion (EC), and segment dirt area and scratch area. Compared with the normal injector valve, the dirt area of the valve has smaller gray value and irregular shape. The main characteristics of the scratch defects are that the gray value is similar to the normal area, the contrast ratio is low, and the defects are distributed in strip. The electrochemical corrosion is characterized by small pits in large area of the injector valve, and according to the experimental observation, EC will cause pits in all areas of the whole injector valve, segmentation is meaningless for the defect of EC. Therefore, it is not necessary to segment the defect of EC. These defects have a negative impact on the quality of the injector. The above defects are difficult to be found by manual observation, and some surface defects can only be observed by high resolution cameraTherefore, it is necessary to detect the surface defects of injector valve for product quality control.

The existing surface defect inspection methods based on machine vision mainly targeted at four types of surface: 1) non-textured surface; 2) repeated pattern surface; 3) uniform textured surface; 4) non-uniform textured surface. Cylindrical metal workpieces can be classified as non-textured surface. For detecting this kind of surface defects, the traditional machine vision methods are statistical-based method and model-based method. Statistical-based methods include first-order statistical methods (Gray mean, Gray variance), second-order statistical methods, LBP [1] and gray level co-occurrence matrix (GLCM) [2]. The model-based method uses a hand-designed feature extractor to extract features and uses classifiers (SVM [3], Random Forest [4]) to classify defects. The statistical-based method is simple in technology and low in computation cost, but the accuracy is hard to meet the industry requirements. The model-based method has relatively good effect on the surface defect inspection, but the traditional machine vision method relies too much on the hand-designed feature descriptors such as LBP, HOG, GLCM which have poor robustness for various defects in Cylindrical metal workpieces.

In recent years, deep learning method has shown its advantages in defect segmentation [5][6], defect detection and classification [7][8][10][11][12][13][15]. Deep learning has the characteristics of high precision and wide application by automatically extracting image features. In the field of defect inspection, the common surface defect inspection tasks include fabric defect detection, steel defect detection, solar panel defect detection, LED chips defect detection and so on. In this field, deep learning can be mainly applied to defect classification and detection and defect segmentation. Aiming at the problem of defect classification and detection, Xu [8]proposed roller bearing defect classification based on SDD-CNN. This method uses SSAD method for data expansion, and uses InceptionV3[9] as the classification network, and its classification accuracy reaches 99.56%. Chen [10] et al. proposed a multispectral CNN structure for surface defect classification of solar cells. Three convolutional neural networks are established for three spectral channels for classification, the classification accuracy is 88.24%. Cheon [11] proposed a CNN structure for wafer surface defect classification, which can detect unknown defects by clustering the eigenvectors of the same defect types. He [12] et al. proposed a semi-supervised defect classification method to make up for the shortage of samples in supervised training. This method uses GAN to generate a large number of unlabeled data. However, the above methods only classify defects, and do not get the location information of defects and the feature information of defects (such as size, distribution, etc.), so some researchers add object detection and image segmentation methods to defect detection. Li [13] et al. used the improved YOLO [14] network to detect six kinds of steel surface defects, and achieved 97.55% mAP and 95.86% recall. Su [15] et al. designed a novel bidirectional attention feature pyramid network structure and embedded it into the regional proposal network to improve the efficiency of Faster-RCNN [16] in detecting surface defects of solar cell. The target detection method needs to use the label with bounding box for training.

Using image segmentation for defect inspection can obtain more accurate defect boundary and shape. Tabernik [17] et al. proposes a segmentation framework for crack

segmentation, which first segments the defect and then classifies defects combined with the segmentation feature map. The experimental results show that the average accuracy of this method is at least 1.9% higher than that of deeplabv3 + [18] and u-net [19]. Tao [20] et al. uses an encoder-decoder structure to segment defects. After that, the segmented region is cropped out for classification. Wang [21] et al transformed the inference of CRF into CNN operation, which fully combined CRF with deep convolutional neural network and significantly improved the defect segmentation effect. However, in the industrial environment, the variety and complexity of cylindrical metal workpieces defects, and its contrast with the background of workpieces fluctuates, making samples annotating time-consuming and be of high cost which is the main weakness of image segmentation using full convolution neural network.

Based on the above shortcomings, weakly supervised learning can effectively solve the leakage problem of labeled defect samples and realize the detection or segmentation of defect image only by using image level annotation. Among them, CAM [22] is an important method to realize weakly supervised learning, and CAM can generate a saliency map to show the probability that each pixel belongs to a certain defect category. Lin [23] used CAM method to visually predict the depth convolution neural network and thus locate the LED defect location. Chen [24] et al proposed a robust weakly supervised learning method for surface defect detection, which uses transfer learning to get CAM. Xu [25] et al proposed a weakly supervised detection framework, in which CNN model was trained to identify surface cracks in motor commutators. The method achieved 99.5% recognition accuracy in Kolektor SSD dataset. The above researches all use CAM method, but CAM needs to replace the fully connected layer with global average pooling layer and then retraining, which leads to the high training cost of the model and limits the use scenarios of the model. Grad-CAM [26] and Grad-CAM++ [27] solve the above problems. The above two methods use the gradient information of the last layer of CNN feature map to assign values to each neuron. The difference lie in that Grad-CAM++ uses a weighted combination of the positive partial derivatives of the last convolutional layer feature maps with respect to a specific class score as weights to generate a visual explanation for the corresponding class label. For the case of multiple defects in one image, Grad-CAM++ can produce more accurate saliency maps. Therefore, Grad-CAM++ has better robustness and interpretability.

The saliency map generated by Grad-CAM++ still has a large noise effect on defect segmentation. Adding attention module in the backbone can effectively alleviate the problem. Attention module can emphasize the weak features of the defect, and suppress the noise in the image. Due to the complex surface and uneven light distribution of the cylindrical metal workpieces, adding attention module is suitable for the surface defect inspection of the injector valve. Attention module includes channel attention module and spatial attention module. These two methods realize "what to look" and "where to look" by calculating the weight of channel direction and spatial position to the feature graph. Channel attention module can make CNN network select the feature map related to the defect, so as to suppress useless background information. Spatial attention module can make CNN network pay attention to the defect location information in the feature map. At present, attention mechanism has been widely used in various tasks [28][29][30]. Hu [28] et al. used global mean aggregation characteristics to calculate channel direction attention in their squeeze and excitation modules. Ji [29] et al. used attention module to align the context information between the feature maps at different scales and the final prediction of the saliency map. Woo [30] et al. proposed CBAM, which infers the attention map in turn along two independent dimensions (channel and space), and then multiplies the attention map by the input feature graph to refine the adaptive feature. Inspired by the structure of CBAM, this paper proposes the integrated attention module (IAM). IAM enables classifiers to focus more on defect-related feature maps and regions.

This paper presents a framework for detecting the defects on the outer Cylindrical Mental surface of fuel injector valve, the contributions of this paper are as follows.

1. Aiming at the defect inspection of cylindrical metal workpieces, an Integrated residual attention convolutional neural network (IRA-CNN) is proposed. IRA-CNN is composed of multiple IRA-Block. Two residual maps are included in IRA-Block to improve the robustness. Every IRA-Block combines a distinctive attention module to improve the interpretability of network. Therefore, the network structure has a good classification effect for the defects of cylindrical metal workpiece.

2. We proposed an integrated attention module (IAM) which is composed of a channel attention submodule and a spatial attention submodule. IAM adaptively selects the input proportion of GAP layer and GMP layer to adjust the output of different IRA-Block. IAM makes full use of the comprehensive characteristics of gap layer and GMP layer. Therefore, IAM can suppress the background of the image and highlight the defect area of the cylindrical metal workpiece.

3. In this paper, a weakly supervised learning strategy is proposed to segment the defects of injector valve. The strategy can segment the defects only by using image level labels. After classification by IRA-CNN, the saliency map is generated by Grad-CAM++. The pixel level segmentation of defects is based on the saliency map, which greatly simplifies the task of image segmentation. The time of pixel level labeling is saved, and the segmentation accuracy is better for the defects of cylindrical metal workpieces.

The rest of this study is as follows. The second section introduces the methodology in detail. The third section introduces the experiments and related performance evaluation. Finally, the fourth part is the conclusion and discussion.

2 Methodology

In this section, the research of cylindrical metal workpiece defects inspection is mainly based on the injector valve. surface defects inspection of cylindrical metal workpiece based on weakly supervised learning structure is shown in Fig.2. We mainly introduce the architectures of the classification module, Integrated attention module, and segmentation module.

2.1 Classification module

Convolutional neural network extracts local feature information from image by convolutional layers and compresses feature information by pooling layers. When detecting the surface defects of fuel injector, the sliding window is used to process the image with the window size of 192×192, which can detect the defects more accurately without reducing the resolution of the original image.

In this paper, IRA-CNN is proposed to extract the features of the outer surface of the injector valve. IRA-CNN is composed of multiple IRA-Blocks in series, in which the attention module (IAM) is integrated. In the next section, the structure of IAM will be explained in detail. EfficientNet [31] and MobileNet [32] also have similar structure, that is, the attention module is integrated into the convolution extraction block. In the IRA-Block, we use 3 × 3 convolution kernel to extract features, and stack multiple 3 × 3 convolution kernels to expand Receptive field. IRA-Block is shown in the Fig. 3.

The function of convolution layer (Conv) is to extract the feature of a local region. Different convolution kernels are equivalent to different feature extractors. The defect features of different positions are extracted by sliding the Conv block in the image. The convolution function is as follows

$$Y = \Phi\left(W \otimes X + b\right) \tag{1}$$

Where, X is the input an image. W represents the weight of the Conv filter. b is the bias of the Conv filter. After convolution operation, the output Y is obtained by using nonlinear activation function $\Phi(\cdot)$.

In order to obtain a larger receptive field and reduce parameter, we set four 3 × 3 convolution stacks to extract features. The number of convolution kernel channels is doubled compared with the previous IRA-Block. After feature extraction, IAM is used to suppress the background noise and highlight the defect area. Finally, Max-pooling is used to down sample and compress features. There is a residual mapping between the input and output of IRA-Block.

IRA-CNN uses three IRA-Blocks to extract the features continuously, the last feature map is sent to the fully connected layer for feature aggregation. The last layer of the fully connected layer is the output of the whole network. The output has the same number of neurons as the labels and can be classified by Softmax classifier. Y_i represents the input of the Softmax classifier while $P(Y_i)$ represents the output probability. Softmax classifier formula is as follows

$$P\left(Y_i\right) = \frac{\exp\left(Y_i\right)}{\Sigma_{i=1}^{K} \exp\left(Y_i\right)} \tag{2}$$

The above is the basic structure of IRA-CNN. IRA-CNN can effectively extract the defect features of injector valve, IRA-CNN network structure is shown in Table 1.

Table 1 This table contains the structure of the IRA-CNN, including the size of each feature layer, and the operations performed by each layer.

Layer	Kernel Size	filters	Stride	Output
-	1×1	32	1	192×192×32
-	1×1	32	1	192×192×32
IRA-Block	1×1,3×3	64	1	192×192×64
Max-Pooling	2×2	-	2	96×96×64
IRA-Block	1×1,3×3	128	1	96×96×128
Max-Pooling	2×2	-	2	48×48×128
IRA-Block	1×1,3×3	256	1	48×48×256
Max-Pooling	2×2	-	2	24×24×256
FC Layer1	1024	FC	-	1024
FC Layer2	1024	FC	-	1024
Softmax	4	Softmax	-	4

2.2 Integrated Attention Module

Because of the randomness of the surface defect location of the injector valve, IAM is proposed and integrated

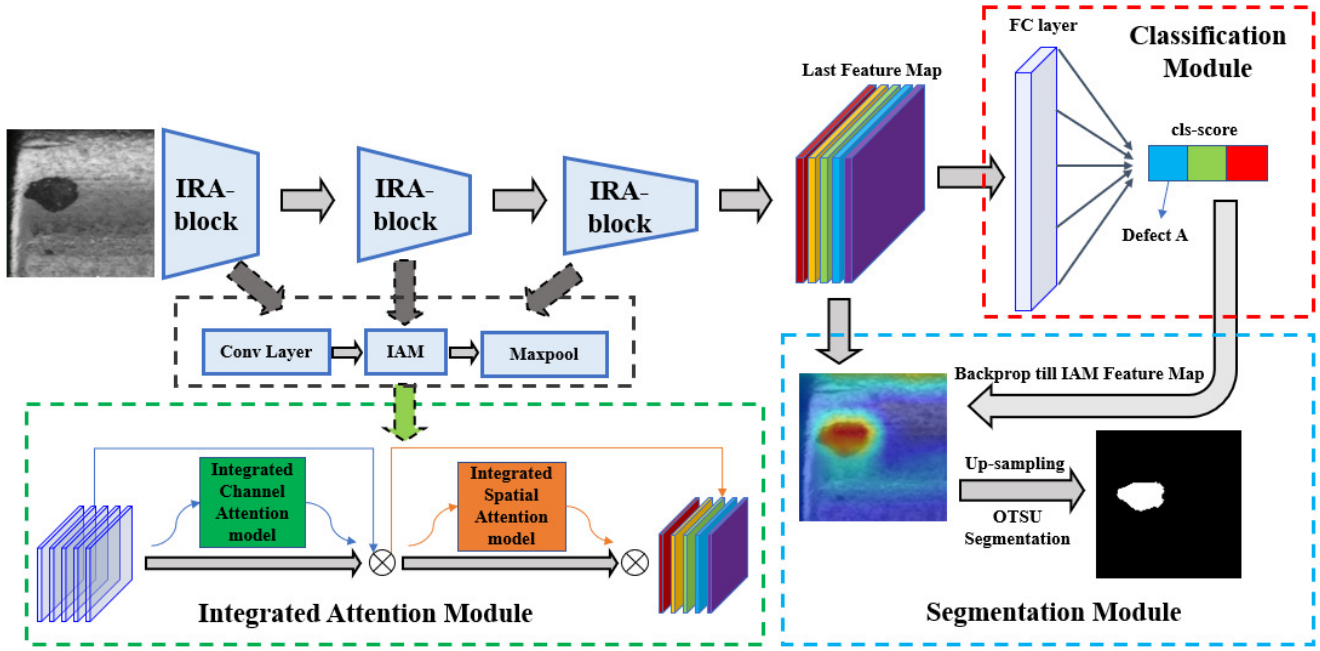

Fig. 2 Framework structure of cylindrical metal workpiece surface defect inspection based on weakly supervised learning.

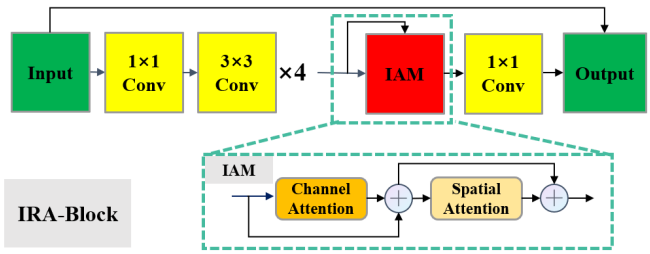

Fig. 3 The structure of IRA-Block, Including IAM and feature extraction.

into the IRA-Block. The proposed Integrated Attention Module(IAM) cascades the channel attention submodule with spatial attention submodule. The input and channel attention map contain a residual connection. Relatively, channel attention map and spatial attention map also contain a residual connection. IRA-CNN can suppress the unnecessary background area and highlight the spatial location of the surface defects of the injector valve.

IAM is inspired by CBAM. CBAM uses channel attention modules to cascade spatial attention modules. In channel attention modules, CBAM compresses the spatial information of feature maps through global max pooling (GMP) and global average pooling (GAP), and then sends GAP and GMP into a shared MLP respectively to generate corresponding feature vectors, and then merge the output feature vectors using element-wise summation. CBAM compulsorily add the GAP layer and GMP layer after MLP feature extraction can't well combine the information of GAP layer and GMP layer. The reasons are as follows: CBAM is biased to GMP. GAP and GMP use shared layer to extract features in CBAM, the inference part can be regarded as GAP and GMP added directly and sent to shared layer. Generally speaking, the value of GMP is larger than GAP. The reason is that the ReLU activation function will filter out a large number of negative values, resulting in many values of 0 in the feature map, which leads to the decline of GAP. Therefore, it is biased for CBAM to directly send GAP and GMP into MLP without feature processing. The proposed IAM multiplies GAP layer and GMP layer by trainable weights and then sends them to MLP layer. The structure of CBAM is equivalent to the case of and removal of all residual connections in IAM. IAM can make channel attention network adaptively select the proportion of GAP or GMP input to MLP. This channel attention module can get channel weights after only one inference. Moreover, residual connection is added between input and channel attention map, and between channel attention map and spatial attention map, which makes network training easier to converge. These are the reasons why IAM can achieve better performance than CBAM.

As shown in Fig.4, Given an intermediate feature map $M_{input} \in \mathbf{R}^{C\times H\times W}$ as input, is the final output attention map. All operations above can be expressed by the following formula

$$M_{ICAM} = (F_{CW} \otimes M_{Input}) + M_{Input} \tag{3}$$

$$M_{IAM} = (F_{SW} \otimes M_{ICAM}) + M_{ICAM} \tag{4}$$

Where $\otimes$ denotes element-wise multiplication. $\mathrm{F_{CW}} \in \mathbf{R}^{C\times1\times1}$ and $\mathrm{F_{SW}} \in \mathbf{R}^{1\times\mathrm{H}\times\mathrm{W}}$ represent channel attention weight and spatial attention weight respectively. $\mathrm{M_{ICAM}}$ is channel attention map. IAM makes IRA-CNN have better classification performance and model interpretability, and improves the accuracy to locate defects.

2.2.1 Integrated Channel Attention Submodule

In this paper, an integrated channel attention submodule is proposed to obtain the channel attention feature map. This submodule extracts information from squeezing each channel feature map, so that the feature layer with stronger semantic information has higher weight. The ways of squeeze include GAP and GMP, which are proved in [22] and [33] to improve the representation ability of the network. Aiming at the cylindrical metal workpiece inspection, GAP and GMP have different representation ability. GMP focuses on the most significant region in the image to compensate the global region which GAP focus on. Therefore, this paper thinks that GAP and GMP should be jointly input into a network. We multiply GAP and GMP by a trainable weight in the input stage and then adds them, and then sends them to a network to extract information. Finally, a residual connection is added between the input and output, which is helpful to the convergence of the network.

The specific implementation is as follows, as shown in Figure 4. Firstly, the input feature map is squeezed into GAP layer $\mathrm{F_{gap}} \in \mathbf{R}^{\mathrm{C}\times1\times1}$ and GMP layer $\mathrm{F_{gmp}} \in \mathbf{R}^{\mathrm{C}\times1\times1}$. Set a parameter $\frac{e^{\alpha}}{e^{\alpha}+1}$ that is greater than 0 and no more than 1 and as the weight of $\mathrm{F_{gap}}$. Relatively, $\frac{1}{e^{\alpha}+1}$ is the weight of $\mathrm{F_{gap}}$, e is the natural constant. So we can ensure that the weight of training is positiveThen $\frac{e^{\alpha}}{e^{\alpha}+1}\mathrm{F_{gmp}}$ and $\frac{1}{e^{\alpha}+1}\mathrm{F_{gap}}$ are added and sent to a hidden layer network (MLP) to extract information. The hidden layer of MLP contains neurons, C/k is the number of channels and C is the reduction rate. Finally, sigmoid function is used to activate the last layer to get the channel weight $\mathrm{F_{CW}}$. The integrated channel attention map $\mathrm{M'_{ICAM}}$ is obtained by multiplying the weight with $\mathrm{M_{Input}}$ by element. The final channel attention map also contains a residual connection to the . The calculation of $\mathrm{M_{ICAM}}$ can be simplified as the following formula

$$\mathrm{F_{CW}} = sigmoid\left(MLP\left(\begin{array}{c}\frac{1}{e^{\alpha}+1}\mathrm{F_{gap}}+\\ \frac{e^{\alpha}}{e^{\alpha}+1}\mathrm{F_{gmp}}\end{array}\right)\right) \tag{5}$$

$$\mathrm{M'_{ICAM}} = \mathrm{F_{CW}} \otimes \mathrm{M_{Input}} \tag{6}$$

$$\mathrm{M_{ICAM}} = \mathrm{M'_{ICAM}} + \mathrm{M_{Input}} \tag{7}$$

2.2.2 Spatial Attention Submodule

Spatial attention submodule can make the network focus on "where" the defect is, which is a supplement to channel attention submodule. The input of the spatial attention submodule is $\mathrm{M_{ICAM}}$. Firstly, the average-pooling and max-pooling along the channel axis are applied to calculate two feature map $\mathrm{F^c_{avg}}$ and $\mathrm{F^c_{max}}$,then the two feature maps are connected and input into a convolution layer to extract features and activate them with activation function. The convolution kernel size is 7×7, and the output of the convolution layer is spatial weights . and are multiplied by the corresponding elements to get . Before getting the final output , there is a residual connection between and . The above operation can be simplified as the following formula

$$\mathrm{F_{SW}} = sigmoid\left(f^{7\times7}\left(\left[\begin{array}{c}Avgpool\left(\mathrm{M_{ICAM}}\right);\\ Maxpool\left(\mathrm{M_{ICAM}}\right)\end{array}\right]\right)\right) \tag{8}$$

$$\mathrm{M'_{IAM}} = \mathrm{F_{SW}} \otimes \mathrm{M_{ICAM}} \tag{9}$$

$$\mathrm{M_{IAM}} = \mathrm{M'_{IAM}} + \mathrm{M_{ICAM}} \tag{10}$$

Where *Avgpool* and *Maxpool* are average pooling and max-pooling along the channel axis respectively.

2.3 Segmentation module

2.3.1 Saliency Map by Grad-CAM++

IAM has made classification network learn "where to look" and "what to look", so this paper uses Grad-CAM++ to generate saliency map. The probability of whether a pixel value belongs to a defect can be ensured in saliency map. Therefore, saliency map is not only the basis to judge whether the network really recognizes the defect feature, but also an important basis for segmentation.

Grad-CAM++ uses the gradient information flowing into the last convolutional layer of the CNN to assign importance values to each neuron for a particular decision of interestThe contribution of each pixel value in the feature map to the classification score is obtained by gradient conduction. The method of obtaining gradient weight is different from that of Grad-CAM. Detailed derivation of gradient weight in [**?**]. Grad-CAM fails to properly localize defects in an image if the image contains multiple occurrences of the same class. Grad-CAM++ solves this problem. Review the calculation method of Grad-CAM++. Suppose that the classification score of a defect classification is S_c (before Softmax operation), (i, j) and (a, b) are iterators over the same activation map A^k are used to avoid confusion. The

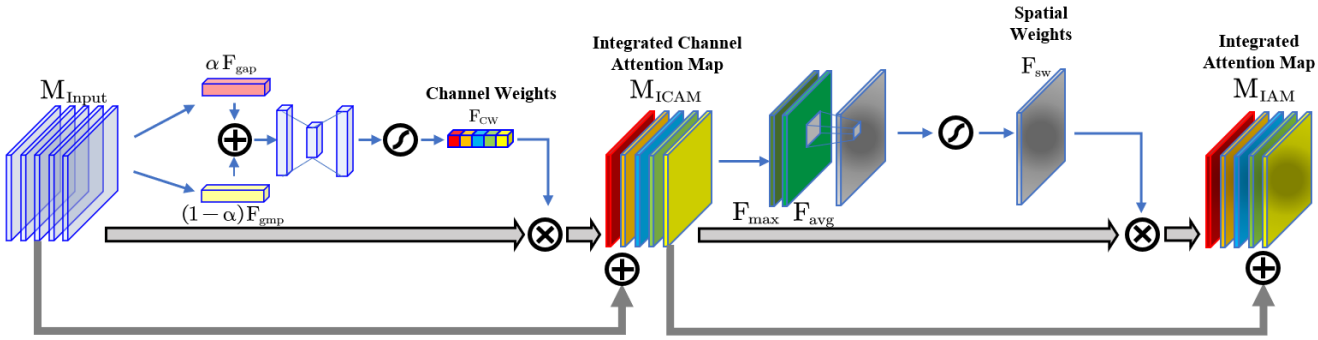

Fig. 4 Frame structure of IAM, IAM includes channel attention submodule and spatial attention submodule.

gradient weight w_k^c of the feature map is calculated as follows

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 S_c}{\left(\partial A_{ij}^k\right)^2}}{2\frac{\partial^2 S_c}{\left(\partial A_{ij}^k\right)^2} + \sum_a \sum_b A_{ab}^k \left\{\frac{\partial^3 S_c}{\left(\partial A_{ij}^k\right)^3}\right\}} \tag{11}$$

$$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc} relu\left(\frac{\partial S_c}{\partial A_{ij}^k}\right) \tag{12}$$

Finally, the gradient weight of the feature map is multiplied by the feature map to obtain the saliency map . The formula is as follow

$$L_{ij}^c = \sum_k w_k^c \cdot A_{ij}^k \tag{13}$$

2.3.2 Segmentation Framework

The saliency map obtained by Grad-CAM++ represents the probability of whether each pixel is a defect, so it is necessary to binarize the saliency map to obtain the contour of the defect area. Binarization methods include fixed threshold method, maximum entropy threshold segmentation method, OTSU segmentation [**?**] method and so on. In these methods, the fixed threshold method needs to manually select the threshold. The maximum entropy threshold segmentation will retain many pixels with low defect probability, which has poor robustness. This paper chooses OTSU method to segment saliency image. OTSU method is based on the idea of clustering. Firstly, the image is divided into panorama and background based on the gray value. Since variance is a measure of the uniformity of gray distribution, the greater the variance between the background and the foreground, the greater the difference between the two parts of the image. When part of the foreground is wrongly divided into background or part of the background is wrongly divided into foreground, the difference between the two parts will become smaller. Therefore, the segmentation that maximizes the variance between classes means the minimum probability of misclassification. OTSU method can highlight the regions with high defect probability in saliency map and suppress the regions with low defect probability to form binary image. The image after using OTSU method is defect segmentation image. Saliency map without defect image will be forced to set to 0, and the gray value of defect segmentation image will also be set to 0.

3 Experiment

3.1 Image acquisition scheme and dataset construction

In this study, the following injector valve machine vision system is used to collect images and detect defects. This system includes Image Acquisition System, Image Analysis System, Mechanical and Electrical Control System. The Mechanical and Electrical Control System uses a Siemens S7-1500 PLC and two S7-200 PLC, as well as a number of motor control injector valves feeding and sorting. The Image Analysis System uses industrial computer with GPU for defect detection, and uses TCP/IP communication protocol to interact with PLC. This system can not only collect and analyze the surface defect image, but also detect whether the outside diameter reach standard, but these are not the focus of this paper. The image acquisition arrangement scheme is shown on the right side of Fig. 5, and the camera, lens, parallel light source and spherical light source are respectively from left to right. The parallel light source irradiates the middle part of the cylindrical valve and the spherical light source irradiates both sides of the cylindrical valve. In order to make it easy for readers to understand this system, the injector valve machine vision system is shown in Fig. 5.

There are few datasets of cylindrical metal workpiece captured by industrial cameras. We build a dataset of surface defects of injector valve, which is used to train

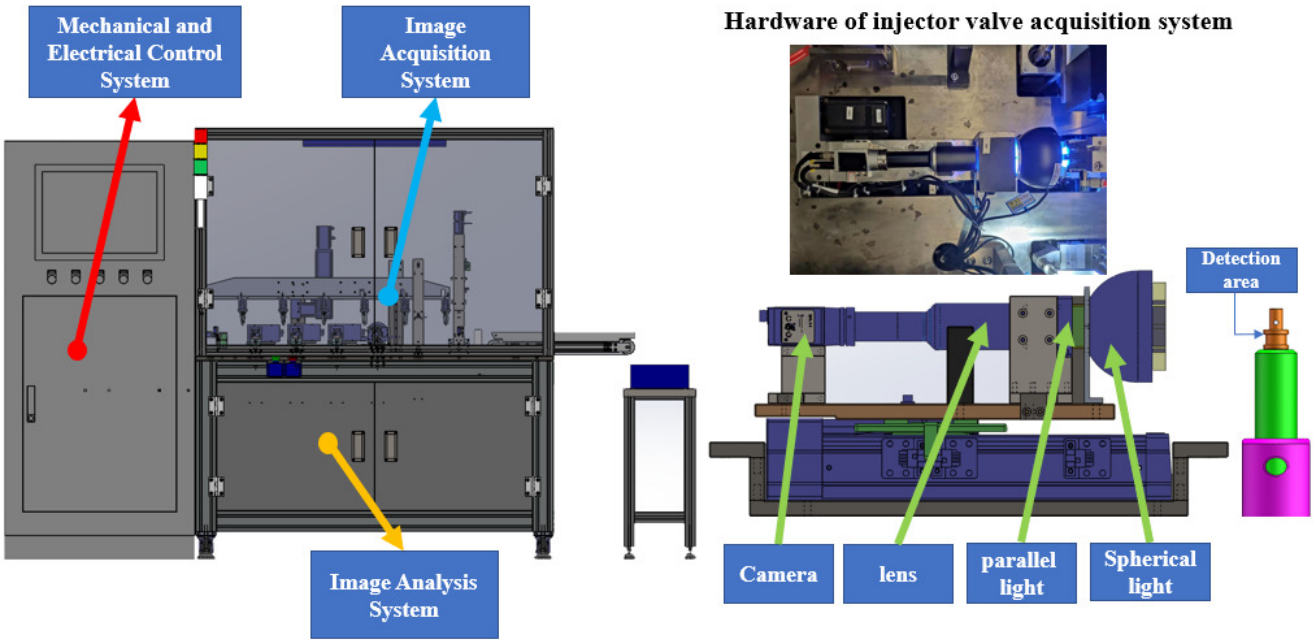

Fig. 5 Injector valve machine vision system, Including Mechanical and Electrical System, Image Acquisition System. Image Analysis System. The right side of figure is the hardware of acquisition system.

and evaluate the proposed model. The dataset was collected by Basler aca2440-20gm camera, and the camera exposure rate was set to 2300. The size of the original image collected by the camera is 2448 × 2048. Since the injector valve image only accounts for a part of the image, we segment the original image to get 768 × 384 valve regions, the segmentation method is obtained by using fixed threshold binarization method. Each valve region is cut into 8 images to get 192 × 192 valve slices as the image in the dataset. Defects often account for only a small part of the image, so the image segmentation into 8 regions is conducive to classification and generating accurate saliency map. The operation of acquiring the dataset is shown in Fig.6. There are 6747 images in the dataset, and the training data and testing data are allocated according to the ratio of 2:3. Among them, pixel level labels are included in the testing data of dirt and scratch to evaluate the segmentation accuracy. Table 2 shows the dataset distribution of training data and testing data.

Table 2 Distribution of surface defects of injector valve dataset. Types of defects include Dirt, Scratch, Electrochemical Corrosion (EC). And dirty, scratch with pixel level annotation.

Dataset (192×192)	Defect type			Good images	Total number of images
	Dirt	Scratch	EC		
Training	652	732	505	810	2699
Testing	978	1098	757	1215	4048
Total	1630	1830	1262	2025	6747
PL-annotation	√	√	×	×	

3.2 Evaluation metrics

We use Precision(P), Recall(R) and F-measure to evaluate classification performance. Moreover, pixel accuracy (PA) and Mean Intersection over Union (MIoU) are used to evaluate the performance of segmentation.

The formula of the above evaluation indicators is as follows

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

$$Pixel\ accuracy = \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} P_{ij}} \tag{17}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{j=0}^{k} P_{ij} + \sum_{j=0}^{k} P_{ji} - P_{ii}} \tag{18}$$

Where TP is the number of defect images which are predicted to be correct by the model; In contrast, FP is the number defect images which are predicted to be false. FN is the number of the actual non-defect which

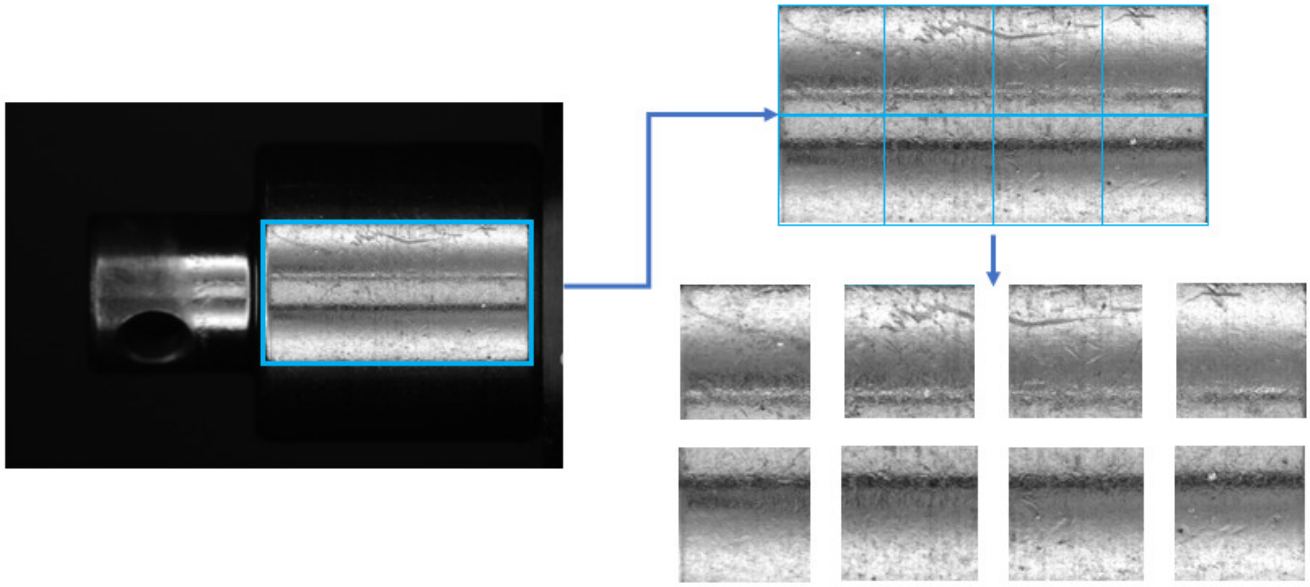

Fig. 6 Dataset building method: each image collected by camera image acquisition is divided into 8 images. These 8 images are input into the dataset.

is mistakenly classified as defect. $\sum_{i=0}^{k} P_{ii}$ represents the number of pixels belonging to class and predicted to be class i as well. $\sum_{i=0}^{k} P_{ij}$ represents the number of pixels that belong to class but are predicted to be class j .

3.3 Implementation Details

The experimental platform is an industrial computer equipped with a NVIDIA GeForce RTX2080 (8G) graphics board. The CPU of the industrial computer is Intel i7-9800x. The training and inference are completed on the industrial computer. The model framework is based on the Pytorch. During the experiment, we use cross entropy loss function. We use Adam optimizer to optimize the learning rate. Among them, the initial learning rate (λ) is 0.001, the first estimated exponential decay rate (β_1) is 0.9, and the second estimated exponential decay rate (β_2) is 0.999 and the epoch of training is 100. Batch size is set to 32. In IAM, a is initially set to 0, so the weight of F_{gmp} and F_{gmp} are initialized to 0.5. Reduction rate set to 2. In the spatial attention submodule, the kernel size of convolution layer is 7×7. In the following ablation experiments, it is shown that increase IAM can significantly improve the classification accuracy.

3.4 Evaluation

This subsection shows experimental results. The evaluation consists of four parts. The first part evaluates

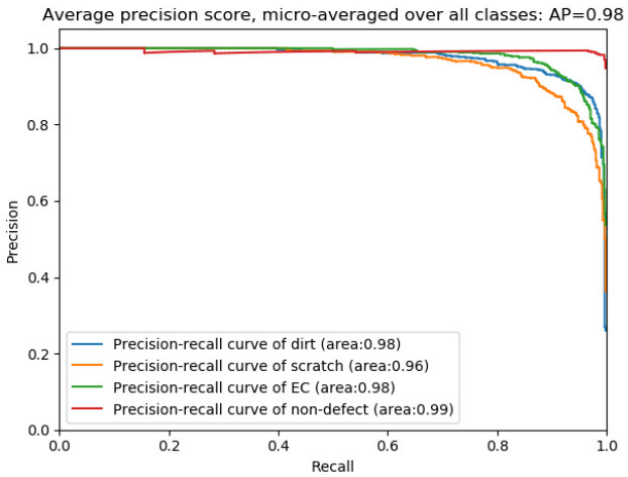

Fig. 7 Classification results of ten methods on the dataset of surface defects of injector valve.

classification of injector valve defects by IRA-CNN. The second part is to evaluate the proposed segmentation framework. The third part is overall evaluation. The last part is time-efficiency evaluation.

3.4.1 classification evaluation

For evaluating classification results, since there is no article on the surface defect detection of injector valve before and the injector valve is made of stainless steel. We compare our method with metal and steel defect classification method. The artificial features of traditional metal defect classification are as follows, (1) GLCM: this method is a classical method to extract texture

features. The angle parameter is set to 0,45,90,135 and the distance parameter is set to 1,2,4,8,16 (2) MLBP: MLBP have the advantages of rotation invariance and gray invariance. Statistical histogram of LBP feature spectrum is used as feature vector for classification. (3) HOG: the feature is formed by calculating and counting the histogram of the gradient direction of the local region of the image. For its parameter setting, block is set to 8 × 8, cell is set to 4 × 4, and stride is set to 4.SVM and MLR are selected in classifier, among which SVM selects linear kernel function. In addition, we compare the deep learning classification methods, such as VGG16 [35], ResNet50. We also compare with s Other weakly supervised defect detection frameworks such as Decaf [36] and RWSLDC [24], in which RWSLDC also has attention module. We also changed IAM to CBAM for training, and added it to the result comparison Results as shown in Fig.8, the classification performance of IRA-CNN is higher than all other models in the test set, reaching 97.7%. GLCM is often used to extract texture feature information, so it is not effective in the area where the texture feature is not obvious. the performance of classification using GLCM is the weakest. SVM classifier has good classification effect in training dataset, but its generalization ability is poor. Compared with the feature extraction method based on manual design, the deep learning method has higher classification accuracy. Compared with RWSLDC which also has attention module, RWSLDC only uses spatial attention module in the last feature layer, and its network depth is shallow, so its classification performance is not as good as IRA-CNN. IRA-CNN has the best classification performance for EC defects, followed by dirt and finally scratch, because scratch account for a small area of the image and have irregular shape. The specific data are shown in Table 3. Precision-Recall curve of each class is drawn in Fig. 7. With the increase of recall, each precision has a downward trend. The results showed that the AP of each class were close to 1.00. The results show that IRA-CNN has impressive classification performance.

In addition, Fig. 9 shows the confusion matrix. The confusion matrix can be used to get the TP, FP and TN of each kind of defects, so as to calculate the precision and recall. It can be observed from Fig. 68that some Dirt defects are easily confused with scratches. These pictures are labeled as 1, 2, 3 and 4 in Fig. 10. In Fig. 10, it can be observed that the pictures labeled as 1 and 2 are actually dirt type, but they are similar to scratch type in shape. The pictures labeled 3 and 4 are of scratch type but similar in shape to the dirty type. In contrast, EC has obvious features. EC is hard to confuse with other defect type. Therefore, we conclude

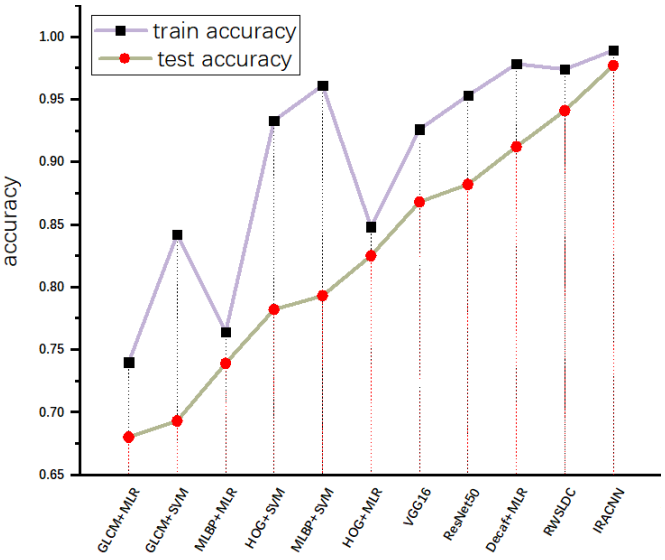

Fig. 8 PR curve of IRA-CNN.

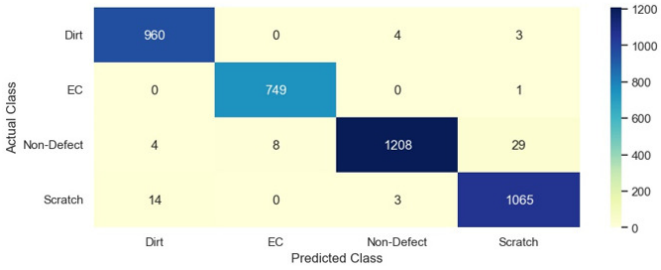

Fig. 9 The confusion matrix of IRA-CNN.

Table 3 Classification accuracy of IRA-CNN in various types of defects.

Defect type	Precision	Recall	F-measure
Dirt	0.981	0.993	0.986
Scratch	0.969	0.984	0.976
EC	0.989	0.998	0.993
Non-Defect	0.994	0.967	0.98

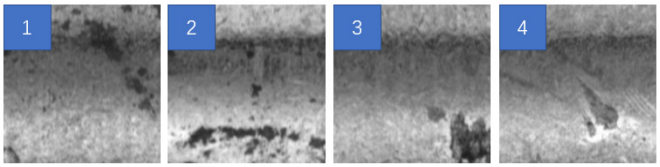

Fig. 10 Dirt surface labeled 1 and 2 have similar characteristics to scratch surface. Scratch surface labeled 3 and 4 have similar shape to dirt surface.

that labeled sample images with distinct features are crucial to achieving high classification accuracy.

3.4.2 Segmentation framework evaluation

According to a large number of experimental observations, EC will cover the whole injector valve, so it does not need to be segmented. Therefore, only dirt

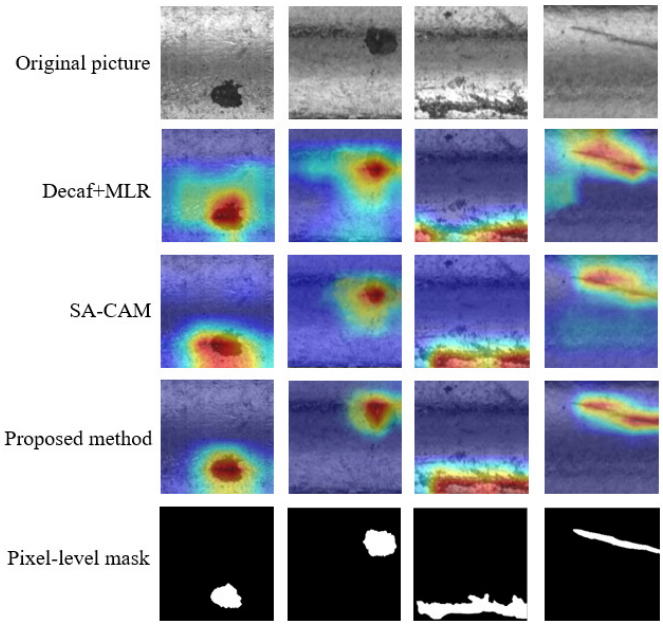

Fig. 11 Comparison of the proposed method with Decaf + MLR and SA-CAM, the last row shows Pixel-level annotation mask.

and scratch defects are selected for segmentation in this experiment. The pixel level annotation of dirty and scratch is labeled with LabelMe. We compare our segmentation module with Ren's and Chen's. Chen's SA-CAM is also a weakly supervised segmentation method, Ren's method is similar to Chen 's. so we choose these two methods as benchmark methods and compare it with CAM and Grad-CAM in the ablation experiment. As shown in Fig. 11 and Table 4, the area segmented by Decaf + MLR usually has a high recall rate due to its large coverage area, but the IOU value is low. Compared with SA-CAM, the precision of our segmentation framework is improved about 4.6%. This is because the segmentation method based on CAM is only sensitive to the large defect area, while our proposed framework focuses on all defect areas in the image, so it covers smaller area. Compared with the two benchmark methods, the PA and IOU of the dirt defect increased by more than 3.5%. The PA and IOU of scratch defect increased by more than 1.5%.

3.4.3 Overall Evaluation

From above classification and segmentation evaluation of our proposed method. We can conclude that IRA-CNN is superior to other classification method in injector valve classification, and performs well in weakly supervised defect segmentation. The reasons are as follows, IAM is integrated in every layer of IRA-CNN. This module makes IRA-CNN pay more attention to the defect area, so it improves the classification performance. IAM helps Grad-CAM++ to produce more accurate saliency map, and Grad-CAM++ makes all defect areas more prominent, which is conducive to improving the segmentation accuracy.

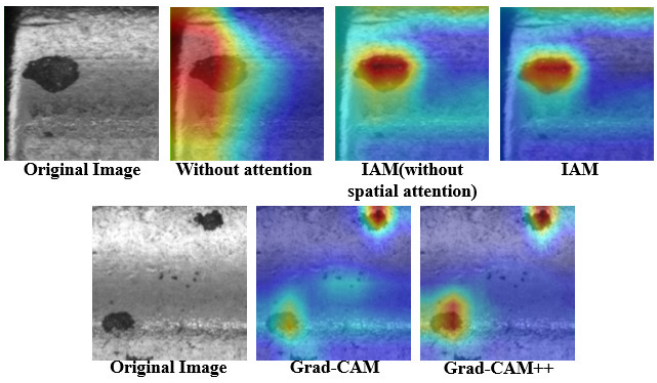

Fig. 12 The above row is the ablation experiment of attention module, and the next row is the comparison between Grad-CAM and Grad-CAM++.

3.4.4 Time-efficiency Evaluation

The results of FPS is the average of 2833 testing defective image. Since the parameter of the IRA-CNN is more than Chen's RWSLDC, the inference time is relatively slow, but compared with SVM and other machine learning classifiers, the inference speed of this method is faster. The time for each method of training and testing is given in Table 5.

3.5 Ablation Studies

3.5.1 Effect of IAM

To prove the effectiveness of IAM module, the attention module is added or replaced in the ablation studies. As shown in table 6, The effectiveness and superiority of IAM can be proved by comparing CBAM with reducing IAM. Fig.12 shows the saliency map generated by Grad-CAM++. The effect of saliency with attention module is better than that without attention module and the saliency map generated by IAM is more accurate, the defect area is more prominent and the background area is suppressed more obviously.

3.5.2 Effect of Grad-AM++

Table 4 shows the advantages of Grad-CAM++ over Grad-CAM and CAM. In the training of CAM, we add a GAP layer to the last feature layer, and all the previous layers are frozen for retraining. The effect of CAM

Table 4 The following table is the segmentation result evaluation. Including Decaf + MLR and SA-CAM and proposed method.

	PA	IOU	Precision	Recall	F-score
Decaf+MLR					
dirt	53.24	38.11	58.54	44.96	0.509
scratch	60.07	48.15	66.63	71.32	0.689
SA-CAM					
dirt	65.21	55.54	60.72	39.88	0.481
scratch	81.55	71.83	72.51	68.29	0.703
IRA-CNN + CAM					
Dirt	48.31	39.27	52.24	42.32	0.468
Scratch	62.98	49.08	65.42	55.47	0.6
IRA-CNN + Grad-CAM					
dirt	63.28	58.26	62.15	45.02	0.522
scratch	82.32	69.11	73.34	70.04	0.717
IRA-CNN + Grad-CAM++					
dirt	69.32	62.9	67.24	58.25	0.624
scratch	85.01	73.2	75.25	76.2	0.757

Table 5 Training time and FPS of IRA-CNN and other method.

Methods	Training time(s)	FPS
IRA-CNN	6432	18.22
RWSLDC	5323	19.88
HOG-SVM	9423	2.06
MLBP+SVM	7325	3.24

Table 6 Classification accuracy of different attention modules in IRA-CNN.

Method	attention	accuracy
IRA-CNN	Without attention	92.21
IRA-CNN	IAM(without spatial attention)	96.18
IRA-CNN	CBAM	96.24
IRA-CNN	IAM	97.71

is the worst, because adding a gap layer to the last layer may lose some semantic information and the classification accuracy. This can be shown in Table 4, and the Precision of CAM is greatly reduced. The effect of Grad-CAM is better than CAM, but as shown in Fig.12, There are many dirt defects in the picture, and Grad-CAM does not cover all of them. Therefore, Grad-CAM++ has better robustness and adaptability in the task of segmentation of injector valve, especially in the task of dirty defect segmentation.

4 Conclusion

In this paper, an end-to-end weakly supervised learning framework is proposed to classify and segment defects. This framework solves the problem of cylindrical metal surface defect classification and segmentation based on the data set of injector valve coren the task of classification, IAM is proposed to s suppress the interference of useless background area and highlight the defect area. In addition, IRA-CNN is designed and integrated with IAM to improve the classification accuracy by more than 1.5% compared with other methods. In the task of segmentation. Using Grad-CAM++ to generate saliency map. The defect pixel level segmentation based on saliency map greatly simplifies the pixel level segmentation task of image segmentation. The segmentation precision is at least 4.6% higher than other models. the segmentation precision is improved and the labeling time is saved. This weakly supervised learning framework only needs image level annotation. The cost of manual annotation is reduced.

There are some limits in our framework. For example, this framework can't classify and segment unknown defects. This is a common problem in industrial environment. In the future, we will study how to integrate unclassified defects into classification networks and realize segmentation and continue to optimize our IRA-CNN, and try to use different convolutional kernel forms to improve classification accuracy.

Acknowledgements The authors would like to express their appreciation to the developers of the Pytorch and the developers of Grad-CAM and Grad-CAM++.

Declarations

funding

This work was supported in part by National Natural Science Foundation of China (No. 51805312); in part by Shanghai Sailing Program (No.18YF1409400).

Ethical approval

No ethical approval was required for this research

Consent to participate

Not applicable.

Consent for publication

All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests

References

1. Kechen Song, Yunhui Yan.(2013) A noise robust method based on completed local binary patterns for hot-rolled steel strip surface. Applied Surface Science 285:858-864
2. Apostolos, et al.(2016) Feature selection for surface defect classification of extruded aluminum profiles. Int J Adv Manuf Technol 83:33-41
3. Shumin D, Zhoufeng L, Chunlei L.(2011) AdaBoost learning for fabric defect detection based on HOG and SVM. In:2011 IEEE International conference on multimedia technology,pp 2903-2906
4. Kwon BK, Won JS. Kang DJ.(2015) Fast defect detection for various types of surfaces using random forest with VOV features. Int J Precis Eng Manuf 16:965–970
5. Han, Hui, et al.(2020) Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks. Pattern Recognition Letters 130: 234-241
6. Augustauskas R, Lipnickas A.(2020) Improved Pixel-Level Pavement-Defect Segmentation Using a Deep Autoencoder. Sensors 20(9): 2557
7. Baumgartl H, Tomas J, Buettner R, et al.(2020) A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring. Progress in Additive Manufacturing 1-9.
8. Xu X, Zheng H, Guo Z, et al.(2019) SDD-CNN: Small data-driven convolution neural networks for subtle roller defect inspection. Applied Sciences 9(7): 1364
9. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z.(2016) Rethinking the inception architecture for computer vision. In:Proceedings of the IEEE conference on computer vision and pattern recognition,pp 2818-2826
10. Chen H, Pang Y, Hu Q, et al.(2020) Solar cell surface defect inspection based on multispectral convolutional neural network. Journal of Intelligent Manufacturing 31(2): 453-468
11. Cheon S, Lee H, Kim C O, et al.(2019) Convolutional neural network for wafer surface defect classification and the detection of unknown defect class. IEEE Transactions on Semiconductor Manufacturing 32(2): 163-170
12. He Y, Song K, Dong H, et al.(2019) Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Optics and Lasers in Engineering 122: 294-302
13. Li J, Su Z, Geng J, et al.(2018) Real-time detection of steel strip surface defects based on improved yolo detection network. IFAC-PapersOnLine 51(21): 76-81
14. Redmon J, Farhadi A.(2018) Yolov3: An incremental improvement. arXiv:1804.02767
15. Su B, Chen H, Zhou Z.(2020) BAF-Detector: An Efficient CNN-Based Detector for Photovoltaic Solar Cell Defect Detection. arXiv:2012.10631, 2020
16. Ren S, He K, Girshick R, et al.(2015) Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv:1506.01497
17. Tabernik Domen, et al.(2020) Segmentation-based deep-learning approach for surface-defect detection. Journal of Intelligent Manufacturing 31(3): 759-776
18. Chen L C, Zhu Y, Papandreou G, et al.(2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV),pp 801-818
19. Ronneberger O, Fischer P, Brox T.(2015) U-net: Convolutional networks for biomedical image segmentation. In:International Conference on Medical image computing and computer-assisted intervention,pp 234-241
20. Tao X, Zhang D, Ma W, et al.(2018) Automatic metallic surface defect detection and recognition with convolutional neural networks. Applied Sciences 8(9): 1575
21. Wang M, Cheng J C P.(2020) A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. Computer-Aided Civil and Infrastructure Engineering 35(2): 162-177
22. Zhou B, Khosla A, Lapedriza A, et al.(2016) Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition,pp 2921-2929
23. Lin H, Li B, Wang X, et al.(2019) Automated defect inspection of LED chip using deep convolutional neural network. Journal of Intelligent Manufacturing 30(6): 2525-2534
24. Chen H, Hu Q, Zhai B, et al.(2020) A robust weakly supervised learning of deep Conv-Nets for surface defect inspection. Neural Computing and Applications 1-16
25. Xu L, Lv S, Deng Y, et al.(2020) A weakly supervised surface defect detection based on convolutional neural network. IEEE Access 8: 42285-42296
26. Selvaraju R R, Cogswell M, Das A, et al.(2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp 618-626
27. Chattopadhay A, Sarkar A, Howlader P, et al.(2018) Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In:2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp 839-847
28. Hu J, Shen L, Sun G.(2018) Squeeze-and-excitation networks.In:Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132-7141
29. Ji Y, Zhang H, Wu Q M J.(2018) Salient object detection via multi-scale attention CNN. Neurocomputing 322:130-140
30. Woo S, Park J, Lee J Y, et al.(2018) CBAM: Convolutional block attention module. In:Proceedings of the European conference on computer vision (ECCV), pp 3-19
31. Tan M, Le Q.(2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In:International Conference on Machine Learning, pp 6105-6114
32. Sandler M, Howard A, Zhu M, et al.(2018) Mobilenetv2: Inverted residuals and linear bottlenecks. In:Proceedings of the IEEE conference on computer vision and pattern recognitionm, pp 4510-4520
33. Lin M, Chen Q, Yan S.(2013) Network in network. arXiv:1312.4400

34. Otsu N.(1979) A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9(1): 62-66
35. Simonyan K, Zisserman A.(2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
36. Ren R, Hung T, Tan K C.(2017) A generic deep-learning-based approach for automated surface inspection. IEEE transactions on cybernetics 48(3): 929-940

Figures

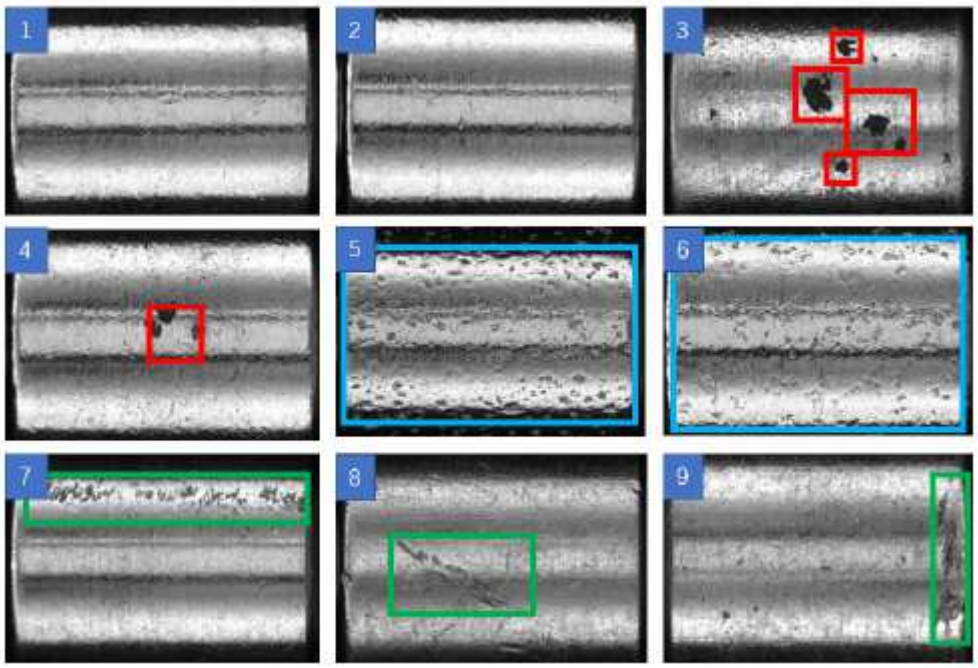

Figure 1

Three kinds of defects in injector valve: Non-Defect:1,2. Dirt:3,4. EC:5,6. Scratch :7,8,9. It can be seen EC covers the whole valve area, so it does not need to be segmented.

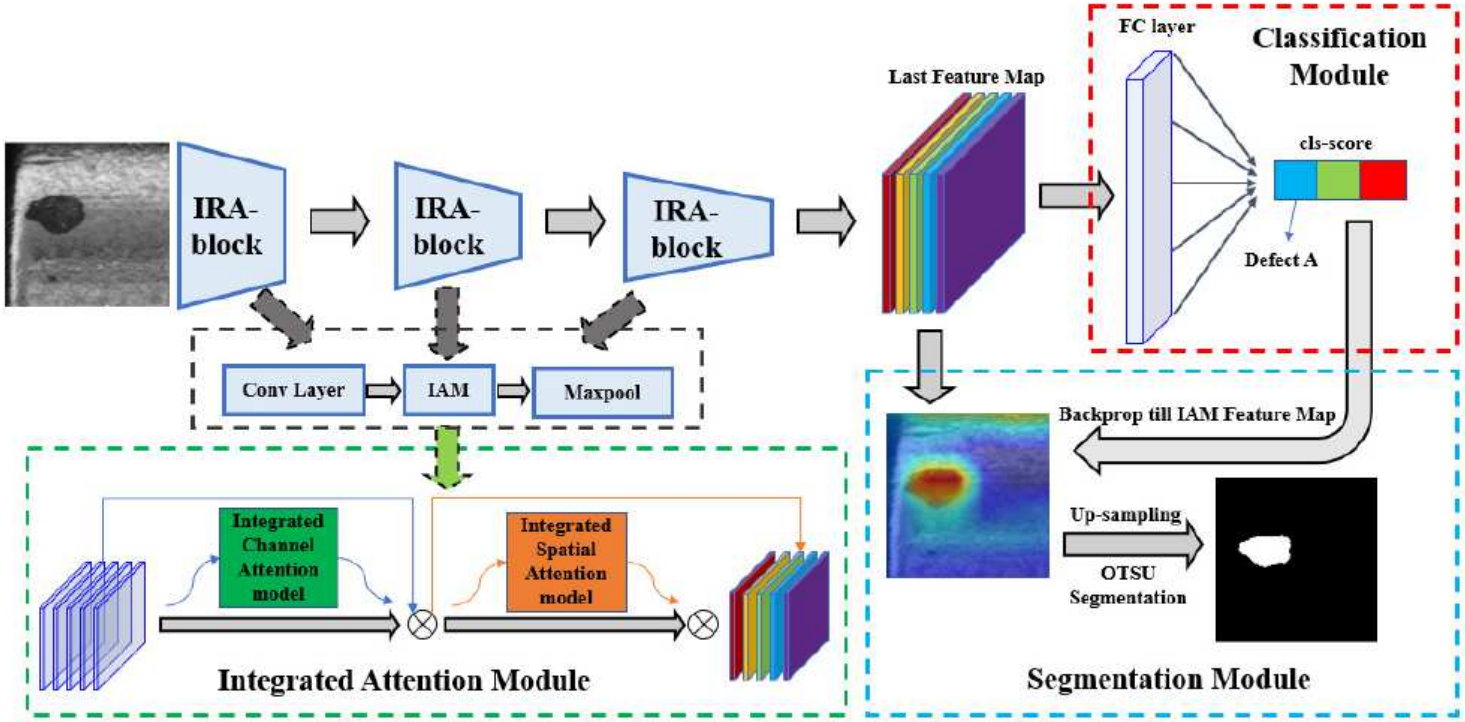

Figure 2

Framework structure of cylindrical metal workpiece surface defect inspection based on weakly supervised learning.

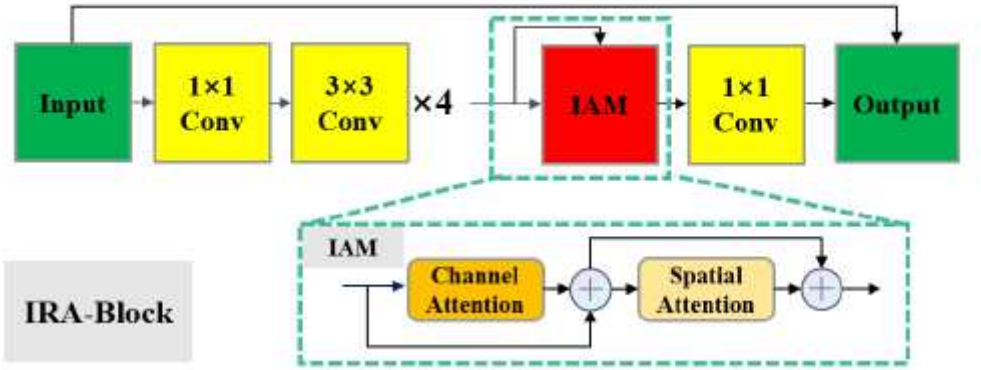

Figure 3

The structure of IRA-Block, Including IAM and fea- ture extraction.

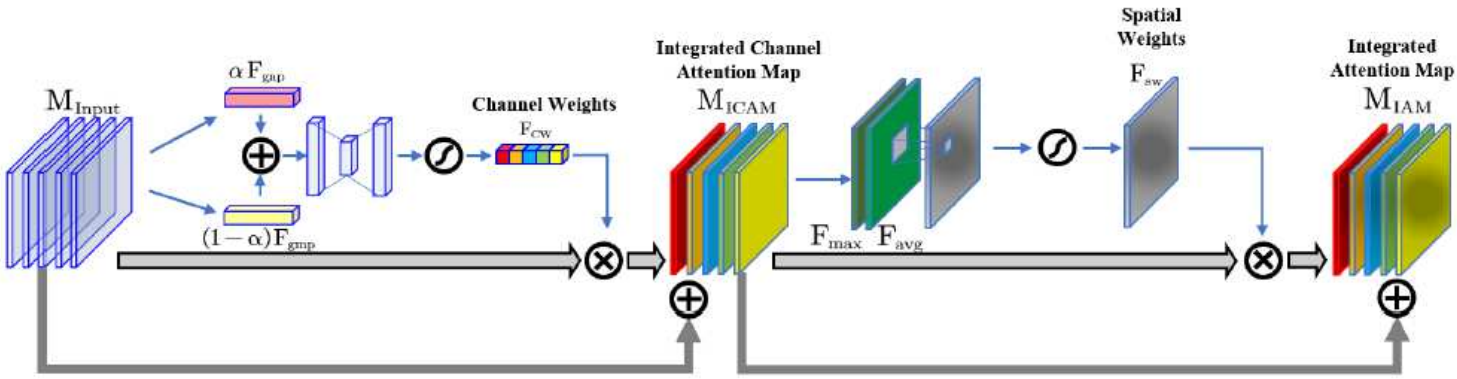

Figure 4

Frame structure of IAM, IAM includes channel attention submodule and spatial attention submodule.

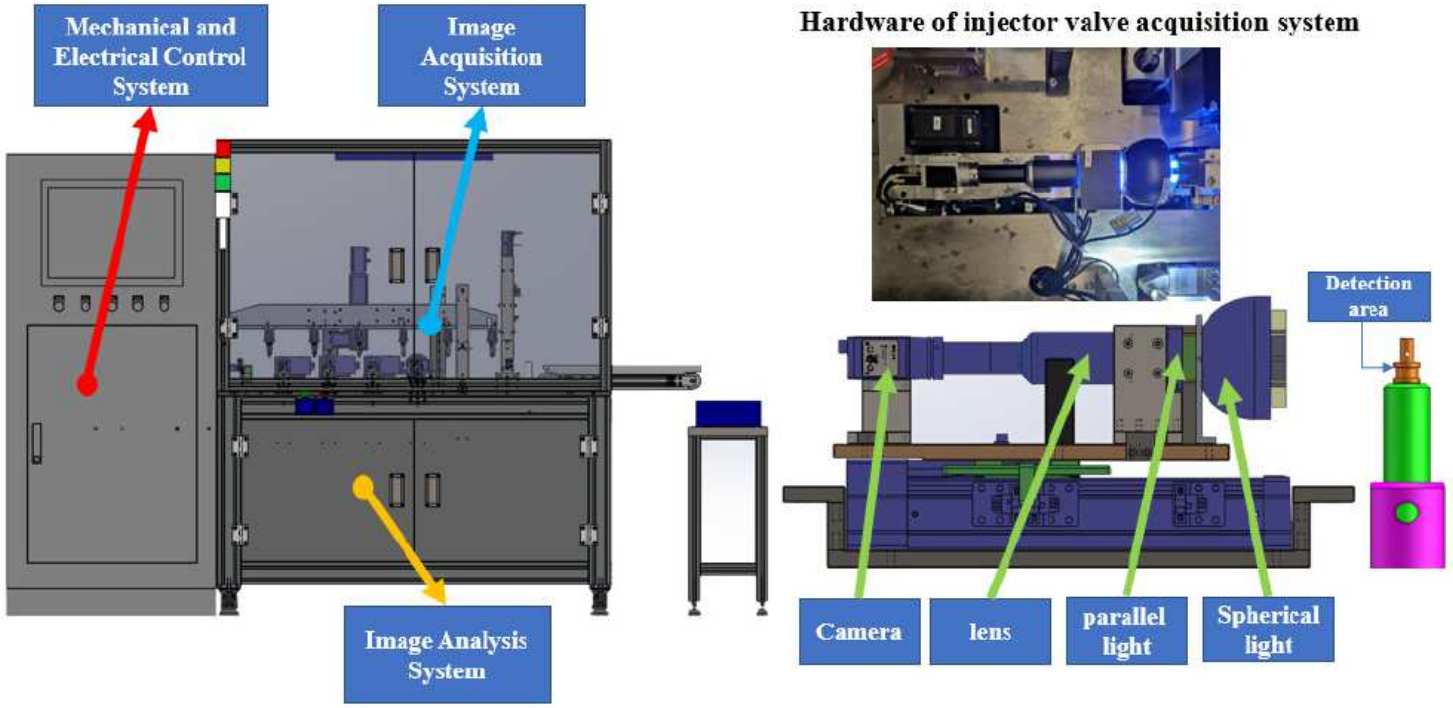

Figure 5

Injector valve machine vision system, Including Mechanical and Electrical System, Image Acquisition System. Image Analysis System. The right side of figure is the hardware of acquisition system.

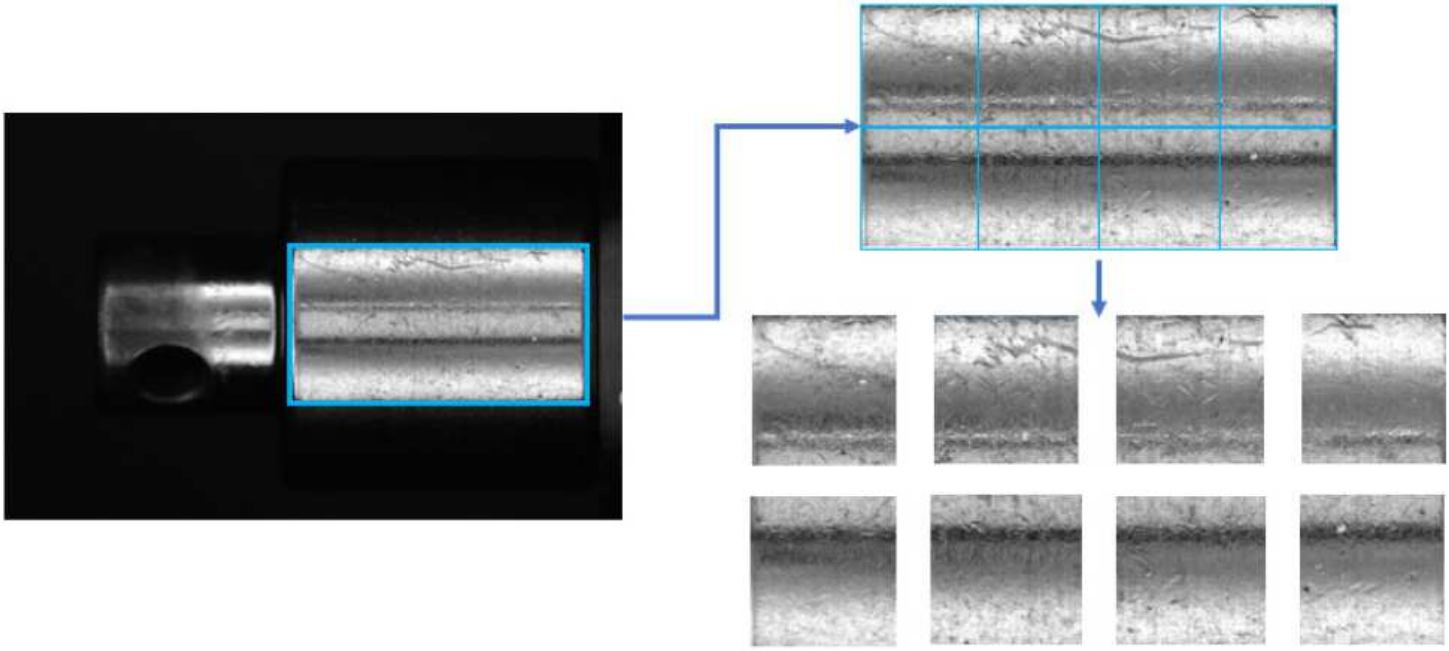

Figure 6

Dataset building method: each image collected by camera image acquisition is divided into 8 images. These 8 images are input into the dataset.

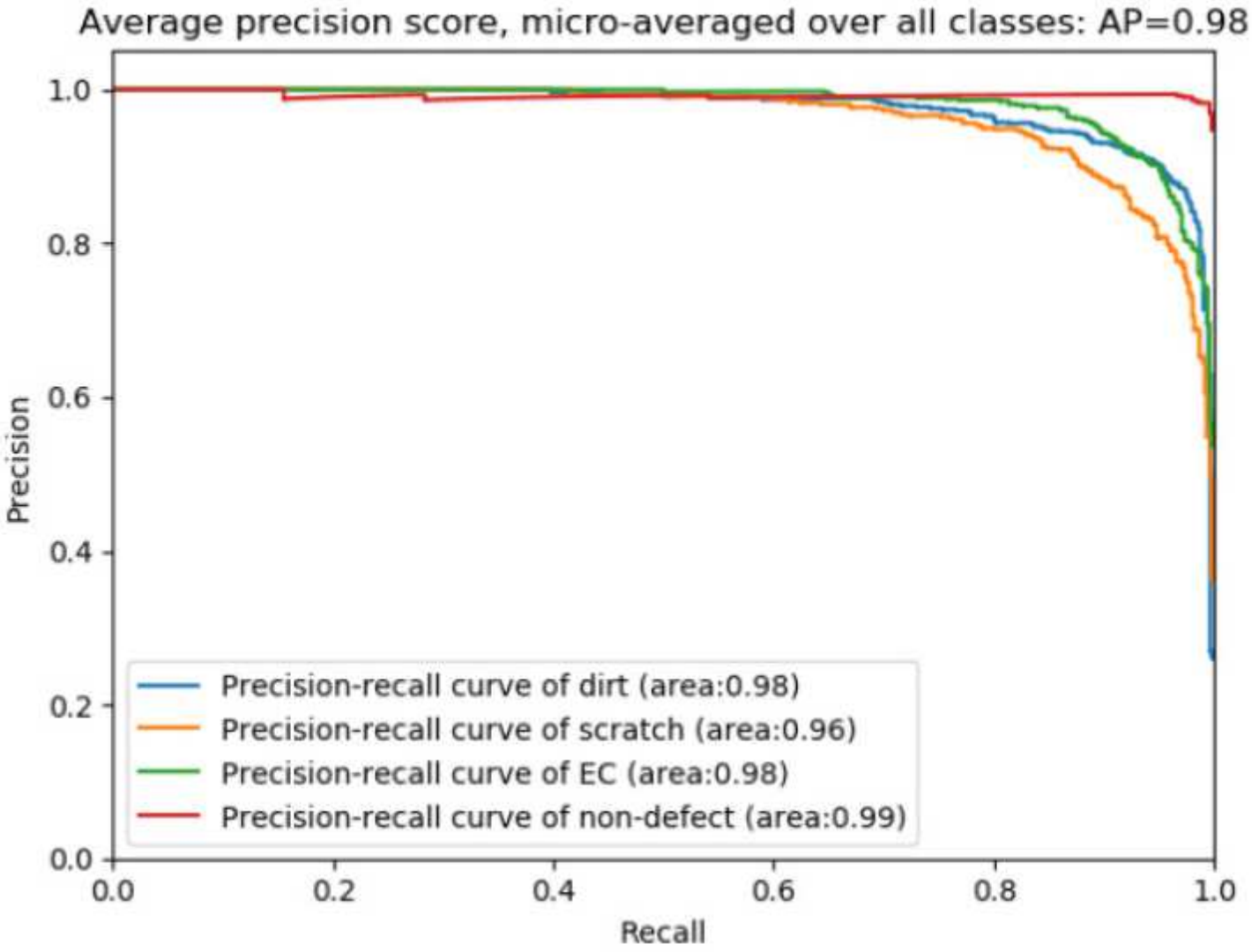

Figure 7

Classification results of ten methods on the dataset of surface defects of injector valve.

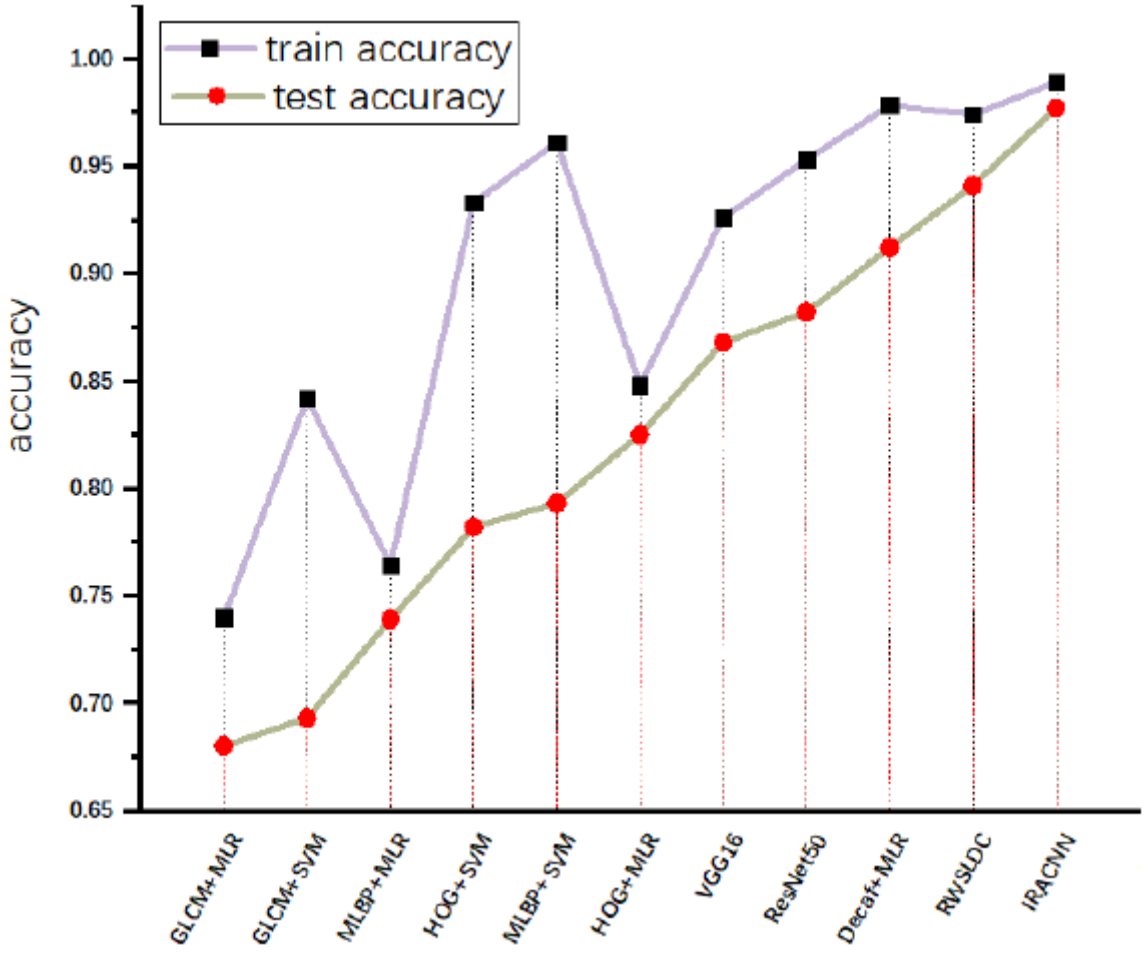

Figure 8

PR curve of IRA-CNN.

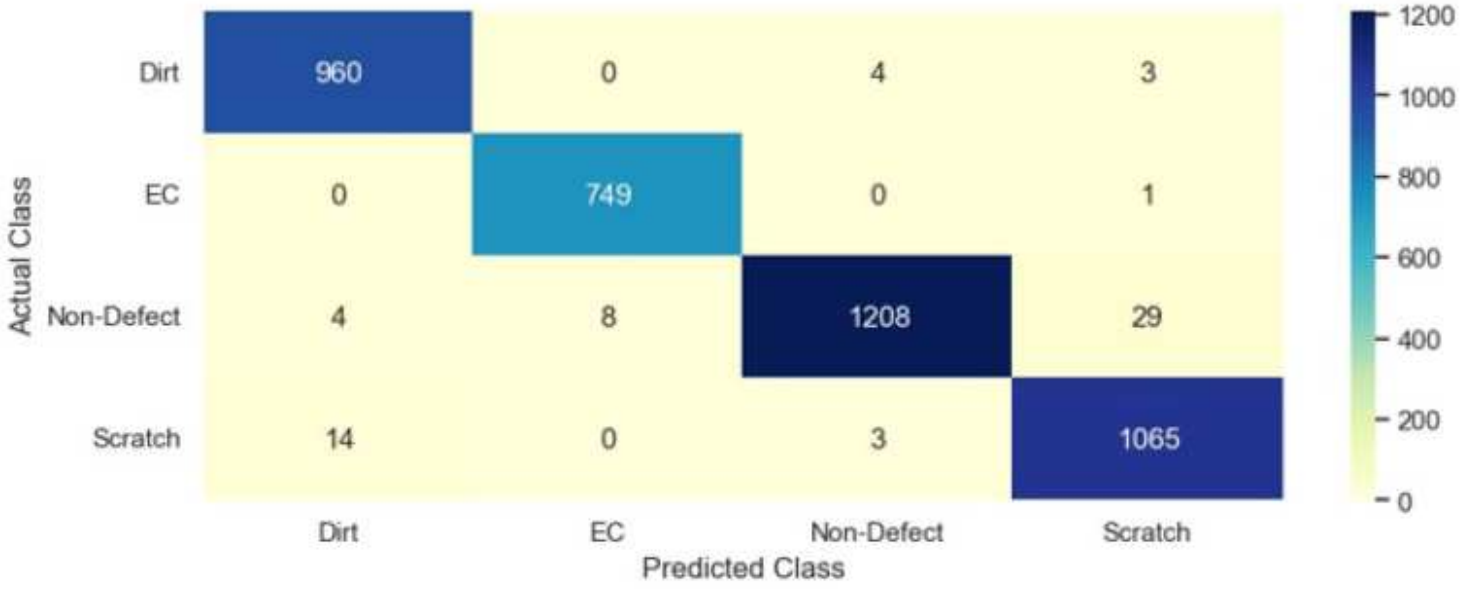

Figure 9

The confusion matrix of IRA-CNN.

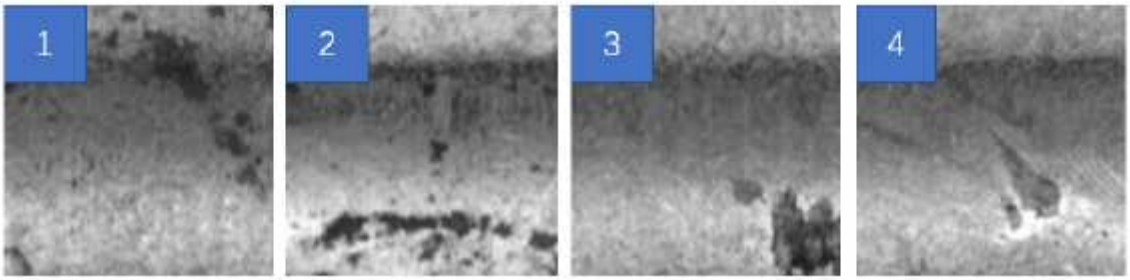

Figure 10

Dirt surface labeled 1 and 2 have similar characteristics to scratch surface. Scratch surface labeled 3 and 4 have similar shape to dirt surface.

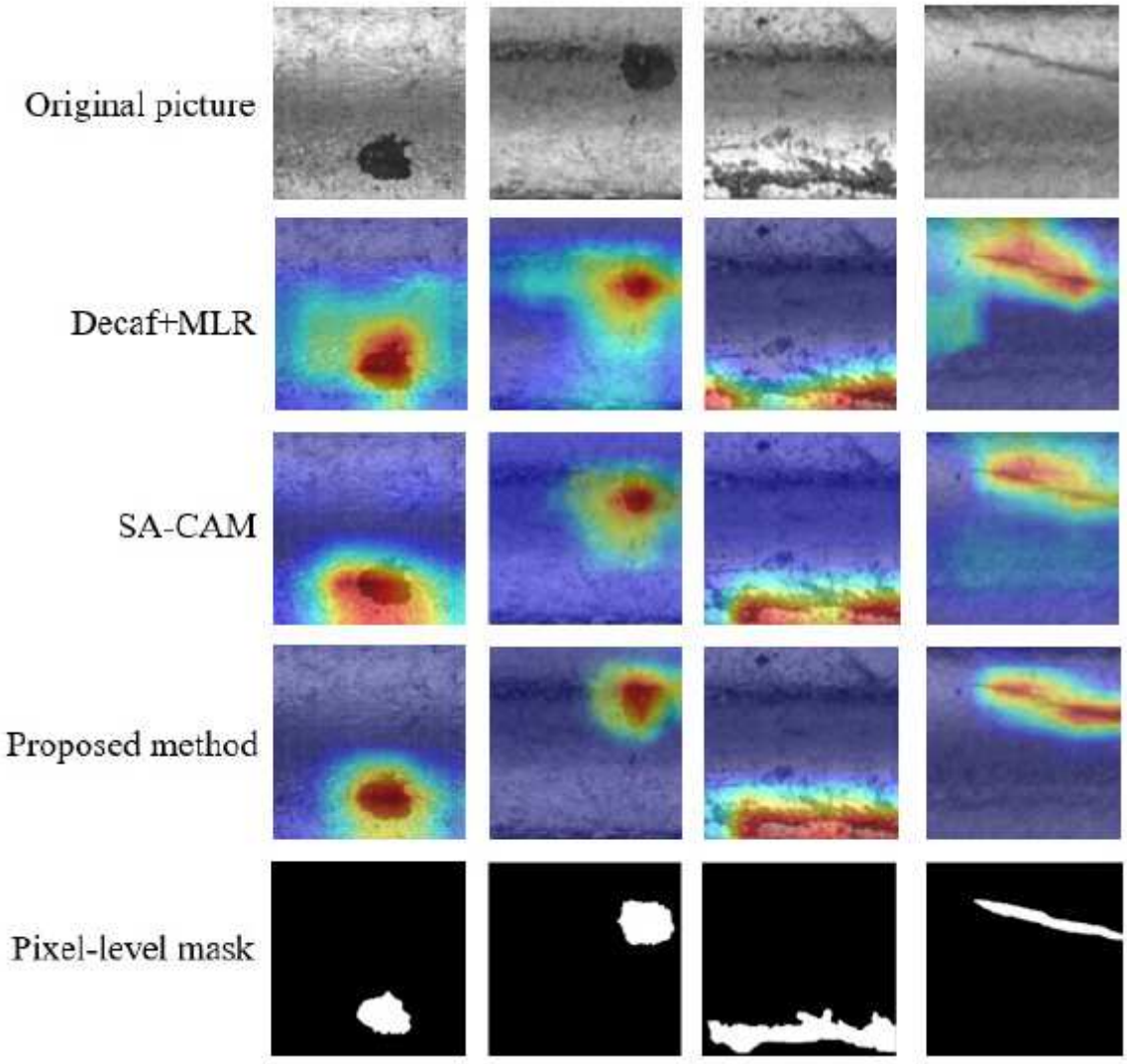

Figure 11

Comparison of the proposed method with Decaf + MLR and SA-CAM, the last row shows Pixel-level annotation mask.

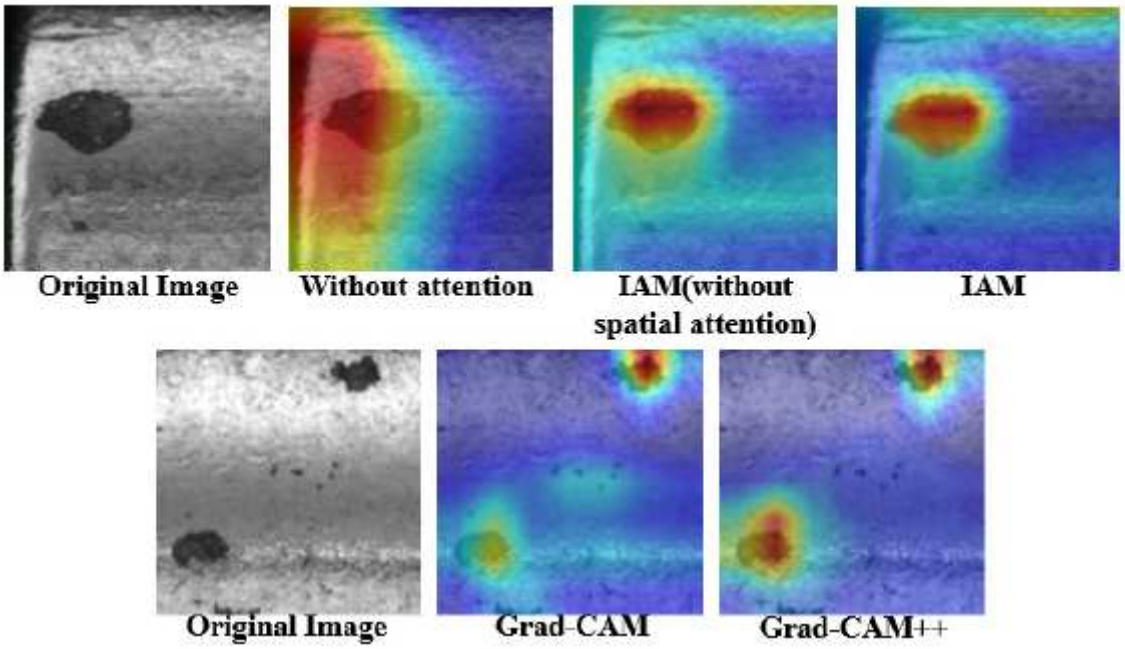

Figure 12

The above row is the ablation experiment of atten- tion module, and the next row is the comparison between Grad-CAM and Grad-CAM++.