IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 April 2024, accepted 25 April 2024, date of publication 29 April 2024, date of current version 3 June 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3395119*

RESEARCH ARTICLE

# A Visual Inspection and Classification Method for Camshaft Surface Defects Based on Defect Similarity Measurement

**JINSONG CAO[1], HANYANG WU[1], WEIYONG WANG[2], TEHREEM QASIM[3], AND DONGYUN WANG[1]**

[1]Key Laboratory of Urban Rail Transit Intelligent Operation and Maintenance Technology and Equipment of Zhejiang Province, College of Engineering, Zhejiang Normal University, Jinhua 321004, China
[2]Zhejiang Jinfei Machinery Company Ltd., Zhejiang 321004, China
[3]Department of Robotics and Artificial Intelligence, Islamabad 44000, Pakistan

Corresponding authors: Hanyang Wu (wuhanyang@zjnu.edu.cn) and Dongyun Wang (zsdwdy@zjnu.edu.cn)

This work was supported in part by the Key Research and Development Program of Zhejiang Province under Grant 2022C01139, and in part by the Key Research and Development Program of Jinhua.

**ABSTRACT** Industrial parts are usually shaped with complex surface structures and they require high surface quality. The surface defects influence the precision and performance of them. The efficient optical inspection equipment and defect detection and classification method are important for improving t productivity. There are two main difficulties in the visual detection of surface defects of industrial parts. One is difficult to develop visual inspection equipment for complex surfaces. Second is that the optimal feature set extracted by the commonly used feature fusion classification methods is not sensitive to defects with high defect similarity, which affects the defect classification accuracy. To solve difficulty 1, We have researched the principles of refraction and reflection when a light beam reaches the interface of a non-homogeneous medium, and develop a high-performance optical inspection equipment, which is available for 360° image acquisition of the external surface of an industrial part. To solve difficulty 2, a feature fusion classification method based on defect similarity measurement is proposed. The main idea is to measure the similarity between various types of defects by calculating the Euclidean distance between class centers and to classify defects with similarity higher than a threshold into one category. Subsequently, a two-step feature extraction and classification strategy is adopted. In the first step, the lower similarity types are separated, and then in the second step, the higher similarity types are separated. Deep learning methods and traditional machine learning methods are used for validation, respectively. Experimental results indicate that the proposed method achieves a 5% improvement in overall classification accuracy, and also has significant improvements in precision, recall and many other indicators compared to several other advanced classification methods.

**INDEX TERMS** Complex surfaces, defect classification, feature set, multi-feature fusion, similarity, two-step feature extraction and classification.

## I. INTRODUCTION

With the improvement of imaging technology and computer performance, computer vision techniques are widely used in various scenarios, such as camera calibration [1], [2], face recognition [3], target detection [4], etc. Recognition of defect types in the products used in the industry is of great significance for improving the production processes. A large number of scholars have already applied vision inspection technology in practical engineering, e.g. rail surface [5], electronic components [6], metal pipes [7] etc.

For the research on the surface optical properties of industrial products and optical inspection equipment, many scholars have conducted relevant studies. For example,

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

Chang et al. [8] developed an optical inspection platform combining parallel image processing with high-resolution opto-mechanical modules for defect detection in touchscreen glass. Dongyun et al. [9]developed an optical device for detecting cracks in metal wheels. Yinchao et al. [10] developed an automated optical inspection machine for defect detection of glass micro-optical elements. Wang et al. [11] investigated a non-contact optical automatic inspection system to identify defects in manufactured paper cups. Lee and Kim [12] developed an advanced 3D stacked visual placement inspection system for high-precision positioning of electronic packages. Comprehensive research of many scholars, it can be found that the following two conclusions: (1): Non-contact optical inspection equipment in the industrial product surface defect detection is increasingly widely used, is a field worthy of in-depth research. (2): for different detection of objects, generally need to design special optical inspection equipment, many of the current and the equipment structure is relatively simple, it is difficult to adapt to the shape of the complex structure of the industrial products of the automated detection.

In the research of defect detection classification, a fusion of multiple types of image-based features can lead to better accuracy of classification algorithms [13]. The key issue in a multi-feature fusion classification task is how to achieve better feature fusion. With different implementation methods, it can be divided into traditional machine learning methods and deep learning methods.

Some representative traditional machine learning methods are as follows. The traditional feature fusion method is relatively simple. One way is to concatenate all features into one feature. However, this method has several shortcomings including the ''curse of dimensionality'' and information redundancy. The local linear embedding-based method (LLE) can be employed to reduce the features' dimensions. However, this method ignored the metric of the original structure [14]. Local canonical correlation analysis can maximize the relevance of local neighbors [15]. Canonical correlation analysis (CCA) and its improvements maximize the correlation of the multi-feature sets and enhance the classification accuracy through the correlation of different feature sets [16]. However, it does not process the interrelationships in the same feature sets. Autoencoder (AE) is a class of neural network models used for dimensionality reduction, feature extraction and reconstruction of data. However, AE is easily overfitted with small amounts of data [17]. Multiview spectral embedding (MSE) can encode various features using different approaches to achieve a physically meaningful embedding. Ren et al. [18] proposed a tensor-based multi-view embedding algorithm to fuse features from different spaces. Their method can obtain unique features for subsequent classification. However, MSE does not mine the intrinsic structure of multiple features. Qin et al. [19] proposed the Semi-supervised Structured Subspace Learning algorithm for clustering data points from multiple sources (SSSL-M). Their method regularizes multiple view-specific affinity matrices into a shared affinity matrix based on reconstruction through a unified framework consisting of backward encoding networks and self-expressive mapping. The shared affinity matrix is comprehensive and can flexibly encode complementary information from multiple view-specific affinity matrices. Pairwise constraint-based multi-view subspace learning (PC-MSL) [20] considers both outer and inter-geometric structures, thus it maintains the discriminatory properties effectively. Manifold learning is used to reduce the number of features in high-dimensional data while preserving its topological structure [21]. However, it does not involve different metrics and structure fusion. If feature structure fusion is performed directly, a single metric cannot be adapted for the structure measurement of various features [22]. The above discussed methods could achieve multi-feature fusion classification by determining the optimal feature set. However, in the majority of defect classification tasks, the similarity between some types of defects is high. In these situations, the optimal feature set by single extraction often ignores some weak details which leads to its insensitivity to these high similarity defects, rendering the classification results unsatisfactory.

Compared to traditional defect detection methods, deep learning-based defect detection methods overcome the drawbacks of low accuracy and manual features that require extensive expert knowledge, providing an end-to-end solution for industrial vision inspection. As a representative of the two-stage detection approach, Faster RCNN introduces a region proposal network (RPN), which makes region proposal almost costless [23]. However, Faster RCNN performing downstream tasks on the last layer of features often leads to feature underutilization and decay of texture information [24]. In terms of feature enhancement, the feature pyramid network (FPN) realizes multi-scale information fusion by simple but effective lateral addition. On this basis, Liu et al. [25] proposed the balanced feature pyramid module, which makes full use of the information flow between multiple layers and improves the accuracy of steel plate defect detection. To further enhance the ability to recognize the appearance of defects between different classes. Although the feature pyramid network is capable of fusing multi-scale feature information, it can only be trained for a particular resolution. The attention mechanism can be regarded as a module for refining feature information. It enhances meaningful features and suppresses unnecessary features. Yeung and Lam [26]. propose a fused-attention network (FANet) for detecting various steel surface defects. It applies an attention mechanism to a single balanced feature map, rather than multiple feature maps. This can improve the accuracy and preserve the detection speed of the detection network. Qilong et al. [27] designed an efficient channel attention (ECA) module to perform channel attention with low model complexity. Sanghyun et al. [28]constructed a convolutional block attention module (CBAM) to perform

channel attention and spatial attention. However, the attention mechanism can only process the features transmitted from the backbone network, so it is especially important to choose what kind of features to give to the attention enhancement module. In Transformer, the traditional CNN and RNN are abandon, and the whole network structure is entirely composed of Attention mechanism, which was firstly applied in the field of natural language processing. Transformer's morphing networks Vision Transformer and Swin-Transformer have a wide range of applications in visual detection tasks Wang et al. [29] propose a novel ViT (BuildFormer), with a dual-path structure. In this method, they designed a spatial-detailed context path to encode rich spatial details and a global context path to capture global dependencies. Besides, they designed a window-based linear multihead self-attention whose complexity is linear with the window size. Cheng et al. [30] proposed an end-to-end detection network based on Swin-Transformer, called SwinCrack, to solve the problem of limited receptive field in pure CNN-based crack detection networks. Compared to CNN-based detection models, SwinCrack can produce more accurate and continuous descriptions of pavement cracks by modeling long-range interactions and adaptive spatial aggregation. The above research provides very useful guidance for improving our deep learning network models.

In order to solve the problem of acquiring images of the outer surface of complex workpieces, we researched the transmission characteristics of light particles in non-uniform media, and developed an automated optical inspection equipment that can acquire images of the outer surface of workpieces in 360°. For the defect detection classification problem, a feature fusion classification method based on defect similarity measurement (FFCM-DSM) is proposed. It can calculate the Euclidean distance between the class centers of each type of defects, so the similarity between each type of defects can then be measured. Subsequently, those types of defects which over the set threshold are classified into one category. In the end, a two-state feature extraction and classification strategy is adopted. The surface defect dataset of camshafts was collected using our developed vision inspection equipment, and defect classification experiments were conducted using FFCM-DSM. Experimental results indicate that the proposed method performs with high accuracy.

The contributions and innovations of this paper mainly include the following two points: (1): the refraction and reflection characteristics of light on the camshaft surface were theoretically researched, and based on this, a set of automated optical inspection equipment was designed and developed, which can satisfy the inspection of camshaft surface defects under the actual status working conditions. (2): the classification of camshaft surface defects was researched, and the FFCM-DSM method was proposed, which improves the accuracy of the classification of the defects.

The rest of the paper is organized as follows, in Section II, the principle of optical reflection and refraction on the camshaft surface and the optical inspection equipment are introduced. In Section III, the proposed defect classification method FFSM-DSM is explained in detail. In Section IV, the experimental results and discussion are given. In Section V, conclusions are drawn.

## II. OPTICAL REFLECTION AND REFRACTION PRINCIPLES AND INSPECTION EQUIPMENT

### A. A NON-UNIFORM MEDIUM SUB-TRANSFER FUNCTION CONSTRUCTION

The radiative transfer factor describes the proportion of the total radiant energy in an optical system that is emitted from a surface Si or a body Vi that is ultimately absorbed by a statistical unit j after being reflected and refracted by a single unit within the system or multiple times, and is defined as shown in (1) [31]. Refraction and reflection occur when a light beam reaches the interface of a non-uniform medium, and an interface model is constructed using Monte Carlo method to describe the redistribution ratio of the radiant energy of the light field at the boundary. The reflectivity of unpolarized light at the boundary is calculated according to Snell's law of refraction for angles of incidence, reflection and refraction of $\alpha$, $\beta$ and $\gamma$, respectively, as shown in (2) [31]. According to the classical electromagnetic field theory, to solve the incident light field through the refractive index n1 and n2, respectively, after the redistribution of the surface of the medium after the redistribution of energy, the energy density of its incident and refracted light, respectively, as shown in (3), (4) [32].

$$RD_{ij} = \frac{N_{i,j}}{N_i} \tag{1}$$

$$\rho(\alpha) = \frac{\rho_p(\alpha) + \rho_s(\alpha)}{2} = \frac{1}{2}\left[\frac{tan^2(\alpha-\gamma)}{tan^2(\alpha+\gamma)} + \frac{sin^2(\alpha-\gamma)}{sin^2(\alpha+\gamma)}\right] \tag{2}$$

$$\omega_f = \left(\frac{2cos\alpha}{cos\alpha + \sqrt{n_{21}^2 - sin^2\alpha}}\right)^2 n_{21}^2\omega_s + \left(\frac{2n_{21}^2 cos\alpha}{n_{21}^2 cos\alpha + \sqrt{n_{21}^2 - sin^2\alpha}}\right)^2 n_{21}^2\omega_p \tag{3}$$

$$\omega_t = \left(\frac{2cos\alpha}{cos\alpha + \sqrt{n_{21}^2 - sin^2\alpha}}\right)^2 \omega_s + \left(\frac{n_{21}^2 cos\alpha - \sqrt{n_{21}^2 - sin^2\alpha}}{n_{21}^2 cos\alpha + \sqrt{n_{21}^2 - sin^2\alpha}}\right)^2 \omega_p \tag{4}$$

### B. B 360° OPTICAL INSPECTION EQUIPMENT

The experimental equipment is a four-workstation rotary detection equipment. The overall setup is shown in Fig 1.

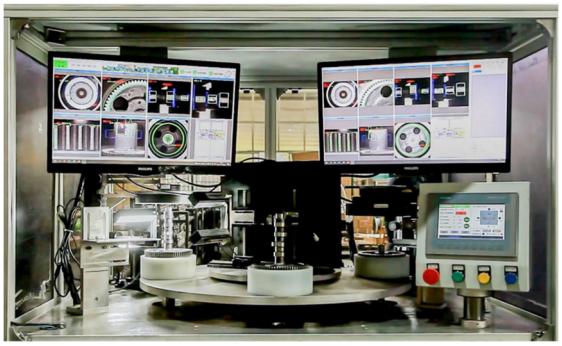

**FIGURE 1.** A snapshot of the equipment used in the experiments.

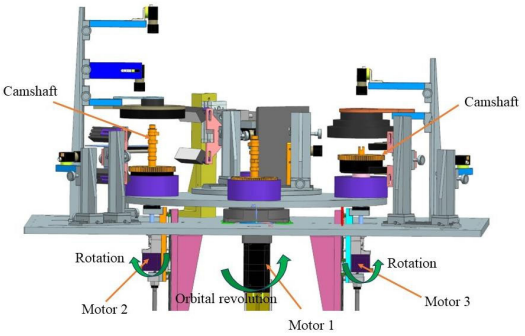


**FIGURE 2.** 3D model diagram of the optical inspection platform.

Servo-driven motors are equipped in the middle of the machine to drive the workstation in a male rotation. Workstation 1 is the loading and unloading station. Workstation 3 is the turning station. Workstation 2 and 4 are the camshaft front and reverse detecting stations, their bottom is equipped with a high-precision servo motor to drive the camshaft completing 360° rotation motion. 3D model diagram of the optical inspection platform is shown as Fig 2. Workstation 2' and Workstation 4' 3D model diagram is shown in Fig 3, cameras and light sources model are shown in Tab 1.

There are two main innovations in the above research. (1) Through the construction of non-uniform medium sub-transfer functions, the reflection and refraction characteristics of light of different wavelengths on surfaces with different properties of camshafts can be investigated, which provides theoretical guidance for the selection of adapting different types of light sources as well as the angle of incidence. (2) The designed 4-station inspection equipment is able to acquire all-round high-definition images of the camshaft outer surface.

## III. A FEATURE FUSION CLASSIFICATION METHOD BASED ON DEFECT SIMILARITY MEASUREMENT (FFCM-DSM)

As shown in Fig 4, there are 4 steps involved in the proposed FFCM-DSM.

Step 1: Extraction of multi-type features to construct a feature pool;

**TABLE 1.** The model for cameras and light sources on workstation 2 and workstation 4.

| Equipment | Model | Brands |
|---|---|---|
| Camera 1 | Aca 1300-30gm | basler |
| Camera 1 | Aca 2500-14gm | basler |
| Camera 1 | Aca 2040-35gm | basler |
| Camera 1 | Aca 2040-35gm | basler |
| Camera 1 | Aca 1300-60gm | basler |
| Camera 1 | Aca 3800-10gm | basler |
| Camera 1 | Aca 640-120gm | basler |
| Light source 1 | XWD-TS-135-32-W | KCS |
| Light source 2 | XWD-TS-135-32-W | KCS |
| Light source 3 | XWD-TS-135-32-W | KCS |
| Light source 4 | XWD-TS-135-32-W | KCS |
| Light source 5 | VLBHA30D214R2-120°-DF-24V | Vlight |
| Light source 6 | R-180-20 | KCS |
| Light source 7 | R-60-90 | KCS |
| Light source 8 | VLBHA30D214R2-120°-DF-24V | Vlight |
| Light source 9 | VLBHA30D214R2-120°-DF-24V | Vlight |
| Light source 10 | AOI180 | KCS |

Step 2: Calculation of the class centers of defective samples based on the normal distribution property;

Step 3: Similarity Calculation and Reclassification;

Step 4: Using the two-stage classification strategy for classification operations

Step 2, step 3 and step 4 (Method 1: Multilayer classification networks based on attention mechanisms and Method 2: PCA feature dimensionality reduction and SVM classification) will be explained in the following.

### A. STEP 2: CALCULATION OF THE CLASS CENTERS OF DEFECTIVE SAMPLES BASED ON THE NORMAL DISTRIBUTION PROPERTY

The long tail effect generally exists in industrial defect sample data sets, especially when the amount of data is small. The sample data distributed in the tail ends up influencing the statistical characteristics. Therefore, we propose an approach for calculating the cluster centers of defect samples based on the properties of a normal distribution. The steps involved in the implementation are as follows:

1: First, we normalize all feature data, then a point $u' = (u'_1, u'_2, \ldots u'_p, \ldots, u'_k)$ among all sample points of each type of defect is identified, which can minimize the sum of distances between it and all sample points, as shown in the following:

$$E = \sum_{p=1}^{k} \sum_{q=1}^{n} |x_p^q - u'_p| \tag{5}$$

where $E$ is the minimum sum of the distance between each data point obtained and the corresponding category center. $p$ represents the value of the $p$-$th$ feature, and $k$ is the number of features. $q$ represents the $q$-$th$ sample, and n is the number of defective samples in this category, so $x_p^q$ represents the value of the $p$-$th$ feature of the $q$-$th$ sample, $\mu'_p$ is computed

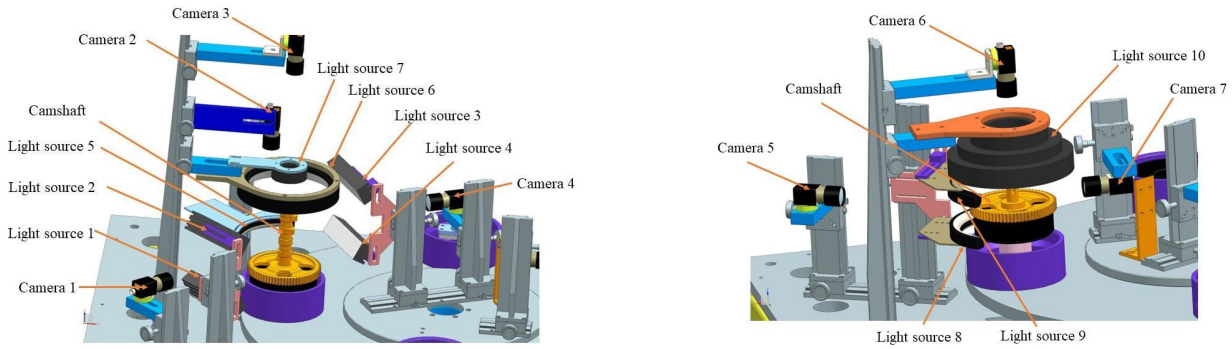


**FIGURE 3.** Cameras and light sources settings for Workstation 2 and Workstation 4.

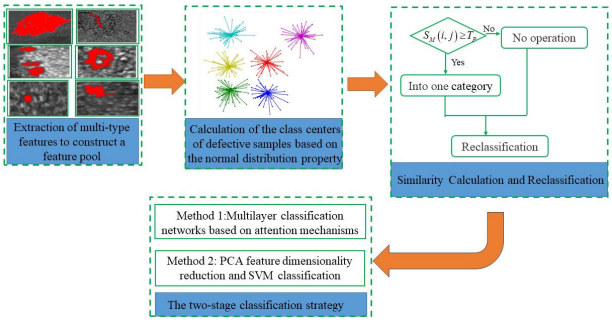


**FIGURE 4.** FFCM-DSM flow diagram.

as follows.

$$u_p^{'} = \frac{1}{n}\sum_{p=1}^{n} x_p^q \tag{6}$$

generally, the smaller the value of E, the higher the similarity of the samples in the class.

2: The distance $d_q$ between the *q-th* sample point and the pseudo-class center $u'$ is calculated as follows:

$$d_q = \sqrt{\sum_{q=1}^{k} \left(x_p^q - \mu_i'\right)^2} \tag{7}$$

constructing a normal distribution of $dq$ is defined in the following.

$$d_q \sim f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{8}$$

3: According to the normal distribution characteristics, the area enclosed by curves and coordinate axes accounts for 99.73% of the total area in the interval of $x \in (\mu - 3\sigma, \mu + 3\sigma)$, and the event that $x$ falls outside $(\mu - 3\sigma, \mu + 3\sigma)$ can be considered as an ''impossible event''. For the industrial defect data sets, it can be considered as an anomaly [33], so the sample can be intercepted by $d_q \in (\mu - 3\sigma, \mu + 3\sigma)$.

4: The above explained intercepted sample is reused to calculate the class center $u_i$ after the removal of the anomaly.

### B. B STEP 3: SIMILARITY CALCULATION AND RECLASSIFICATION

For each type of defect, the class center can be expressed as $\mu = (\mu_1, \mu_2, \ldots\ldots, \mu_k)$, which can be considered as a vector. The Euclidean distance between the class centers of the two types of defects can be measured as follows:

$$D_{ij} = \sqrt{\sum_{p=1}^{k} \left(\mu_p^i - \mu_p^j\right)^2} \tag{9}$$

where $i$ and $j$ represent two different defective types.

The next step is to determine whether a reclassification is necessary. If $D_{mim}$ is too small, then reclassification is not needed. When $D_{\max}/D_{\min} \geq 2$, reclassification is needed. For those situations that need to be reclassified, a new set $d_{ij}$ is obtained by normalizing the Euclidean distance $D_{ij}$. The similarity $S_M(i,j)$ between any two defects is computed as shown in (10). Subsequently, a reasonable threshold $T_P$ is set, which can be determined by the actual scenario. When $S_M(i,j)$ is greater than the threshold, then the defects need to be classified into a category as shown in (11).

$$S_M(i,j) = 1 - d_{ij} \tag{10}$$

$$\begin{cases} if\ S_M(i,j) \geq T_P & take\ i,j\ into\ one\ category, \\ if\ S_M(i,j) < T_P & no\ operation; \end{cases} \tag{11}$$

### C. THE TWO-STAGE CLASSIFICATION METHOD

#### 1) METHOD 1: MULTILAYER CLASSIFICATION NETWORKS BASED ON ATTENTION MECHANISMS

In the case where the data sample is adequate and the sample types are balanced, after the reclassification operation, to fully demonstrate the effectiveness of this strategy, in the classification stage, we propose a multilayer classification network that fuses the attention mechanism. This network is based on the backbone network of YOLO v5, and the

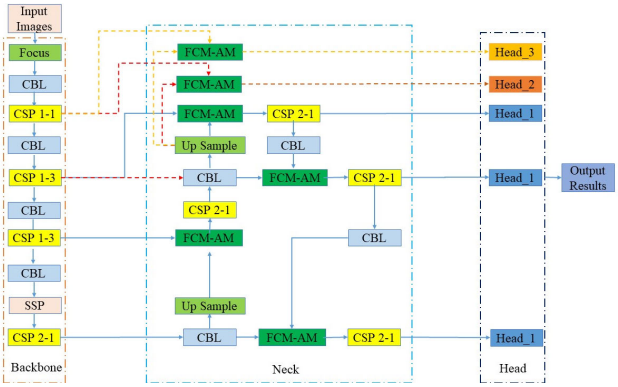


**FIGURE 5.** Structure of the improved YOLO v5s. The blue line is the original structure in the YOLO v5 model, the red and orange lines are the processes we added, the green module is the feature enhancement module based on attention mechanism, and Head_2, Head_3 are the added detection heads.

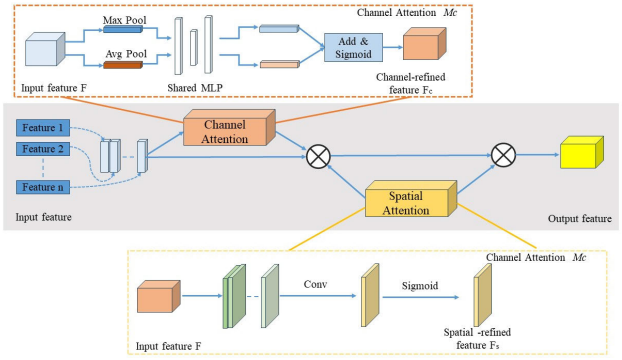


**FIGURE 6.** Structure of the FCM-AM module. The gray part in the middle is the structure of the whole FCM-AM, with the channel attention module on the top and the spatial attention module on the bottom.

feature connection module based on the attention mechanism is introduced in the feature fusion stage. Its structure is shown in the following Fig 5.

As shown in the Fig 5, this layered network, we add Head_2 and Head_3 to the original three Head_1. First, we analyze the effects of Head_1 on our detection task. Head_1 fuses the deep information extracted from the backbone network, which contains more semantic information (global features), so that it facilitates the detection of defects and the labeling of their locations. However, for holes and sand holes, rust spots and dirt, which are several defects with high similarity, discriminating them requires more detailed information (local features), and deep semantic information will result in overfitting. So here we are just using Head_1 to detect the location of defects and will judge polishing, cracking, C1 and C2 type defects. It is proposed that Head_2 and Head_3 only fuse local features at a shallower level, so we will use Head_2 to discriminate rust spots and hair chip defects from C1 and Head_3 to discriminate holes and sand eye defects from C2. Other details of the network remain the same as the traditional Yolo v5s model [28].

In the network, there exists an FCM-AM module which is a feature connection module based on the attention mechanism that we have introduced in the network, it has been proved to be beneficial in improving the performance of the network. The structure of the FCM-AM module is shown in Fig 6.

The introduction of an attention mechanism to weigh the features in the target region makes the feature extraction network selectively focus on the target region that contains important information while suppressing other irrelevant information. The significance of introducing a separate FCM-AM module before Head_2 and Head_3 is that it can enhance the discriminative features of rust spots and dander, holes and sand holes.

#### 2) METHOD 2: PCA FEATURE DIMENSIONALITY REDUCTION AND SVM CLASSIFICATION

PCA is a statistical analysis method that transforms multiple original variables into several new composite variables through a linear transformation, and it is widely used for dimensionality reduction of feature space. It maximizes the retention of information within the data after dimensionality reduction, and determines the direction by measuring the variance of the data in the projection direction [34].

SVM algorithm is widely used to solve image recognition problems, signal processing problems, risk identification problems, etc. [35]. It maps the training set to a high-dimensional feature space and applies a linear regression through a non-linear mapping function, the basic mathematical equations of SVM are shown in (12) and (13) [36].

$$\omega_i^T \times \Phi\left(X_{ijt}\right) + b_i, \ \forall i \tag{12}$$

$$X_{ijt} = \left\{\widetilde{u}_{i1t}, \widetilde{u}_{i2t}, \ldots, \widetilde{u}_{ijt}\right\}, \ \forall i, t \tag{13}$$

where $\omega_i^T$ is the weight vector of the hyperplane and $\Phi\left(X_{ijt}\right)$ is the bias term. According to the structured minimum principle, (12) is equivalent to minimizing the cost function, which is shown below:

$$\begin{cases} \min\left(\frac{1}{2}||\omega_i||^2 + \frac{1}{2}C_i\sum_{t=1}^{T_t}\xi_i^2\right) \\ s.t.y_{it} - \omega_i^T \cdot \psi\left(X_{ijt}\right) - b_i = \xi_i \end{cases}, \forall i \tag{14}$$

where $\xi_i$ is the relaxation variable, $C_i$ is the penalty factor and $\psi\left(X_{ijt}\right) = K\left(X_{ijt}, X'_{ijt}\right)$ is the kernel function. Commonly used kernel functions include linear kernel, polynomial kernel, Gaussian kernel, exponential kernel, Laplace kernel, sigmoid kernel, etc [37]. The application of PCA and SVM algorithms in a two-step classification strategy is shown in Fig 7.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

### A. PREPARATION FOR EXPERIMENTS

#### 1) DATA SETS AND LOSS WEIGHTS

As shown in Tab 2, according to the classification method proposed in this paper, 205 defect samples were acquired from the casting surface of the camshaft, including 42 samples of hole defect, 42 samples of polishing defect, 28 samples of rust spot defect s, 34 samples of sand hole defect,

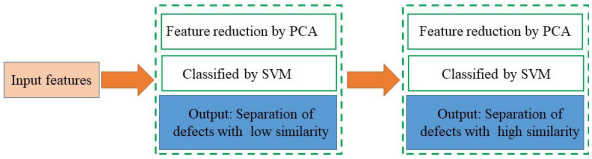


**FIGURE 7. Application of PCA and SVM algorithms in a two-step classification strategy.**

38 samples of dander defect. 21 samples of crack defects. Among these, cracks, holes and sand holes are irreparable defects, rust spots and polishing are repairable defects, and dander are pseudo defects. Their images are shown in Fig 8, the yellow box shows the defects.

For irreparable defects, the weight $W_{ij}$ of the cost of misclassifying as the repairable defects is 1.5, as pseudo defects is 2. For reparable defect, the weight $W_{ij}$ of the cost of misclassifying as pseudo-defect is 1.5, and the weight $W_{ij}$ of the cost of the remaining misclassifications is 1.

#### 2) EXTRACTION METHOD OF THE DEFECT AREA

The surface defects of the camshaft collected in this experiment can be divided into line defects, block defects and texture defects.

The Laplace Gaussian, also known as the second derivative of the Gaussian detects the boundary by finding the zero point of the second derivative in the gray value of an image [38]. It has a strong response value to the crack on the camshaft surface and the boundary characteristics of the matted area. Therefore, the linear defects on the camshaft surface can be detected by the Laplace operator based on scale normalization. The basic idea of the local dynamic threshold segmentation algorithm is to divide an image into several sub-images, and select each sub-image to determine the upper and lower bound thresholds according to the histogram peak segmentation point to achieve local region segmentation [39]. Therefore, the block defects can be extracted by the local dynamic threshold segmentation algorithm. A two-dimensional Garbor filter can extract local frequency features of spatial images for texture feature extraction [40]. Hence, we use a two-dimensional Garbor filter to extract the texture of the defect area. The result of the defective area extraction is shown in Fig 9.

### B. EXPERIMENTAL PROCESS

In the field of visual detection, geometric features, texture features and global features are widely used for the description of defective areas [41], [42]. Therefore, the feature pool for this experiment contains geometric features, texture features and global features. Details are shown in Tab 3. The texture features of the defective region are extracted using a grey-scale co-occurrence matrix [43]. The features contained in the feature pool are shown in Tab 3

As shown in Tab 4, the class centers of 6 defect types using 11 different features are calculated. Then the Euclidean distance $D_{ij}$ between the center of each type of defect is calculated as shown in Tab 5. The similarities $S_M(i, j)$ between the various types of defects are calculated as shown in Tab 6. In Tab 6, the highest similarity to the polishing defect is obtained for rust spot defect, with a value of 0.3704. Additionally, the highest similarity with cracking defects is obtained for dander defects, with a value of 0.5390. It indicates that the similarity between these two types of defects and other types of defects is relatively low. On the other hand, the similarity index between dander and rust spot defects is recorded as 1. This indicates a high degree of similarity between these two types of defects. Additionally, the similarity index between sand hole defects and hole defects is calculated as 0.8890. This indicates that there is also a strong similarity between sand hole defects and hole defects. Using the threshold calculation method discussed in Section II, we set the value of $T_P$ to 0.8. As such, the dander and rust spot defects are placed one category, noted as $C_1$, and the hole and sand hole defect are divided into another category, noted as $C_2$.

### C. COMPARISON AND ANALYSIS

In the assessment of multiclassification tasks, the four key performance metrics are accuracy, precision, recall rate, and F1 score, which are defined is shown below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{18}$$

where true positive ($TP$) means a positive sample was correctly predicted as positive, true negative ($TN$) means a negative sample was correctly predicted as negative, false positive ($FP$) means a negative sample was incorrectly predicted as positive and false negative. ($FN$) means a positive sample was incorrectly predicted as negative.

#### 1) COMPARISON OF DEEP LEARNING METHODS

To verify the superiority of our proposed improved YOLO v5 model in camshaft surface defect detection experiments, we compared the method with several other advanced defect detection methods, they are Faster RCNN Res2Net-101 [44], Vision Transformer (ViT) [45], Swin-Transformer [46], DDN [47], and YOLO v5s [48].

To meet the data amount requirements of deep learning methods, we expanded our dataset by using Cycle GAN, expanding the number of each of the 6 classes of defects to 300, dividing the training set and test set according to a ratio of 2:1, and using a 5-Fold method in the training phase. All comparison models are implemented using a PyTorch-based framework, trained and tested on NVIDIA

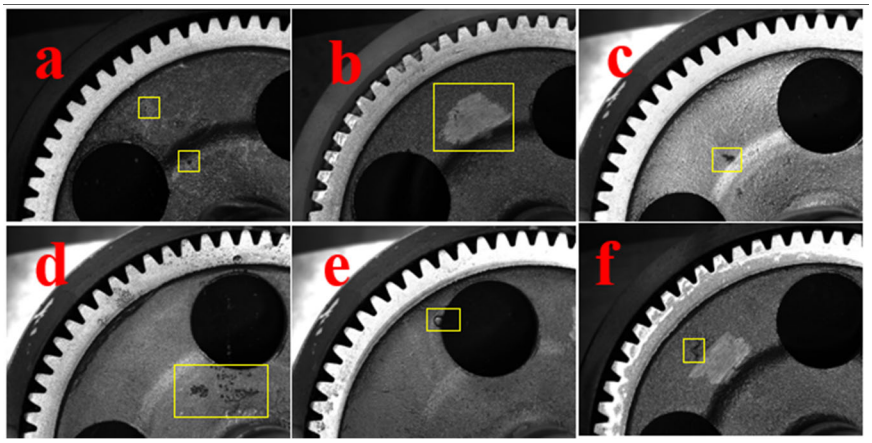


**FIGURE 8.** 6 Images of various defects (a: hole; b:polishing; c:danger; d:rust spot; e:sand hole; f:crack).


**FIGURE 9.** Example of defective area extraction. (a) by the Laplace operator; (b) by the local dynamic threshold segmentation; (c) by the Two-dimensional Garbor filter.

**TABLE 2.** Table of sample statistics for camshaft surface defects data sets.

| Defect type | crack | hole | sand hole | rust spot | polishing | dander |
|---|---|---|---|---|---|---|
| Defective property | Irreparable defect | | | Repairable defect | | Pseudo defect |
| Number of samples | 21 | 42 | 34 | 28 | 42 | 38 |

**TABLE 3.** Table of features contained in the feature pool.

| Geometric feature | Texture feature | Global feature |
|---|---|---|
| Circumference | Entropy | Minimum greyscale value |
| Area | Energy | Maximum greyscale value |
| Convexity | Uniformity | |
| Roundness | Correlation | |
| | Contrast | |

RTX4060. We initialized the parameters of the backbone network using pre-trained weights from ImageNet, and all experiments used uniform parameters. In order to ensure the consistency of the experiment, uniform parameters are used in all comparison experiment. The image size of the input network is fixed to 640 × 640, and 300 epochs are trained. batch size is 8, kernel $k = (3, 3)$, step stride $= 2$, padding $= 0$, object confidence threshold is 0.001, momentum is 0.9, and IoU threshold for NMS is 0.6. Learning rate: 0.001, Weight decay: random weighting, Optimizer: Adam. The activation function is relu. All comparison models are trained and tested on NVIDIA RTX4060. PyTorch library version is 2.0.1.

Fig 10(a) - Fig 10(f) shows the confusion matrices for FFCM-DSM, Faster RCNN Res2Net-101, ViT, Swin-Transformer, DDN, and YOLO v5s. Tab 7 shows the accuracy comparison. In terms of over accuracy metric, FFCM-DSM reaches 0.945, which is higher than Faster RCNN Res2Net-101 (0.867), ViT (0.884), Swin-Transformer (0.890), DDN

**TABLE 4.** Table of the class centers coordinates of the 6 types of defects.

| | Circumference | Area | Convexity | Roundness | Entropy | Energy | Uniformity | Correlation | Contrast | Minimum greyscale value | Maximum greyscale value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Hole** | 0.0064 | 0.0004 | 0.7955 | 0.5721 | 0.4386 | 0.0658 | 0.4879 | 0.3583 | 0.3059 | 0.0274 | 0.0248 |
| **Polishing** | 0.5195 | 0.4760 | 0.7740 | 0.4882 | 0.8126 | 0.0327 | 0.6948 | 0.9758 | 0.0851 | 0.0473 | 0.5967 |
| **Crack** | 0.0744 | 0.0054 | 0.2549 | 0.1004 | 0.4341 | 0.1195 | 0.6215 | 0.5812 | 0.1447 | 0.0949 | 0.0905 |
| **Dander** | 0.0291 | 0.0044 | 0.6914 | 0.4738 | 0.3128 | 0.3074 | 0.8414 | 0.7859 | 0.0249 | 0.2249 | 0.1747 |
| **Rust spot** | 0.0527 | 0.0141 | 0.7431 | 0.5188 | 0.4964 | 0.2223 | 0.7655 | 0.8803 | 0.0660 | 0.3305 | 0.3024 |
| **Sand hole** | 0.0315 | 0.0046 | 0.7312 | 0.5944 | 0.6068 | 0.0570 | 0.5336 | 0.7311 | 0.2785 | 0.0497 | 0.0455 |

**TABLE 5.** Table of Euclidean distances $D_{ij}$ between the centers of 6 types of defects.

| | Hole | Polishing | Crack | Dander | Rust spot | Sand hole |
|---|---|---|---|---|---|---|
| **Hole** | 0 | 1.1992 | 0.7904 | 0.7372 | 0.7820 | 0.4000 |
| **Polishing** | | 0 | 1.1888 | 1.0324 | 0.8662 | 0.9713 |
| **Crack** | | | 0 | 0.7146 | 0.8032 | 0.7474 |
| **Dander** | | | | 0 | 0.3002 | 0.6121 |
| **Rust spot** | | | | | 0 | 0.5585 |
| **Sand hole** | | | | | | 0 |

**TABLE 6.** Table of $S_M(i, j)$ between the 6 types of defects.

| | Hole | Polishing | Crack | Dander | Rust spot | Sand hole |
|---|---|---|---|---|---|---|
| **Hole** | 1 | 0 | 0.4547 | 0.5136 | 0.4641 | 0.8890 |
| **Polishing** | | 1 | 0.0116 | 0.1855 | 0.3704 | 0.2536 |
| **Crack** | | | 1 | 0.5390 | 0.4405 | 0.5026 |
| **Dander** | | | | 1 | 1.0000 | 0.6531 |
| **Rust spot** | | | | | 1 | 0.7172 |
| **Sand hole** | | | | | | 1 |

(0.874), and conventional YOLO v5s (0.884). Tab 8 and Fig 11 show the precision of these six deep-Learning methods for various types of defects. It can be seen that for the four types of defects with high similarity, the precision values are significantly improved after using the FFCM-DSM method. For Hole, Dander, Rust spot, and Sand hole defects, the highest precision of the other five deep learning models reaches 0.846, 0.859, 0.882, and 0.840, respectively, while FFCM-DSM reaches 0.894, 0.918, 0.932 and 0.929, respectively, which are improved by 0.048, 0.059, 0.102, and 0.089. It can be obviously evident that there is a significant improvement in the classification accuracy of FFCM-DSM for these four types of defects with a high degree of similarity. For a more thorough analysis, the three types of real and irreparable defects i.e., hole, crack, and sand hole are analyzed using recall rate. Tab 9 and Fig 12 shows the recall rate of three types of irreparable defects, the results show that FFCM-DSM acquires the maximum recall rate over the hole, crack and sand hole defects compared to the other five methods, with an improvement of 0.044, 0.01, and 0.043 respectively. By comparing the improved YOLO v5s model by using FFCM-DSM with the base YOLO v5s model, it can be observed that after adopting the FFCM-DSM method, there is an improvement of 0.061 in the overall classification accuracy metrics. For the four types of defects with a high degree of similarity, i.e., Hole, Dander, Rust spot, and Sand hole, the classification precision, 0.093, 0.059, 0.102, and 0.116 were improved respectively, which is a significant improvement.

### 2) COMPARISON OF TRADITIONAL MACHINE LEARNING METHODS

Considering the case of inadequate and unbalanced samples, we compare the following five traditional machine learning

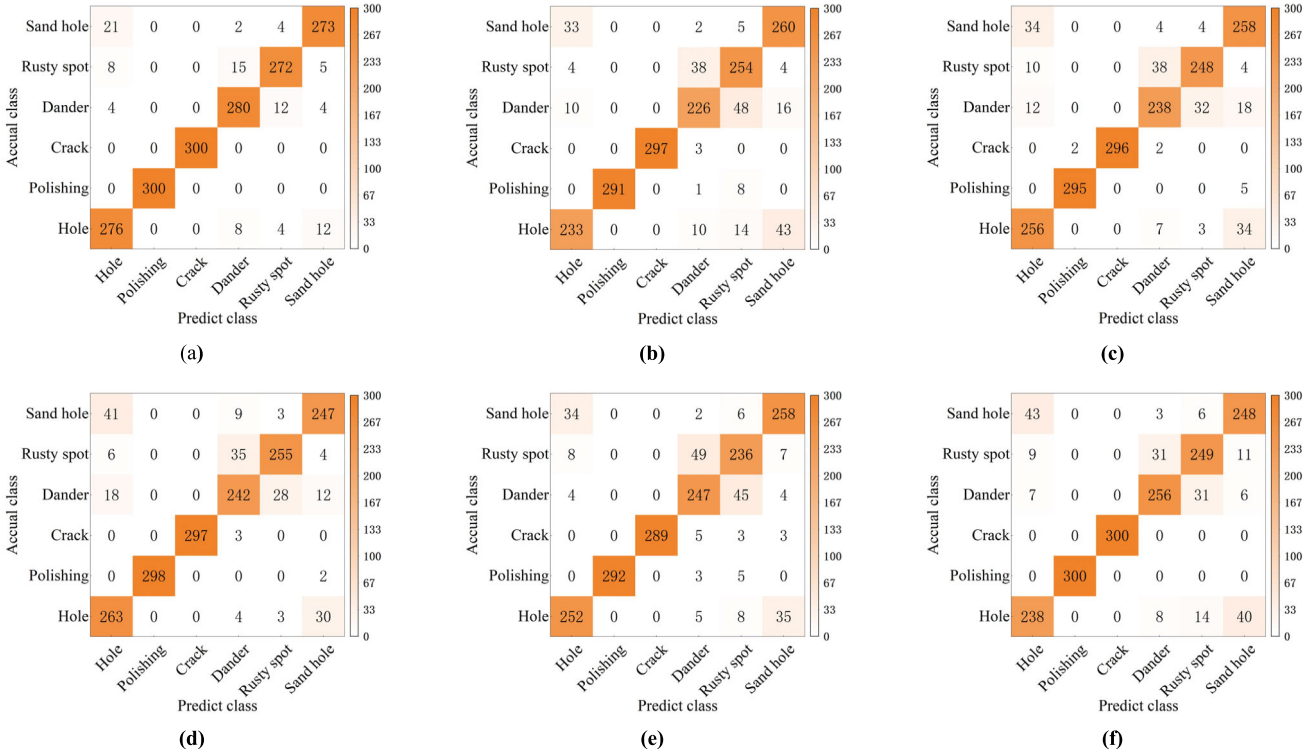


**FIGURE 10.** Confusion matrix of 6 deep learning classification methods. (a) FFCM-DSM; (b) Faster RCNN Res2Net-101; (c) ViT; (d) Swin-Transformer; (e) DDN; (f) YOLO v5s.

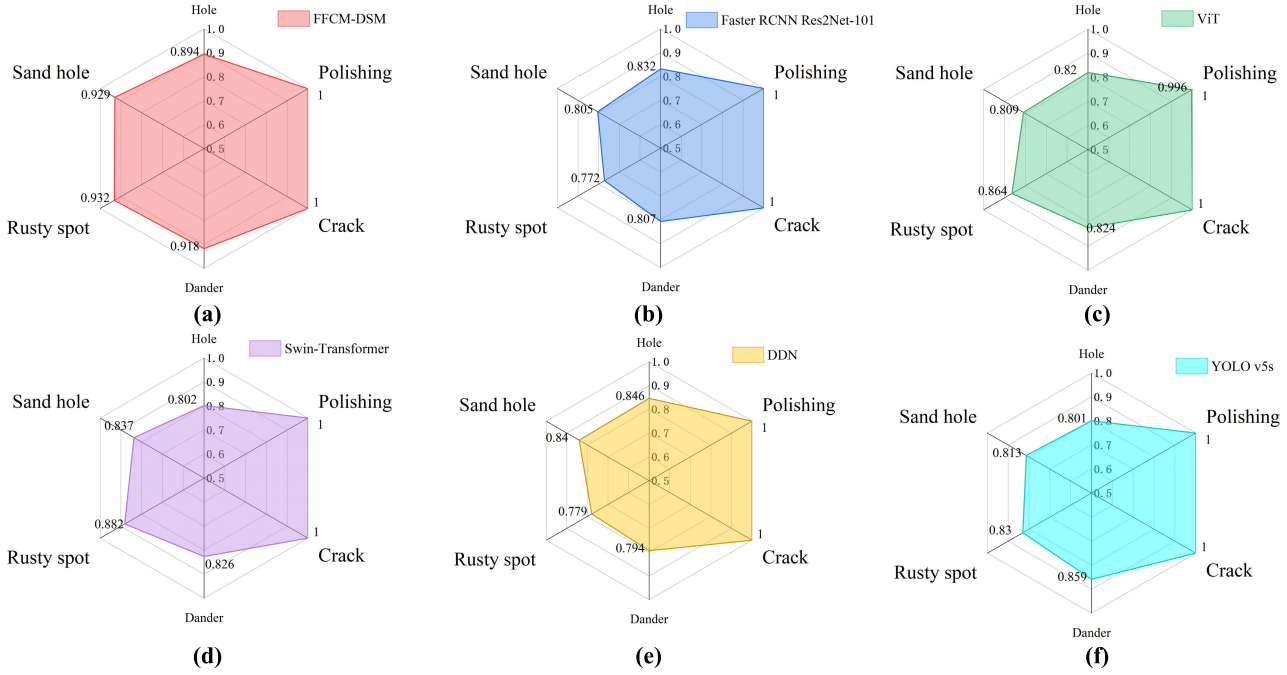


**FIGURE 11.** Comparison graph of precision value for 6 deep learning classification methods. (a) FFCM-DSM; (b) Faster RCNN Res2Net-101; (c) ViT; (d) Swin-Transformer; (e) DDN; (f) YOLO v5s.

algorithms with FFCM-DSM Canonical correlation analysis (CCA) [49]; Concatenate various features to a single feature (CFS) [50]; Autoencoder (AE) [51]; Isometric feature mapping (ISOMAP) [52]; Fisher linear discriminant analysis (FLDA) [53]. For FFCM-DSM, CCA, ISOMAP and FLDA, the number of retained components is 2. For AE, the number of neurons in input and output layers are both 256.

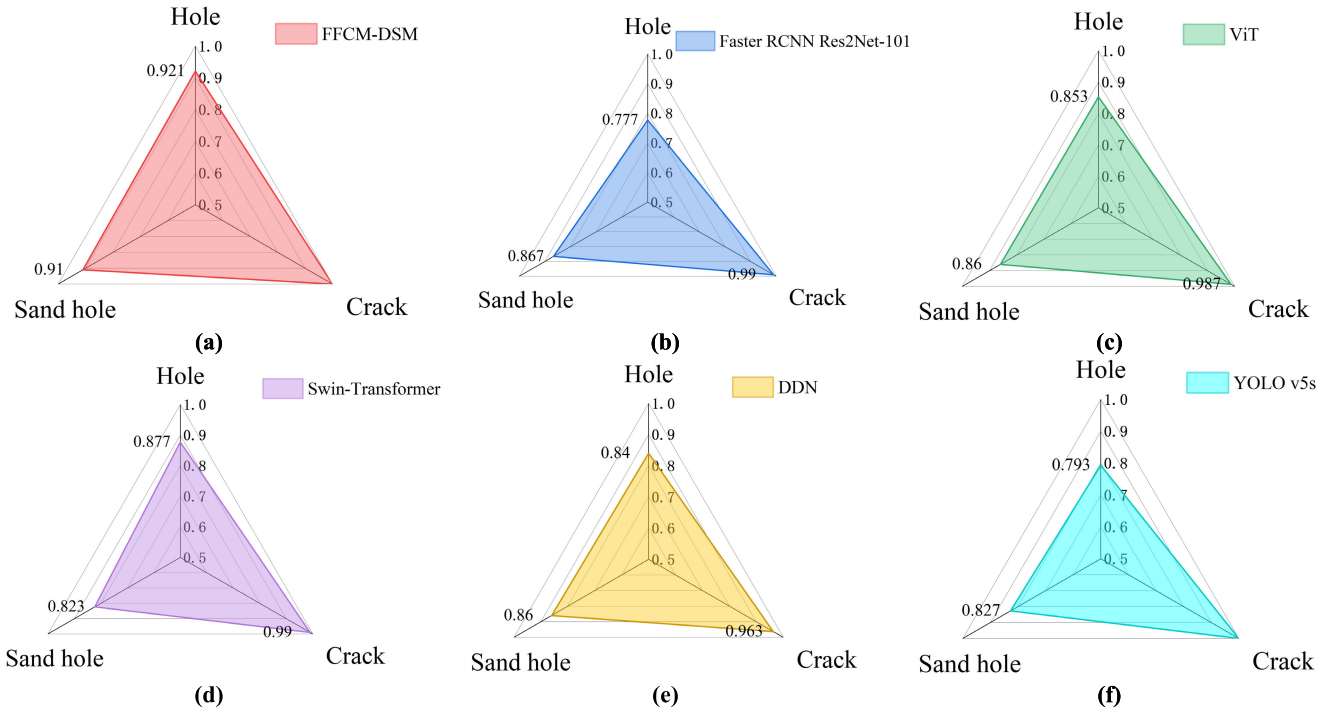


**FIGURE 12.** Comparison graph of recall rate for 6 deep learning classification methods. (a) FFCM-DSM; (b) Faster RCNN Res2Net-101; (c) ViT; (d) Swin-Transformer; (f) YOLO v5s.

**TABLE 7.** The overall accuracy values for 6 deep learning classification methods.

| Method | FFCM-DSM | Faster RCNN Res2Net-101 | ViT | Swin-Transformer | DDN | YOLO v5s |
|---|---|---|---|---|---|---|
| Overall Accuracy | 0.945 | 0.867 | 0.884 | 0.890 | 0.874 | 0.884 |

**TABLE 8.** The precision values of 6 deep learning classification methods for various types of defects.

| | FFCM-DSM | Faster RCNN Res2Net-101 | ViT | Swin-Transformer | DDN | YOLO v5s |
|---|---|---|---|---|---|---|
| Hole | 0.894 | 0.832 | 0.820 | 0.802 | 0.846 | 0.801 |
| Polishing | 1.000 | 1.000 | 0.996 | 1.000 | 1.000 | 1.000 |
| Crack | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Dander | 0.918 | 0.807 | 0.824 | 0.826 | 0.794 | 0.859 |
| Rust spot | 0.932 | 0.772 | 0.864 | 0.882 | 0.779 | 0.830 |
| Sand hole | 0.929 | 0.805 | 0.809 | 0.837 | 0.840 | 0.813 |

**TABLE 9.** The recall rates of the six classification methods for the three non-repairable defects.

| | FFCM-DSM | Faster RCNN Res2Net-101 | ViT | Swin-Transformer | DDN | YOLO v5s |
|---|---|---|---|---|---|---|
| Hole | 0.921 | 0.777 | 0.853 | 0.877 | 0.840 | 0.793 |
| Crack | 1.000 | 0.990 | 0.987 | 0.990 | 0.963 | 1.000 |
| Sand hole | 0.910 | 0.867 | 0.860 | 0.823 | 0.860 | 0.827 |

The corresponding confusion matrices are illustrated in Fig 13 (a)-Fig 13 (f). As shown in Tab 10, similar to the results obtained with deep learning methods, in terms of overall accuracy, FFCM-DSM achieves the highest accuracy index of 0.946 and outperforms the other five classification methods. The remaining fusion methods exhibit decreasing

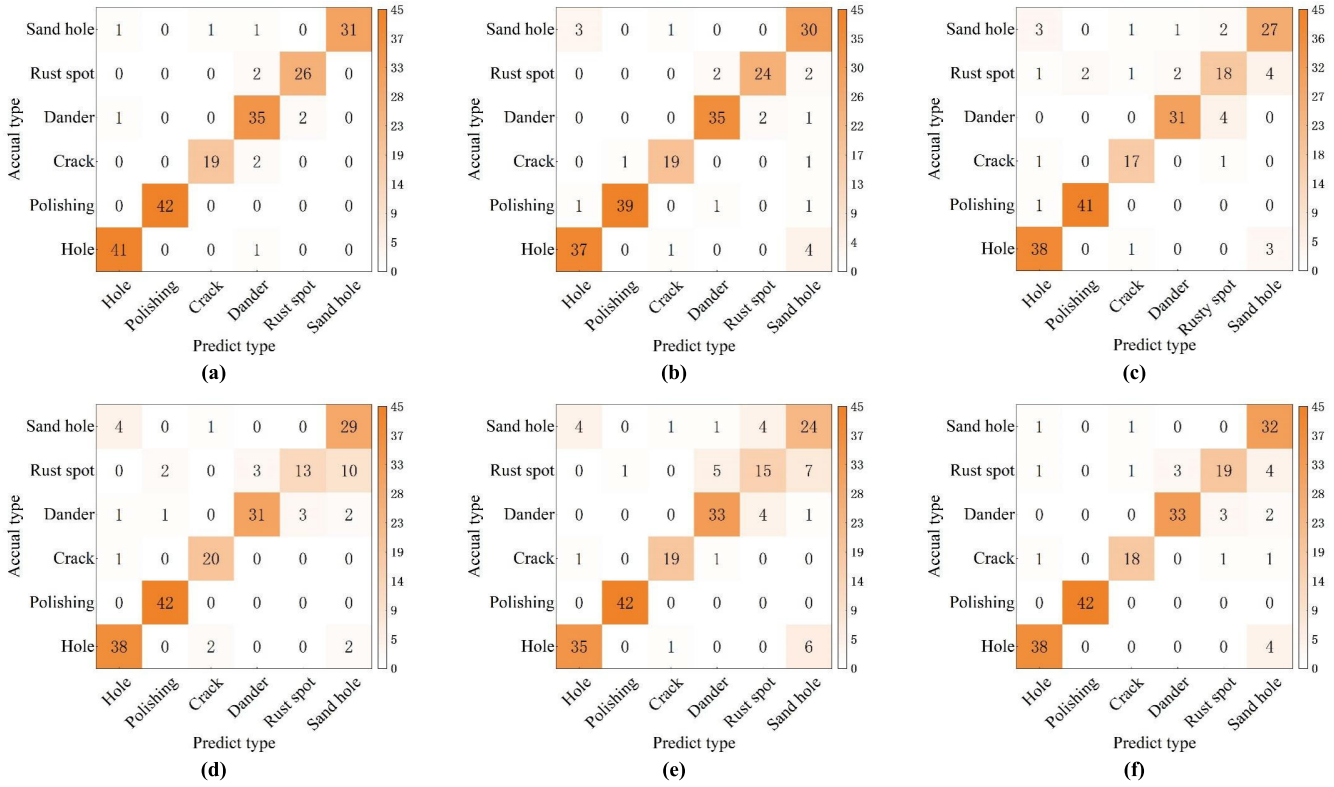


**FIGURE 13.** Confusion matrix of 6 traditional machine learning methods. (a) FFCM-DSM; (b) CCA; (c) CFS; (d) AE; (e) ISOMAP; (f) FLDA.

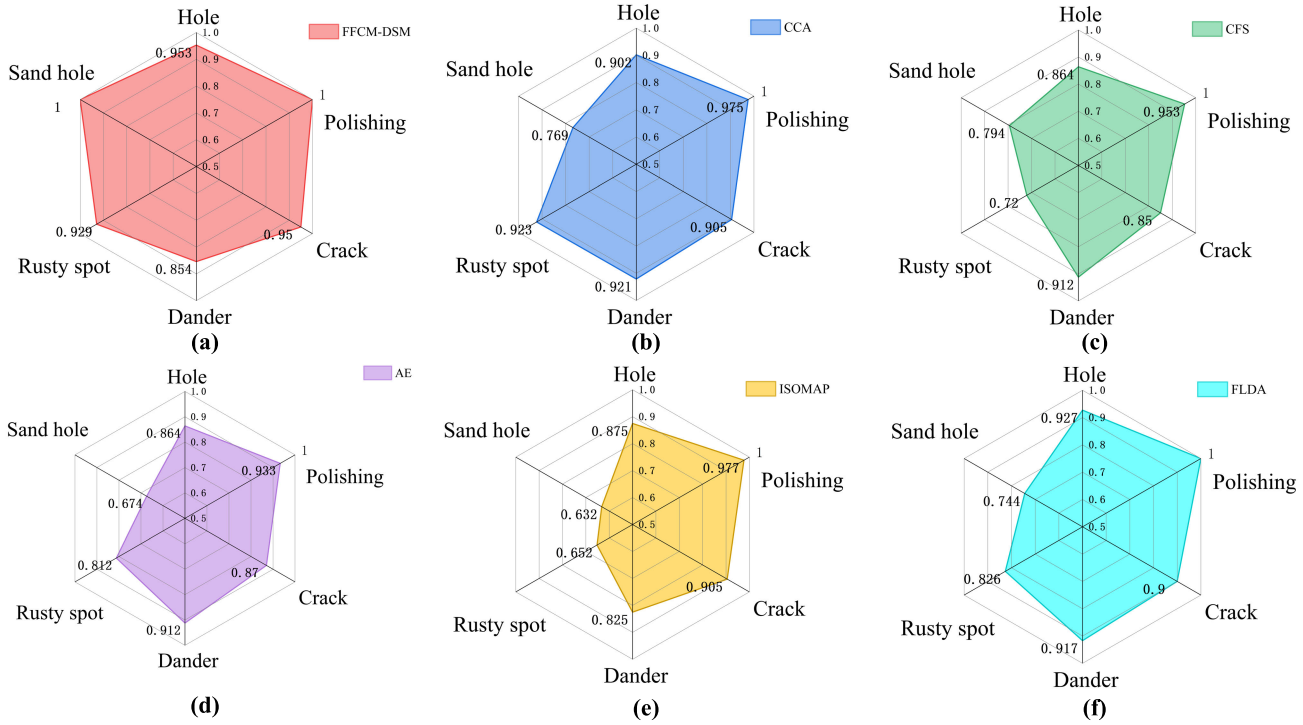


**FIGURE 14.** Comparison graph of precision value for 6 traditional machine learning methods. (a) FFCM-DSM; (b) CCA; (c) CFS; (d) AE; (e) ISOMAP; (f) FLDA.

accuracy indices in the following order: CCA (0.897), ISOMAP (0.888), AE (0.844), CFS (0.839), and FLDA (0.820).

As can be seen in Tab 11 and Fig 14, in terms of precision value, FFCM-DSM performs better than the other 5 classification methods for the five types of defects. Only on the

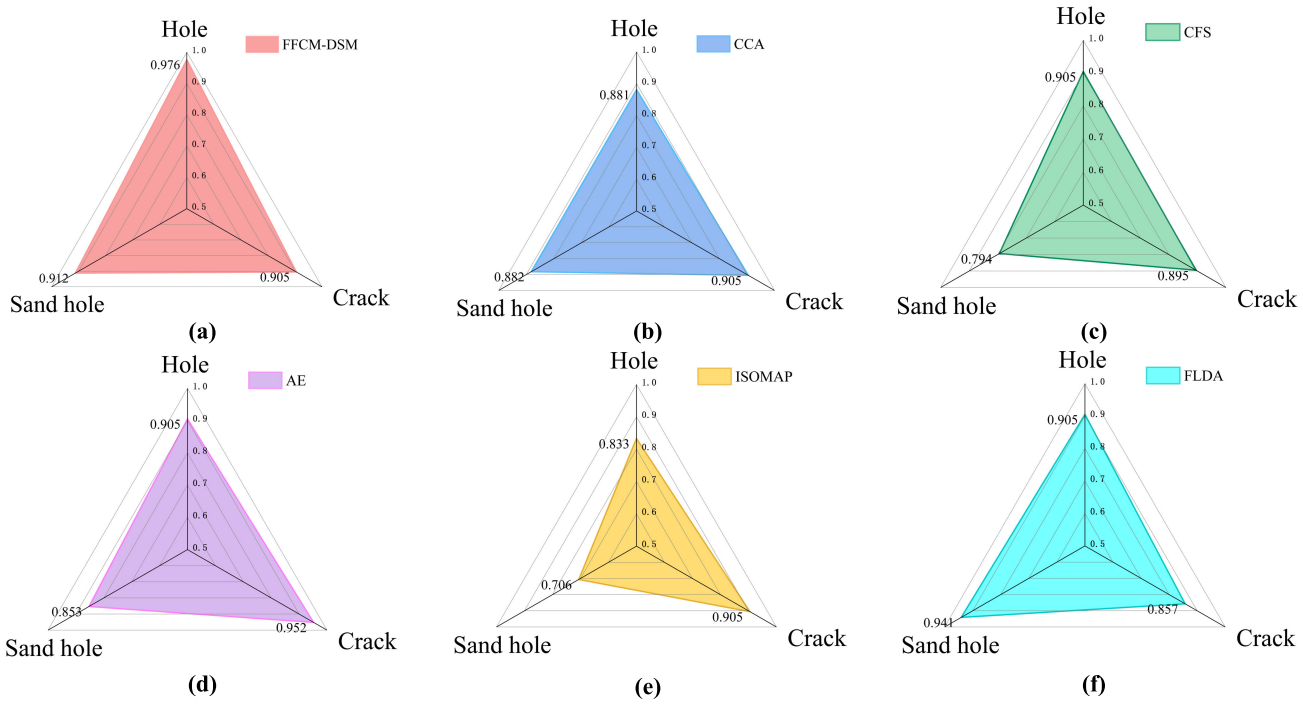


**FIGURE 15.** Comparison graph of recall rate for 6 traditional machine learning methods. (a) FFCM-DSM; (b) CCA; (c) CFS; (d) AE; (e) ISOMAP; (f) FLDA.

**TABLE 10.** The overall accuracy for 6 traditional machine learning methods.

| Method | FFCM-DSM | CCA | CFS | AE | ISOMAP | FLDA |
|---|---|---|---|---|---|---|
| Overall Accuracy | 0.946 | 0.897 | 0.839 | 0.844 | 0.888 | 0.820 |

**TABLE 11.** The precision value of each type of defects for 6 traditional machine learning methods.

| | FFCM-DSM | CCA | CFS | AE | ISOMAP | FLDA |
|---|---|---|---|---|---|---|
| Hole | 0.953 | 0.902 | 0.864 | 0.864 | 0.875 | 0.927 |
| Polishing | 1.000 | 0.975 | 0.953 | 0.933 | 0.977 | 1.000 |
| Crack | 0.950 | 0.905 | 0.850 | 0.870 | 0.905 | 0.900 |
| Dander | 0.854 | 0.921 | 0.912 | 0.912 | 0.825 | 0.917 |
| Rust spot | 0.929 | 0.923 | 0.720 | 0.812 | 0.652 | 0.826 |
| Sand hole | 1.000 | 0.769 | 0.794 | 0.674 | 0.632 | 0.744 |

**TABLE 12.** The recall rate of irreparable defects for 6 traditional machine learning methods.

| | FFCM-DSM | CCA | CFS | AE | ISOMAP | FLDA |
|---|---|---|---|---|---|---|
| Hole | 0.976 | 0.881 | 0.905 | 0.905 | 0.833 | 0.905 |
| Crack | 0.905 | 0.905 | 0.895 | 0.952 | 0.905 | 0.857 |
| Sand hole | 0.912 | 0.882 | 0.794 | 0.853 | 0.706 | 0.941 |

**TABLE 13.** The AUC value for 6 traditional machine learning methods.

| Methods | FFCM-DSM | CCA | CFS | AE | ISOMAP | FLDA |
|---|---|---|---|---|---|---|
| AUC | 0.99 | 0.97 | 0.95 | 0.95 | 0.99 | 0.99 |

Dander defect, only 0.854 was achieved, lower than CCA (0.921) and CFS (0.912). In terms of the recall rate, as shown in Tab 12 and Fig 15, for Hole and Sand hole, FFCM-DSM achieved 0.976 and 0.912, respectively, performing better than the other 5 classification methods. Based on the precision and recall rate, the AUC value are calculated. As can be observed from Tab 13, the AUC value of FFCM-DSM, FLDA and ISOMAP is 0.99, which is better than AE (0.95), CFS (0.95) and CCA (0.97).

Table 14 shows the PCA-SVM classification method without FFCM-DSM in terms of is overall classification accuracy as well as classification precision metrics for each type of

**TABLE 14.** The overall accuracy and precision value for PCA-SVM methods.

| | Types of defects | Value |
|---|---|---|
| **Overall Accuracy** | / | 0.790 |
| **Precision value** | Hole | 0.706 |
| | Polishing | 1.000 |
| | Crack | 0.895 |
| | Dander | 0.778 |
| | Rust spot | 0.586 |
| | Sand hole | 0.714 |

defect. It can be seen that the overall classification accuracy of the method is 0.790, which is much lower than 0.946 using the FFCM-DSM. It can be seen that the overall classification accuracy of the method is 0.790, which is much lower than that of 0.946 with the FFCM-DSM. For the six types of defects counted, the precision value for polishing defects reaches 1.000, which is the equal to the method with the FFCM-DSM, while for the other five types of defects, the values are 0.706, 0.895, 0.778, 0.586 and 0.714, all lower than the method using FFCM-DSM. The results of the above ablation experiments show that after the application of the FFCM-DSM method, the PCA-SVM classification method shows a large improvement in several classification metrics.

By comparing the above with several other types of advanced deep learning methods and traditional machine learning methods, we can observe that FFCM-DSM has significant improvements in several metrics such as overall accuracy, precision, and recall rate etc.

## V. CONCLUSION

In this paper, our contributions are mainly the following two points: (1) We explore the propagation characteristics of light in a non-uniform medium on the surface of a camshaft, based on it, we design a high efficiency optical inspection equipment. (2) We propose FFCM-DSM for the task of defect classification. FFCM-DSM can quantify the similarity between various types of defects and classify the defects with high similarity into one category. After separating the low similarity defects, for different categories, feature extraction and classification operations are implemented respectively to separate high similarity defects. FFCM-DSM achieves high-precision classification of highly similar defects. To verify the effectiveness of the FFCM-DSM, a camshaft surface image acquisition device is built and comparative experiments are carried out. The results show that both for deep learning methods and traditional machine learning methods, FFCM-DSM has gained significant improvements, it improves the overall classification accuracy by approximately 5% and also has the best performance in terms of precision, recall etc. Compared to several other advanced classification methods, especially for defects with high similarity. The proposed method provides a new way for the application of multi-feature fusion classification methods to surface defects detection of industrial parts.

However, there are a few limitations of the proposed FFCM-DSM method. When calculating similarity, feature values need to be extracted manually, which requires vast experience to select useful features for classification.

## REFERENCES

[1] Z. Meng, H. Zhang, D. Guo, S. Chen, and J. Huo, "Defocused calibration for large field-of-view binocular cameras," *Autom. Construction*, vol. 147, Mar. 2023, Art. no. 104737.

[2] J. Huo, H. Zhang, Z. Meng, F. Yang, and G. Yang, "A flexible calibration method based on small planar target for defocused cameras," *Opt. Lasers Eng.*, vol. 157, Oct. 2022, Art. no. 107125.

[3] Y.-H. Huang and H. H. Chen, "Deep face recognition for dim images," *Pattern Recognit.*, vol. 126, Jun. 2022, Art. no. 108580.

[4] L. Zeng, B. Sun, and D. Zhu, "Underwater target detection based on faster R-CNN and adversarial occlusion network," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104190.

[5] M. Nieniewski, "Morphological detection and extraction of rail surface defects," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6870–6879, Sep. 2020.

[6] M. Zhao, W. Qiu, T. Wen, T. Liao, and J. Huang, "Feature extraction based on Gabor filter and support vector machine classifier in defect analysis of thermoelectric cooler component," *Comput. Electr. Eng.*, vol. 92, Jun. 2021, Art. no. 107188.

[7] J. I. Fajardo, C. A. Paltán, L. M. López, and E. J. Carrasquero, "Textural analysis by means of a gray level co-occurrence matrix method. case: Corrosion in steam piping systems," *Mater. Today, Proc.*, vol. 49, pp. 149–154, 2022.

[8] M. Chang, B.-C. Chen, J. L. Gabayno, and M.-F. Chen, "Development of an optical inspection platform for surface defect detection in touch panel glass," *Int. J. Optomechatronics*, vol. 10, no. 2, pp. 63–72, Apr. 2016.

[9] D. Wang, J. Yin, H. Wu, and B. Ge, "Method for detecting internal cracks in joints of composite metal materials based on dual-channel feature fusion," *Opt. Laser Technol.*, vol. 162, Jul. 2023, Art. no. 109263.

[10] Y. Du, J. Chen, H. Zhou, X. Yang, Z. Wang, J. Zhang, Y. Shi, X. Chen, and X. Zheng, "An automated optical inspection (AOI) platform for three-dimensional (3D) defects detection on glass micro-optical components (GMOC)," *Opt. Commun.*, vol. 545, Oct. 2023, Art. no. 129736.

[11] P. Wang, Y.-H. Lee, H.-W. Tseng, and C.-F. Yang, "Research and evaluation on an optical automatic detection system for the defects of the manufactured paper cups," *Sensors*, vol. 23, no. 3, p. 1452, Jan. 2023.

[12] H. Lee and M. Y. Kim, "Automatic optical inspection system with telecentric optics and phase-measuring profilometry for highly accurate localization of electronic packages," *Int. J. Control, Autom. Syst.*, vol. 18, no. 8, pp. 2120–2130, Aug. 2020.

[13] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proc. IEEE*, vol. 103, no. 9, pp. 1560–1584, Sep. 2015.

[14] K. George and P. Sankaran, "Computer assisted recognition of breast cancer in biopsy images via fusion of nucleus-guided deep convolutional features," *Comput. Methods Programs Biomed.*, vol. 194, Oct. 2020, Art. no. 105531.

[15] J. Yin and S. Sun, "Multiview uncorrelated locality preserving projection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3442–3455, Sep. 2020.

[16] K. Zhang, Y. Li, J. Wang, Z. Wang, and X. Li, "Feature fusion for multimodal emotion recognition based on deep canonical correlation analysis," *IEEE Signal Process. Lett.*, vol. 28, pp. 1898–1902, 2021.

[17] H. El-Fiqi, M. Wang, K. Kasmarik, A. Bezerianos, K. C. Tan, and H. A. Abbass, "Weighted gate layer autoencoders," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7242–7253, Aug. 2022.

[18] B. Ren, B. Hou, J. Chanussot, and L. Jiao, "Modified tensor distance-based multiview spectral embedding for PolSAR land cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2095–2099, Dec. 2020.

[19] Y. Qin, H. Wu, X. Zhang, and G. Feng, "Semi-supervised structured subspace learning for multi-view clustering," *IEEE Trans. Image Process.*, vol. 31, pp. 1–14, 2022.

[20] J. Yu, D. Tao, Y. Rui, and J. Cheng, "Pairwise constraints based multiview features fusion for scene classification," *Pattern Recognit.*, vol. 46, no. 2, pp. 483–496, Feb. 2013.

[21] W. Wang, V. Aggarwal, and S. Aeron, "Tensor train neighborhood preserving embedding," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2724–2732, May 2018.

[22] G. Lin, H. Zhu, X. Kang, C. Fan, and E. Zhang, "Feature structure fusion and its application," *Inf. Fusion*, vol. 20, pp. 146–154, Nov. 2014.

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

[24] X. Hou, M. Liu, S. Zhang, P. Wei, and B. Chen, "CANet: Contextual information and spatial attention based network for detecting small defects in manufacturing industry," *Pattern Recognit.*, vol. 140, Aug. 2023, Art. no. 109558.

[25] Z. Liu, R. Tang, G. Duan, and J. Tan, "TruingDet: Towards high-quality visual automatic defect inspection for mental surface," *Opt. Lasers Eng.*, vol. 138, Mar. 2021, Art. no. 106423.

[26] C.-C. Yeung and K.-M. Lam, "Efficient fused-attention model for steel surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[27] W. Qilong, W. Banggu, Z. Pengfei, L. Peihua, Z. Wangmeng, and H. Qinghua, "ECA-net: Efficient channel attention for deep convolutional neural networks," 2019, *arXiv:1910.03151*.

[28] W. Sanghyun, P. Jongchan, L. Joon-Young, and K. I. So, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*.

[29] L. Wang, S. Fang, X. Meng, and R. Li, "Building extraction with vision transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5625711.

[30] C. Wang, H. Liu, X. An, Z. Gong, and F. Deng, "SwinCrack: Pavement crack detection using convolutional swin-transformer network," *Digit. Signal Process.*, vol. 145, Feb. 2024, Art. no. 104297.

[31] S. E. Han, "Transport mean free path tensor and anisotropy tensor in anisotropic diffusion equation for optical media," *J. Opt.*, vol. 22, no. 7, Jul. 2020, Art. no. 075606.

[32] S. S. Park, Y. Jung, and Y. G. Lee, "Spectral dependence on the correction factor of erythemal UV for cloud, aerosol, total ozone, and surface properties: A modeling study," *Adv. Atmos. Sci.*, vol. 33, no. 7, pp. 865–874, Jul. 2016.

[33] C. F. F. Karney, "Sampling exactly from the normal distribution," *ACM Trans. Math. Softw.*, vol. 42, no. 1, pp. 1–14, Mar. 2016.

[34] F. L. Gewers, G. R. Ferreira, H. F. D. Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. D. F. Costa, "Principal component analysis," *ACM Comput. Surv.*, vol. 54, pp. 1–34, May 2021.

[35] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020.

[36] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: A review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 803–855, Aug. 2019.

[37] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system," *Expert Syst. Appl.*, vol. 67, pp. 296–303, Jan. 2017.

[38] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 1794–1807, Jun. 2020.

[39] Y. Cai, S. Mi, J. Yan, H. Peng, X. Luo, Q. Yang, and J. Wang, "An unsupervised segmentation method based on dynamic threshold neural p systems for color images," *Inf. Sci.*, vol. 587, pp. 473–484, Mar. 2022.

[40] S. Shi, H. Si, J. Liu, and Y. Liu, "Facial expression recognition based on Gabor features of salient patches and ACI-LBP," *J. Intell. Fuzzy Syst.*, vol. 34, no. 4, pp. 2551–2561, Apr. 2018.

[41] C. Singh and J. Singh, "Geometrically invariant color, shape and texture features for object recognition using multiple kernel learning classification approach," *Inf. Sci.*, vol. 484, pp. 135–152, May 2019.

[42] K. Prasad and C. B. K. Udupa, "BCHisto-net: Breast histopathological image classification by global and local feature aggregation," *Artif. Intell. Med.*, vol. 121, Nov. 2021, Art. no. 102191.

[43] H. Chen, W. Li, and Y. Zhu, "Improved window adaptive gray level co-occurrence matrix for extraction and analysis of texture characteristics of pulmonary nodules," *Comput. Methods Programs Biomed.*, vol. 208, Sep. 2021, Art. no. 106263.

[44] C. Bowen, W. Yunchao, S. Honghui, F. Rogerio, X. Jinjun, and H. Thomas, "Revisiting RCNN: On awakening the classification power of faster RCNN," 2018, *arXiv:1803.06799*.

[45] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.

[46] J. Ma and H. Chen, "Efficient supervised pretraining of Swin-transformer for virtual staining of microscopy images," *IEEE Trans. Med. Imag.*, vol. 43, no. 4, pp. 1388–1399, Apr. 2024.

[47] J. Kim, H. Yoon, and M.-S. Kim, "Tweaking deep neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5715–5728, Sep. 2022.

[48] X. Zhu, J. Liu, X. Zhou, S. Qian, and J. Yu, "Enhanced feature fusion structure of YOLO v5 for detecting small defects on metal surfaces," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 6, pp. 2041–2051, Jun. 2023.

[49] X. Yang, W. Liu, W. Liu, and D. Tao, "A survey on canonical correlation analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2349–2368, Jun. 2021.

[50] T. Damoulas and M. A. Girolami, "Combining feature spaces for classification," *Pattern Recognit.*, vol. 42, no. 11, pp. 2671–2683, Nov. 2009.

[51] W. Xu, S. Keshmiri, and G. Wang, "Stacked Wasserstein autoencoder," *Neurocomputing*, vol. 363, pp. 195–204, Oct. 2019.

[52] R. Huang, G. Zhang, and J. Chen, "Semi-supervised discriminant isomap with application to visualization, image retrieval and classification," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 6, pp. 1269–1278, Jun. 2019.

[53] J. Yang, J.-Y. Yang, D. Zhang, and J.-F. Lu, "Feature fusion: Parallel strategy vs. serial strategy," *Pattern Recognit.*, vol. 36, no. 6, pp. 1369–1381, Jun. 2003.

**JINSONG CAO** was born in Anhui, China. He received the bachelor's degree in mechanical engineering from Lanzhou University of Technology, in 2019. He is currently pursuing the master's degree in mechanical engineering with the College of Engineering, Zhejiang Normal University. His research interests include the visual detection of surface defects in industrial products and intelligent manufacturing technology.

**HANYANG WU** was born in Zhejiang, China. He received the Ph.D. degree in mechanical engineering from Dalian University of Technology, Dalian, China, in 2020. He is currently a Lecturer with the College of Engineering, Zhejiang Normal University, China. His current research interests include visual image processing, mechanical dynamics, and mechanical optimal design.

**WEIYONG WANG** was born in Zhejiang, China. He received the B.Eng. degree from Northeastern University, in 2011. His research interests include robot technology and intelligent manufacturing technology.

**TEHREEM QASIM** was born in Dera Ismail Khan, Pakistan. He received the B.S. degree in telecommunication systems from Gomal University, Pakistan, in 2012, and the M.Phil. and Ph.D. degrees in electronics from Quaid-i-Azam University, Pakistan, in 2015 and 2021, respectively. Since 2022, he has been an Assistant Professor with the Department of Robotics and Artificial Intelligence, SZABIST University, Pakistan. He has authored nine journals and five conference papers. His research interests include anomaly detection, human activity analysis, biometrics, and remote sensing.

**DONGYUN WANG** was born in Jiangxi, China. He received the Ph.D. degree in mechanical engineering from the College of Mechanical Engineering, Zhejiang University, Hangzhou, China, in 2009, with a focus on hybrid power-train technology on construction machinery.

He is currently a Professor with the College of Engineering, Zhejiang Normal University, China. He has conducted several projects from the National Nature Science Foundation and the Natural Science Foundation of Zhejiang Province. His current research interests include smart manufacturing technologies and machine vision inspection technology.