Article citation info:
Xu S, Cui W, Zhou X, Zhong Q, Wei Y, Wang Y. Machine vision-based surface defect detection method for welds. Diagnostyka. 2025;26(1):2025106. https://doi.org/10.29354/diag/195841.

DIAGNOSTYKA, 2025, Vol. 26, No. 1

e-ISSN 2449-5220
DOI: 10.29354/diag/195841

MACHINE VISION-BASED SURFACE DEFECT DETECTION METHOD FOR WELDS

Su XU [1], Wenxuan CUI [2], Xianzhang ZHOU [3,*], Qiucheng ZHONG [2], Yifan WEI [1], Yipeng WANG [1]

[1] School of Innovation and Entrepreneurship, Jiangsu Ocean University, Lianyungang, 222000, China
[2] School of Mechanical Engineering, Jiangsu Ocean University, Lianyungang, 222000, China
[3] Institute of Vocational Education and Adult Education, Chongqing Academy of Education Science, Chongqing, 400010, China
[*] Corresponding author, e-mail: ccqzxz@outlook.com

Abstract

The texture feature extraction including grayscale co-occurrence matrix and various shape feature extraction methods are adopted in this paper, as well as convolutional neural network based on Visual Geometry Group-16 structure. In particular, the Squeeze-and-Excitation module and dilated convolution technique are introduced to improve the model, aiming to enhance its feature extraction and classification capabilities. On the JPEGWELD dataset, the improved model had 98.7% accuracy in the training set, 97.9% accuracy in the test set, and 98.7% recall rate. In the comparative analysis, although the number of parameters of the improved VGG16 model was 33.64M and the maximum model size was 385MB, the detection time was only 1.3s. The results demonstrated that the model had efficient optimization and computational performance, with a good balance between design and optimization while maintaining a short detection time. The proposed method exhibits high accuracy and efficiency in the detection of various types of weld defects, demonstrating strong universality and adaptability. Its applicability to diverse industrial settings is evident. The study provides an effective solution for industrial automated inspection, which is of great significance to improve the quality control level and production efficiency of manufacturing industry.

Keywords: machine vision; weld seam; surface defects; hollow convolution; gray level co-occurrence matrix; metrology

1. INTRODUCTION

In modern industrial production, welding technology is a part of the key manufacturing process, and the quality of the weld is directly related to the safety and reliability of the product. There are many kinds of weld defects, including but not limited to: cracks, non-fusion, porosity, slag inclusion and burning through. These defects not only show significant visual differences in appearance, but also may produce complex textures and geometric shapes under different welding materials, process conditions and environmental backgrounds. This increases the background complexity of weld defect detection, especially in the case of lighting changes, background interference and different workpiece surface conditions, it is more difficult to accurately detect weld defects. Therefore, it is of great significance to carry out accurate defect detection on the weld. Traditional Weld Defect Detection (WDD) mainly relies on artificial vision with poor efficiency, and it is also easily influenced by operator experience and subjective judgment, resulting in unstable and inaccurate detection results [1]. According to relevant statistics, the incidence of welding defects may be as high as 10% to 30%, depending on the welding material and process. The visual characteristics of weld defects not only depend on the size and shape of the defect, but also are closely related to the surface texture of the weld, lighting conditions, resolution of the camera equipment and other factors [2]. With the development of industrial automation and intelligent manufacturing, an efficient and accurate automatic weld defect detection method is urgently needed. In recent years, machine vision technology has become one of the mainstream methods for automated weld defect detection [3]. In particular, convolutional neural networks (CNNS) have made significant progress in the field of image processing, demonstrating their potential in automated weld defect detection. CNN can automatically extract image features and classify them without manually designing features, but it may face challenges when dealing with complex backgrounds and small defects [4]. The main challenges include how to improve the accuracy of the model in identifying weld defects,

Received 2024-08-22; Accepted 2024-12-11; Available online 2025-01-06

especially in complex backgrounds and different types of defects. Meanwhile, it is necessary to consider how to improve model performance while maintaining or reducing computational costs to adapt to resource constraints in practical industrial applications.

This study aims to improve the accuracy of automatic detection of weld defects by improving CNN models, especially the Visual Geometry Group-16 (VGG16) model. The goal is to develop an MCV system that can accurately identify and classify various weld defects to reduce reliance on manual inspection. To achieve the above objectives, this study adopts texture feature extraction methods including Grayscale Co-occurrence Matrix (GLCM) combined with shape feature extraction technology and discusses a Deep Learning (DL) model based on VGG16. Moreover, the Squeeze-and-Excitation (SE) module and Dilated Convolution (DC) are introduced to enhance the model's feature learning and classification capabilities. The innovation of this study lies in the integration of SE module and cavity convolution technology into the improved VGG16 model. The SE module enhances useful features and removes unimportant features by re-calibrating the feature response of convolutional layers. Cavity convolution increases the receptive field by introducing space interval into the convolution kernel, which makes the model capture a wider range of context information without increasing the computational complexity. The model can effectively improve the recognition ability of weld defects under complex texture background. The research structure is mainly divided into four parts. The first part is to review the existing WDD technology, and the research progress based on MCV, analyze the shortcomings of the existing technology, and provide a basis for putting forward an improvement plan. The second part introduces the methods of weld feature extraction in detail, including texture feature extraction and shape feature extraction. This part explains how to use GLCM and various shape features to extract weld feature. This part describes the structure and feature extraction mechanism of the VGG16 model, introduces the improved method of introducing the SE module and cavity convolution, and expounds upon the working principle and advantages of the improved VGG16 model. The third part proves the effectiveness and superiority of the improved algorithm through experiments, including the performance analysis before and after the improvement and the comparison with other advanced algorithms. This part also shows the experimental results and data analysis in detail. The fourth part is the discussion of the results. The fifth part summarizes the research results, discusses the application prospect of the improved model in industrial WDD, and puts forward the practical significance and future research direction.

2. LITERATURE REVIEW

With the advancement of MCV technology, manual defect detection is gradually being phased out. Numerous scholars are improving MCV and using it in defect detection to achieve intelligent defect detection. Lai proposed a high-precision optical detection system combined with DL technology to accurately detect and analyze various defects in screw heads, including damaged and peeled surfaces. It trained the designed CNN using 3,200 raw images, showing a detection with 92.8% accuracy and average speed of 0.03 s/image. Compared to traditional MCV ways, this system was preferable for inspecting in industrial production lines [5]. Chen et al. proposed a digital twin based on Multi-sensor Fusion (MSF) for in-situ quality monitoring and defect correction during robot laser directed energy deposition process. This method could synchronize and register MSF within the 3D volume of the part. Compared to traditional single-sensor-based monitoring, MSF could gain a deeper understanding of potential process physical characteristics. It paved the way for adaptive additive manufacturing, with superior efficiency, less consumption, and cleaner production [6]. Pratt et al. trained and tested a semantic segmentation model in accordance with the u-net architecture for analyzing EL image of photovoltaic modules made by monocrystalline and polycrystalline silicon solar cells. The focus was to develop a DL method for computer vision. This method was device-independent and could generate EL images regardless of the device used and image quality [7]. Nagata et al. proposed a CNN-based defect detection Support Vector Machine (SVM) technique. They built, trained, and tested deep CNN-based SVM, and promoted visual-based detection to detect various defective manufacturing faults. In this work, pre-trained sssNet and AlexNet were utilized as feature vector extractors for training and showed good detection performance [8].

Eshkevari et al. developed an MCV system for controlling the size characteristics of medical glass bottles. Innovative methods have been developed to capture images of glass bottles and their reflections accurately due to the hardship of doing so. This ensures that appropriate images were gained. Finally, an automated method for measuring the different size characteristics of penicillin bottles was proposed and evaluated using real samples [9]. Iker et al. proposed a method called BoDoC that can improve classification. To evaluate the method, a new dataset was created from the foundry for detecting surface errors in castings with two different defects. The study also introduced a series of techniques for selecting features from images. Through comparison, it was verified that their proposed method improved the direct classification results in real scenes, with an accuracy of 91.305% [10]. Mohamed et al. created a real-time MCV prototype to sort and detect quality parameters of various agricultural products. The

prototype was taken for image acquisition and processing, utilizing color value data from all relevant defects. Python was developed for simple thresholding. The prototype has been successfully used to detect external faults in the tested product with rational accuracy. The detection rates during real-time operation of oranges, potatoes, and peanuts were 96.97%, 98.50%, and 99.09% [11]. Liu et al. proposed a knowledge reuse strategy to train CNN, which transferred knowledge from other visual tasks to industrial defect detection tasks by introducing model-based transfer learning and data augmentation. This strategy had great accuracy with restricted training samples. Experiments on injection molded products have shown that when only 200 images were available for each category, the detection accuracy improved to around 99%. This method was also reliable enough in detecting complex faults with multiple appearances [12]. Zuo et al. proposed an intelligent multi-expert inspection method based on X-ray to solve the problems of efficiency and accuracy in the evaluation of pipeline welding defects. In the actual defect assessment application case, the recall rate of this method is 98.5%, the accuracy is 95.2%, and the accuracy is 97.4% [13]. Jiang et al. proposed a method for welding defect recognition based on convolutional neural networks (CNN) to solve the serial effects of multiple steps such as image preprocessing, region segmentation, feature extraction and type recognition in traditional methods. The results show that this method has higher accuracy than traditional CNN and has been successfully applied to welding defect identification [14].

The above studies show that MCV has a wide range of research in defect detection and has also achieved certain results. Nevertheless, there are still some deficiencies, such as a lack of technical versatility, an inability to be well adapted to different industrial environments, and various types of defect detection. Some high-precision detection methods require a lot of computing resources and complex hardware equipment and are difficult to be applied in resource-limited industrial environments. This study aims to improve the VGG16 model and analyze its application in WDD.

3. SURFACE WDD BASED ON MACHINE VISION

3.1. Weld seam feature extraction

To achieve surface WDD, extracting weld seam features is needed. Therefore, the study starts from the texture and shape characteristics. In texture feature extraction, this study adopts GLCM, which quantifies texture information by examining the spatial relationships of grayscale levels in the image [15-17]. Figure 1 shows the calculation process of GLCM.

Figure 1 shows the grayscale values in the image as the horizontal and vertical coordinates of the matrix, and the quantity of occurrences of grayscale values is statistically analyzed. To reduce the computational complexity of GLCM at high grayscale levels, the grayscale level of the image is often simplified to 8 or 16 levels. This processing significantly reduces computational complexity. Image recognition mainly utilizes statistical features based on GLCM, such as angular Second-order Moments (SDOM). The larger the value of SDOM, the more pixel value sequences each grayscale level contains, resulting in coarser image texture and more prominent feature information. On the contrary, the image texture becomes finer [18]. Through SDOM, weld characteristics can be classified into five types: normal weld, fracture, curling, crack, and burn through.

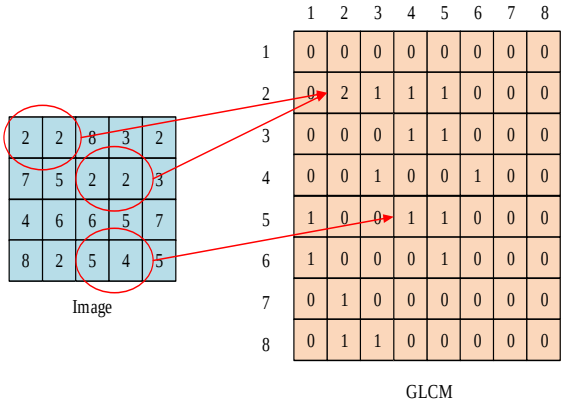


Fig. 1. Calculation of GLCM

Shape is the most intuitive way to describe an object. The shape mainly refers to the closed curve around the object. Shape features are also important features that distinguish different objects. This study focuses on extracting shape features from weld seam images after defect segmentation. The extracted features mainly include weld width, perimeter, area, duty cycle, elongation, and 7 invariant matrices. The expression for the width of the weld seam is shown in formula (1) [19].

$$W = \sum_{i=0}^{n} y_i \tag{1}$$

In formula (1), W represents the image width. The image perimeter is the sum of the pixels of the outer contour of the weld seam, and its expression is formula (2).

$$P = 2 \times (\sum_{i=0}^{m} x_i + \sum_{i=0}^{y} y_i) \tag{2}$$

In formula (2), P represents the circumference of the image. By calculating the girth, a basic geometric feature of the weld can be obtained, which helps to distinguish different types of weld defects. The image area is the amount of pixels included in the weld seam image, as shown in formula (3).

$$Area = \sum_{i=0}^{m} x_i \times \sum_{i=0}^{n} y_i \tag{3}$$

In formula (3), $Area$ represents the area of the image. By calculating the area, it is possible to further understand the size and coverage range of the weld seam, which also helps to accurately classify and

identify weld defects. The image duty cycle is the ratio of the defect area to the minimum bounding rectangle area, as shown in formula (4).

$$Rq = \frac{S}{L*W} \tag{4}$$

In formula (4), Rq is the image duty cycle, S represents the defect area, and L is the length of the bounding matrix. Elongation is the length-width ratio of the min-bounding rectangle of a defect, expressed as formula (5).

$$Rt = \frac{W}{L} \tag{5}$$

In formula (5), Rt represents the image extension. This study uses Hu moment invariants. The order and center moments of the image are represented by formula (6).

$$\begin{cases} m_{pq} = \sum_{1}^{M}\sum_{1}^{N} x^p y^q f(x,y) \\ \mu_{pq} = \sum_{1}^{M}\sum_{1}^{N} (x-\overline{x})^p (y-\overline{y})^q f(x,y) \end{cases} \tag{6}$$

In formula (6), m_{pq} represents the image moment and μ_{pq} represents the image center distance. M and N represent the length and width of the image, respectively. p and q represent the image moment dimension. The center moment is normalized using formula (7).

$$\eta_{pq} = \frac{\mu_{pq}}{\mu^{'}} \tag{7}$$

In formula (7), η represents the normalized standard value of center distance. By combining the center distances of the second and third orders, seven Hu invariant moments can be obtained.

3.2 Classification of weld defects based on CNN

In defect classification problems, machine learning based on image texture and shape features as feature vectors has good results, but its recognition is limited by manually extracted features. Therefore, a method that can automatically extract features for classification is needed. CNN adopts local receptive fields and weight sharing, which improves the model's generalization ability while reducing the training parameters of the neural network. Therefore, using convolution to filter feature values in images has the effect of improving model recognition rate [20]. The structure of the VGG16 network is shown in Figure 2.

Figure 2 is the structure of VGG16, which has 13 convolutional layers and uses a continuous stacking of 3*3 Convolutional Kernels (CKs) instead of the large CKs in AlexNet, reducing parameters and increasing network depth while having the same receptive field [21-22]. The pooling layer adopts the same pooling parameters. The convolutional and pooling layers of the first 16 layers extract image features, and finally complete the classification through 3 Fully Connected Layers (FCLs) and Softmax classification layers to obtain the final result. The convolutional layer in the VGG16 structure extracts features from the data and enhances feature classification. The operation process is shown in Figure 3.

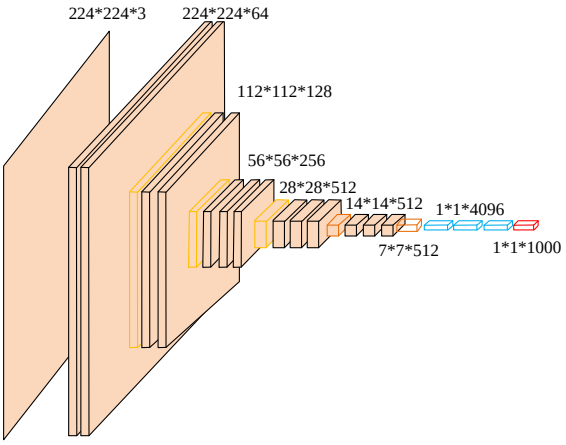


Fig. 2. VGG16 network structure

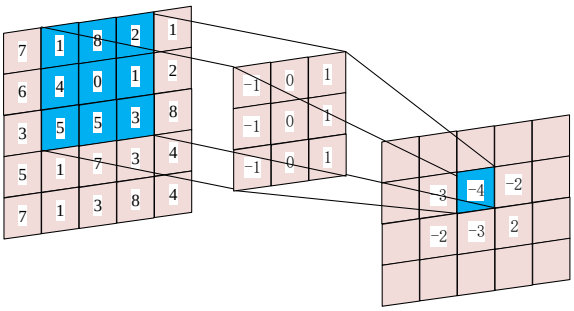


Fig. 3. Convolution operation process

In Figure 3, the essence of the convolution operation is to perform a weighted sum of the images. The CK continuously interacts on the Feature Map (FM), and the weighted sum of the weights and the feature values of each point in the CK region class is used to complete feature extraction. The size of convolutional output features is formula (8).

$$n = \frac{w+2p-f}{s}+1 \tag{8}$$

In formula (8), n and w are the size of the output/input FM. p and f are the size of the edge filled pixels and CK. s represents the stride of the CK sliding. After the convolutional layer, it is usually a pooling layer, which aims to reduce the dimensionality of the feature images output by the convolutional layer. The pooling layer has both average pooling and maximum pooling. The features obtained by average pooling are more sensitive to flat areas, while the features obtained by maximum pooling are more sensitive to edge and corner areas. Figure 4 shows the pooling process of both.

After different pooling methods in Figure 4, it is found that average pooling takes the average feature value within the template area as the output, while maximum pooling takes the max feature value within the template area as the output. Another function of the pooling layer is to expand the receptive field of the network, which refers to the mapping range of

pixels on the output FM to the input FM. The feature values output by CNN represent the information of a certain region in the input FM. The larger the receptive field, the wider the mapping range of the output feature values on the input FM, indicating that the output feature values contain more extensive information. Conversely, the larger the receptive field, the more detailed the information contained in the output feature values.

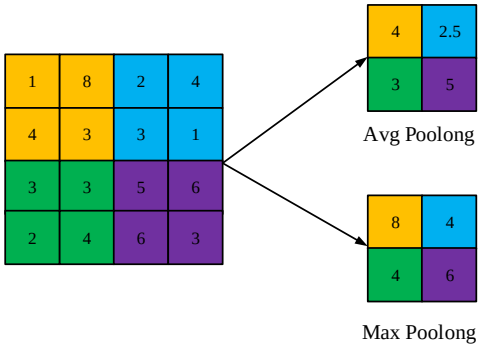


Fig. 4. Maximum pooling and average pooling

3.3. Improved CNN model for VGG16

In the VGG16 network structure, convolutional layers perform undifferentiated feature extraction on welding images. Although a large number of features can ensure detection accuracy, they also introduce features that are not of concern to weld defects. Therefore, this study aims to address the above issues and improve the model by introducing the SE module. Figure 5 shows the basic structure of SE.

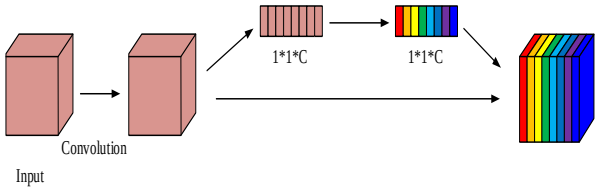


Fig. 5. Basic structure of SE module

In Figure 5, F_{tr} represents convolution and operation, and U represents FM. The image is compressed, stimulated, and the weights are redefined to obtain features. Compression is the global average pooling of FMs, compressing each feature channel into a real number, so that each real number can represent the global receptive field of that channel. After completing the compression, an FM of $1\times1\times C$ is obtained. Incentive is composed of two FCLs. The first FCL completes the dimensionality reduction of the FM and is activated by the Relu function. The second FCL completes the dimensionality enhancement of the FM and is activated by the Sigmoid function. The two fully connected processes continuously learn the weight values of each channel through weight learning.

After completing convolution in the VGG16 model, a pooling layer is added, partly to increase the receptive field of convolution and reduce computational complexity. However, this process also reduces the spatial resolution of the input data. To address this issue and increase the receptive field, DC is used in this study. The difference in convolution between DC and the original CNN model of VGG16 lies in the difference in receptive fields. The receptive field of the convolutional layer in the original VGG16 model is continuous, while DC can take values at intervals. The use of DC can effectively reduce the number of pooling layers and improve Recognition Accuracy (RA). Choosing different DC rates results in different receptive fields for convolution operations, and the scale of the obtained feature data information is also different. Different scales of information are crucial in data processing. The expression for DC is formula (9).

$$g_{ri-r-1,rj-r-1} = h_{ij} \tag{9}$$

In formula (9), h_{ij} represents the convolution kernel in the VGG16 network, and r represents the DC rate. Assuming r is the convolution kernel of 1 and 2, and its DC representation is shown in Figure 6.

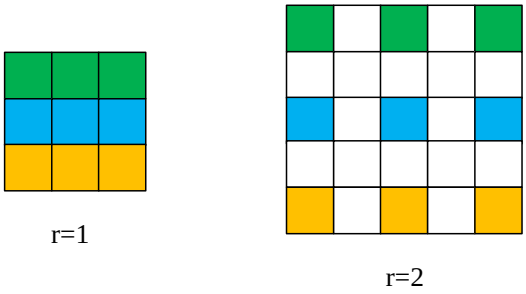


Fig. 6. Dilated convolution

In Figure 6, the colored blocks represent the parameters that the convolution kernel needs to calculate, while the white blocks represent the positions where the matrix is 0. The size of the CK obtained after the expansion of the DC kernel is formula (10).

$$o = k + (k-1)(r-1) \tag{10}$$

In formula (10), o and k are the size of the expanded CK and original CK. Since the presence of zero values does not increase computational complexity, adding zero values between CKs can increase the receptive field of convolutional operations without increasing computational complexity. In summary, in the welding surface defect detection method, the MCV technology and algorithm used in the research include GLCM, CNN and the improved VGG16 model. Firstly, the texture features of welds are extracted by GLCM, which quantifies texture information by examining the spatial relationship of image gray level, and divides welds into five features by statistical features. Secondly, shape feature extraction includes weld width, perimeter, area, duty cycle, elongation and Hu invariant moment. These features are extracted through the segmented weld image to better distinguish different types of weld defects. The VGG16 network structure is used to classify weld

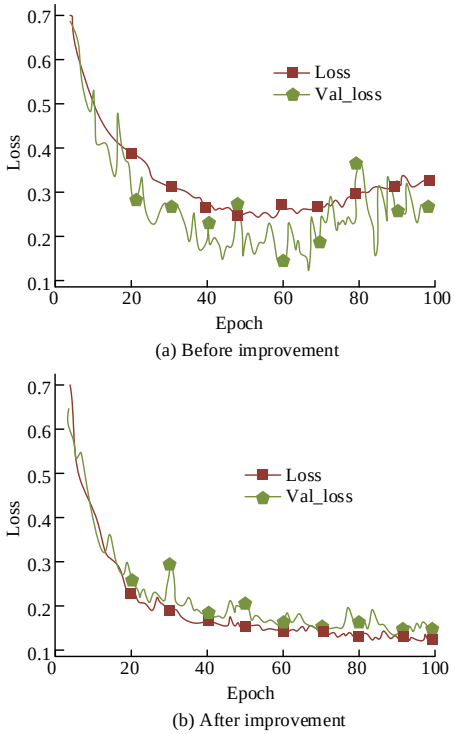
defects. The network is composed of 13 convolutional layers, and image features are extracted through continuous stacking of 3×3 convolutional nuclei, and the classification is completed through the FCL and Softmax classification layer. To enhance the performance of the model, an SE module has been incorporated, which serves to enhance the useful features and to remove the unimportant features by recalibrating the feature response of the convolutional layer. In addition, hollow convolution is used to increase the receptive field without increasing the computational complexity. By introducing space interval into the convolution kernel, the model can capture a wider range of context information, thus improving the identification accuracy of weld defects.

4. ANALYSIS OF WDD EFFECT BASED ON IMPROVED VGG16 CINN MODEL

This study conducts performance comparison experiments to validate the effectiveness of the improved algorithm. To ensure fairness and accuracy, all networks use the same dataset and set the optimal parameters in the experiment. The JPEGWELD data set is selected in this study, which contains 4,000 images of steel plate weld defects. The images of the data set are divided into five defect types: normal weld, fracture, flanging, crack and burn through. The JPEGWELD dataset contains images of weld defects of different types and severity, covering common welding problems. The diversity ensures that the model can be effectively trained and tested on different defect types. The image quality of the data set is high, which ensures the accurate extraction of texture and shape features. Although the dataset contains a variety of weld defect types, the number of images for some defect types can be uneven, resulting in the model learning less about certain types of defects during training. The utilization of the JPEGWELD data set in the experiment is conducive to the promotion of the learning ability and generalization ability of the model, as well as to subsequent comparative experiments. This study first analyzes the performance of the VGG16 CNN model before and after improvement, and evaluates it through loss rate, accuracy, summoning rate, and F1 score indicators. Then, current advanced algorithms are selected for comparison.

4.1. Performance analysis of improved VGG16 CNN model

This study divides the dataset into training and testing sets in an 8:2 ratio and set the initial learning rate of the improved VGG16 model to 0.001 and the batch size to 32. In the process of WDD, high-quality weld images are first captured by high-resolution industrial cameras, multi-spectral imaging and laser scanning technologies. Then, the images are preprocessed by gray-scale, normalization, image enhancement, denoising and data enhancement to improve the clarity of the images and the accuracy of feature extraction. The main data acquisition equipment is as follows: Basler ace1920-40gm high-resolution industrial camera with a resolution of 1920x1200; Laser scanner Leica BLK3D, accuracy ± 1.0mm; The multi-spectral imaging system Headwall Photonics Micro-Hyperspec covers visible light and infrared spectrum band only 400nm-1000nm. Tthe gray-scale processing is to convert the color image to gray-scale image to reduce the computational complexity and concentrate on processing the brightness information of the weld. Image normalization normalizes image pixel values to [0,1] to reduce the impact of light changes on detection results. The contrast of weld and defect is improved by edge enhancement and other methods. In terms of feature extraction of weld defect samples, the study not only extracts the JPEGWELD data set, but also obtains these features through the weld width, depth, Angle and other geometric features through high-precision measurement equipment, which is combined with image data for comprehensive analysis. In the above experimental environments and experimental methods, the loss rate results of the model before and after improvement are shown in Figure 7.



(a) Before improvement

(b) After improvement

Fig. 7. Comparison results of loss curves B&A model improve`ment

Figures 7 (a) and (b) show the loss function curves of the unimproved and improved VGG16 CNN models. In 7 (a), the trend of this curve shows that the model is difficult to converge during the training process. In Figure 7 (b), the improved model has a fast convergence rate in the first 20 iterations. After 60 iterations, the curve gradually stabilizes and the final convergence value is 0.1. Figure 8 is the model accuracy before and after improvement.

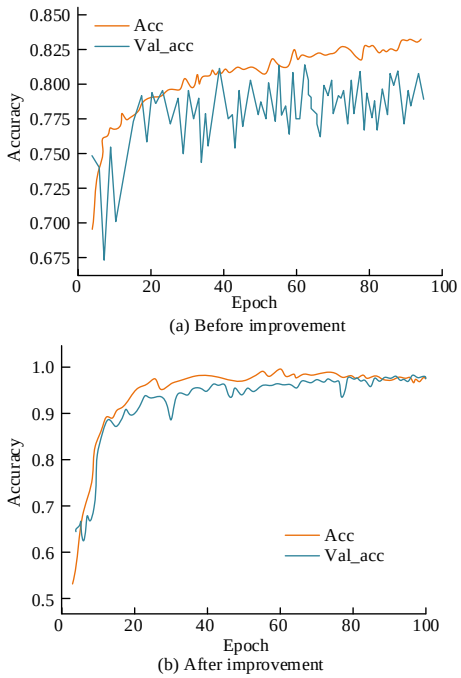


Fig. 8. Comparison results of accuracy curves B&A model improvement

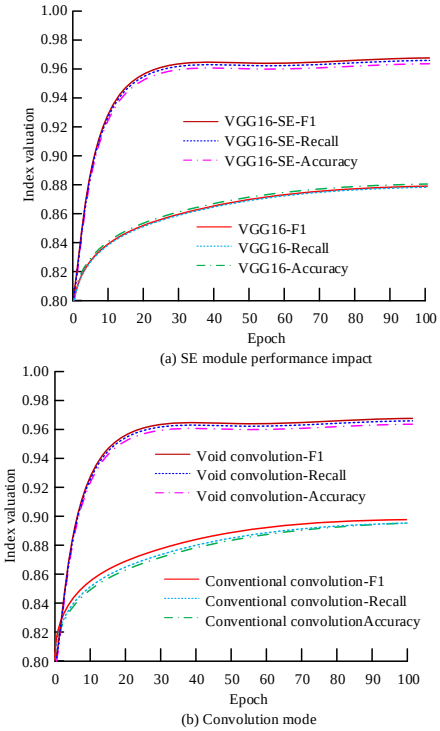


Fig. 9. Indicator evaluation B&A model improvement

Figures 8 (a) and (b) show the accuracy curves of VGG16 before and after improvement. In 8 (a), the curve change results indicate that the RA of the training set of the unimproved model is 82.3%, and the RA of the test set is 77.2%. In 8 (b), the improved model has an accuracy of 98.7% in the training and 97.9% in the test. This study introduces the SE module and improves the VGG16 model using DC instead of traditional convolution. The improvement effect is shown in Figure 9.

Figure 9 (a) shows the evaluation results of indicators before and after the introduction of the SE module. When the SE is added to the improved VGG16 model, the accuracy increases from 0.88 to 0.964, the recall increases from 0.879 to 0.958, and the F1 score increases from 0.879 to 0.961. This indicates that the SE module is crucial for improving model performance. The SE module enhances the ability to learn useful features by dynamically adjusting the weight of feature channels, significantly improving accuracy, recall, and F1 score. Figure 9 (b) shows the evaluation results using DC indicators. The improved VGG16 model using DC achieves accuracy, recall, and F1 score of 0.964, 0.958, and 0.961, while the indicators of traditional convolution are 0.897, 0.896, and 0.896, respectively. DC introduces spatial intervals in the CK, allowing the model to increase the receptive field without increasing computational complexity. In the results of Figure 9, the curves of each model have almost the same changes, which may be caused by the action of SE module. The SE module enhances the classification capability of the model by re-calibrating the feature response of the convolutional layer, enhancing useful features and suppressing unimportant ones. This enhancement effect is global, meaning that there is a significant improvement in all features. Consequently, the accuracy, recall, and F1 score of the model improve simultaneously when detecting different types of weld defects, resulting in the curve of these indicators becoming consistent. In addition, accuracy, recall and F1 scores are evaluated from different angles based on the same classification results. Accuracy measures the proportion of all predictions that are correct. Recall measures the proportion of all actual positive examples that are correctly identified. F1 scores are the harmonic average of accuracy and recall. A model's performance improvement in one category is typically accompanied by a simultaneous improvement in all three indicators, and vice versa. Therefore, there is a certain correlation between these indicators, causing their curve changes to converge. Based on the above, the SE module enhances useful features and suppresses unimportant features by re-calibrating the feature response of the convolutional layers. It provides a more detailed feature control mechanism for the model by learning the importance of each channel, thereby making the model more focused on useful features for WDD tasks and improving classification accuracy. DC allows the model to increase the receptive field without increasing additional computational costs. This method enables the model to capture a wider range of contextual information, which helps to better understand the complex textures and details in weld seam images. This is very meaningful for improving the accuracy of defect detection and reducing missed detection.

4.2. Improving the WDD performance of the model

This study classifies five types of weld seams and compared them using three structures: ResNet, GoogleNet, and MobileNe. Among them, ResNet adopts a very deep network structure, in which each residual block is composed of two or three convolutional layers, and the input is directly bypassed to some layers and added to the output by short-circuit connection, so as to make training more efficient [23]. GoogleNet's network structure is relatively deep, but it also reduces the number of parameters by increasing the width of the Inception module. Each Inception module processes the input data through convolution kernels of different sizes and concatenates the outputs of various convolution layers to obtain information of different scales [24]. MobileNet introduced deep separable convolution and split the standard convolution operation into deep convolution and point-by-point convolution, which significantly reduced the computational effort [25]. Figure 10 shows the classification results of different structures.

Figures 10 (a) to 10 (d) represent the classification results of improved VGG16, ResNet, GoogleNet, and MobileNe, respectively. The classification effect of Figure 10 (a) has a significant distinguishing effect, and different welding types have significant feature distinctions. Other methods have significantly higher misclassification or misidentification rates than VGG16 in classification. Table 1 shows the evaluation data of identification indicators for different welds using different models.

Table 1 shows that the improved VGG16 performs the best in terms of accuracy for all types of welds, especially in identifying normal welds, cuts, and cracks, with an accuracy rate of over 96%. The improved VGG16 also has an advantage in terms of recall rate, especially in the detection of normal welds, with a recall rate of 98.7%. The improved VGG16 has the highest F1 score among all models, indicating a good balance between accuracy and recall. In Table 1, the ResNet performance is relatively weak, which may be due to its structure and parameter settings not being fully suitable for complex WDD tasks. GoogleNet and MobileNet perform well in certain types of weld seam detection, but their overall performance is still inferior to the improved VGG16. The improved VGG16 model significantly outperforms other models in terms of accuracy, recall, and F1 score when dealing with surface WDD tasks. This indicates that its improvement measures have greatly improved the overall performance. To further verify the performance in similar algorithms, the results of this study comparing its detection time, parameter quantity, and model size are listed in Table 2.

In Table 2, the improved VGG16 has the highest number of parameters, which usually means that the model is more complex and may have stronger learning and representation abilities. The improved VGG16 model has the maximum size of 385MB, which is consistent with its high parameter count. Although the improved VGG16 has the most parameters and the largest model size, its detection time is the shortest, only 1.3 seconds. This indicates

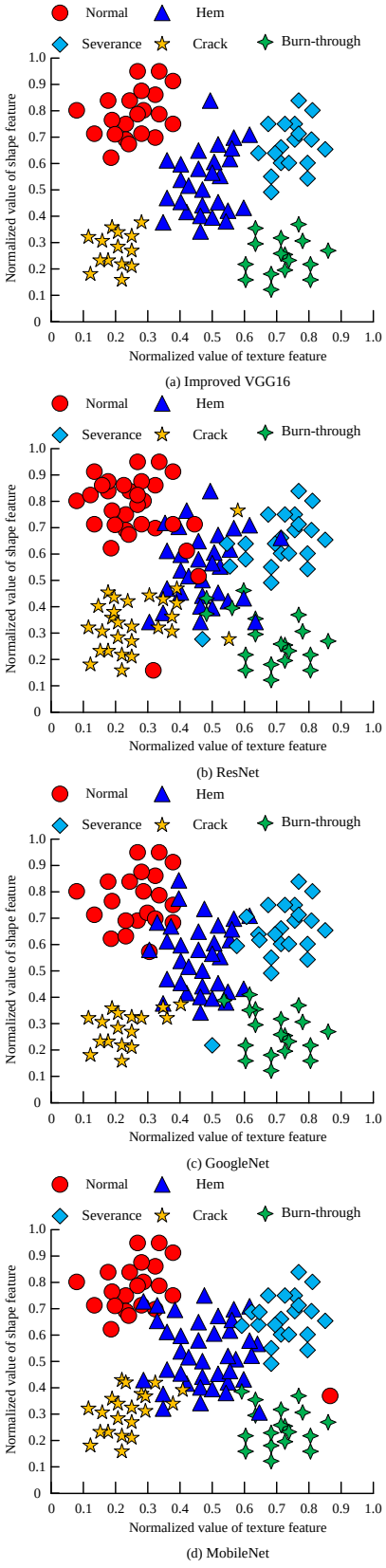


Fig. 10. Classification results of different methods

Table 1. Evaluation results of indicators for different models

Algorithm		Normal weld	Severance	Hem	Crack	Burn-through
Accuracy	ResNet	0.743	0.709	0.694	0.606	0.714
	GoogleNet	0.857	0.806	0.842	0.657	0.714
	MobileNet	0.746	0.758	0.696	0.735	0.687
	Improved VGG16	0.975	0.966	0.951	0.964	0.966
Recall	ResNet	0.687	0.733	0.625	0.666	0.833
	GoogleNet	0.825	0.833	0.800	0.766	0.733
	MobileNet	0.775	0.773	0.575	0.833	0.733
	Improved VGG16	0.987	0.966	0.975	0.900	0.966
F1 score	ResNet	0.714	0.721	0.657	0.635	0.768
	GoogleNet	0.840	0.819	0.820	0.707	0.758
	MobileNet	0.760	0.745	0.630	0.780	0.709
	Improved VGG16	0.981	0.967	0.962	0.930	0.967

Table 2. Comparison of different model parameters, sizes, and detection times

Algorithm	Number of parameters (M)	Model Size (MB)	Detection time (s)
ResNet	25.58	97.9	1.5
GoogleNet	5.94	40	1.8
MobileNet	3.5	21.6	1.4
Improved VGG16	33.64	385	1.3

its efficient optimization and computational performance. The improved VGG16 has the most complex model structure while maintaining a shorter detection time. This indicates that it has achieved a good balance between design and optimization, even though the model is large and has multiple parameters, it can still perform inference calculations quickly. To further compare the advancement and effectiveness of the research methods, the study adopts three mainstream methods for comparative analysis. These detection models include Faster Region-based CNN (Faster-RCNN), Kalman Filter and Convolutional Neural Networks (CNN-ENKF), and Gaussian Mixed Convolutional Autoencoders and Convolutional Neural Networks (GMCA-CNN) [26-28]. Meanwhile, Accuracy, Recall, F1 Score and Detection Time are used for evaluation, and the results are shown in Table 3.

Table 3. Performance comparison of different algorithms

Algorithm	Accuracy (%)	Recall (%)	F1 Score (%)	Detection Time (s)
Faster-RCNN	94.2	93.5	93.8	2.5
CNN-ENKF	96.1	95.4	95.7	2.2
GMCA-CNN	95.5	94.8	95.1	1.8
Improved VGG16	97.9	98.7	97.9	1.3

The results in Table 3 show that the accuracy of the improved VGG16 is 97.9%, which is 3.7%, 1.8% and 2.4% higher than that of Faster-RCNN, CNN-ENKF and GMCA-CNN, respectively. In terms of recall rate and F1 score, the improved VGG16 is 98.7%, which has the best performance among the four methods. In terms of detection time, the detection time of the improved VGG16 is 1.3s, which is reduced by 1.2s, 0.9s and 0.5s, respectively, compared with Faster-RCNN, CNN-ENKF and GMCA-CNN. The results show that compared with the current mainstream methods, the proposed method still has better performance, and its effectiveness and advanced nature have been verified.

5. DISCUSSION

On the JPEGWELD data set, the accuracy of training set and test set of the improved model reached 98.7% and 97.9%, respectively. Across different weld types, the improved VGG16 demonstrated an accuracy and recall rate of more than 96% and an average F1 score of 0.967. The improved VGG16 model was superior to the comparison model in both accuracy and recall rate. For example, the Faster-RCNN model had an accuracy rate of 94.2%, while the improved VGG16 model achieved 97.9%. In terms of recall rate, the improved VGG16 model was 98.7%, while the CNN-ENKF was 95.4%. This result showed that the improved VGG16 model performed well in the identification of various weld defects, especially in the detection of small defects and complex textures. Although the improved VGG16 model parameters and model size were larger, its detection time was the shortest, only 1.3s. This showed that the optimal design of the model maintained the superiority of complex model structure while ensuring high detection efficiency. In contrast, the detection times of ResNet and GoogleNet were 1.5s and 1.8s, respectively, and their accuracy and recall rates were not as good as the improved VGG16 model. The results showed that the introduction of SE module and cavity convolution significantly enhanced the feature extraction capability of the model. By re-calibrating the feature response of the convolutional layer, SE module effectively enhances the weight of useful features and supsets unimportant features. By expanding the receptive field, the cavity convolution technique enables the model to capture a wider range of context information, which is particularly important for complex texture recognition in WDD. In current studies, the high-precision optical detection system based on DL [5] and the semantic segmentation model based on u-net architecture proposed by Pratt et al. [7] have achieved good results in specific fields. However, these methods have some limitations in universality and adaptability and are difficult to be applied to different industrial environments and various types of defect detection. In contrast, the improved VGG16 model has excellent performance on a variety of weld defect types, and has strong versatility and adaptability, which can be applied to

weld inspection requirements in various industrial environments.

6. CONCLUSION

This study explored automated detection methods on the basis of MCV to meet the needs of industrial WDD. This study was based on VGG16 CNN and introduced SE module and DC technology for improvement to enhance the model's capacity to extract weld defect features. Additionally, GLCM and various shape features were used for weld seam feature extraction, and a DL model adapted to WDD was constructed. Through the experimental results, the improved VGG16 model has achieved remarkable results in improving the accuracy and efficiency of WDD, which proves the application value of SE module and cavity convolution in optimizing CNNs. The proposed model improved the accuracy of defect detection, thus improving the quality and reliability of products. The efficient performance was suitable for industrial production lines that require real-time monitoring. At the same time, the error caused by manual detection was reduced. Although the improved VGG16 model performed well, it still had some limitations. The weld image may be affected by different lighting conditions during the acquisition process. For example, too bright or too dark light can make details of welds and defects difficult to capture, affecting the recognition accuracy of the model. Steel plate welds may exhibit different textures and physical properties depending on the material. Images of welds with different materials can pose challenges to the model's ability to generalize, especially if there are not enough material types covered in the training data. The improved VGG16 model has high parameter number and model complexity due to the introduction of SE module and void convolution, which leads to its application in resource-constrained environments. Future research directions include optimizing the model structure, and further reducing the computational complexity and storage requirements of the model through model pruning, quantization and knowledge distillation. It is also possible to expand and diversify training datasets, including weld seam images of different materials and welding processes, to enhance the model's generalization ability. This model has great potential for application in manufacturing quality control fields such as automotive industry, aerospace industry, construction engineering, and electronic product manufacturing.

Source of funding: *The research is supported by: 2022 Jiangsu Provincial College Students' Innovation Training Provincial Key Project SZ202211641649001 – Anti-epidemic medical waste harmless treatment robot under the dual carbon target; 2023 Jiangsu Provincial College Students' Innovation Training Provincial General Project SY202311641651001 – Design of medical supplies rapid delivery system based on unmanned aerial vehicle technology; The author gratefully acknowledges the financial supports by the 2021 Chongqing Municipal Education Commission Science and Technology Research Plan Major Project under Grant numbers KJZD-M20211440; 2022 Chongqing Municipal Research Institute Performance Incentive Guidance Special Project under Grant numbers cstc2022jxjl40004.*

Author contributions: *research concept and design, S.X.; Collection and/or assembly of data, W.C.; Data analysis and interpretation, S.X., X.Z., Q.Z.; Writing the article, X.Z., Q.Z.; Critical revision of the article, Y.W., Y.W.; Final approval of the article, S.X., W.C., X.Z., Q.Z., Y.W., Y.W.*

Declaration of competing interest: *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.*

REFERENCES

1. Lee H, Heogh W, Yang J, Yoon J, Park J, Ji S, Lee H. Deep learning for in-situ powder stream fault detection in directed energy deposition process. Journal of Manufacturing Systems. 2022;62(4):575-587. https://doi.org/10.1016/j.jmsy.2022.01.013.
2. Chen Y G, Shu Y, Li X, Xiong C, Cao S, Wen X, Xie Z. Research on detection algorithm of lithium battery surface defects based on embedded machine vision. J. Intell. Fuzzy Syst. 2021;41(3):4327-4335. https://doi.org/10.3233/JIFS-189693.
3. Wang W, Lu K, Wu Z, Long H, Wang B. Surface defects classification of hot rolled strip based on improved convolutional neural network. ISIJ International. 2021;61(5):1579-1583. https://doi.org/10.2355/isijinternational.ISIJINT-2020-451.
4. Abhilash P M, Chakradhar D. Image processing algorithm for detection, quantification and classification of microdefects in wire electric discharge machined precision finish cut surfaces. Journal of Micromanufacturing. 2021;5(2):116-126. https://doi.org/10.1177/25165984211015410.
5. Lai J Y, Tsao Y R, Liu C Y. High-accuracy detection and classification of defect and deformation of metal screw head achieved by convolutional neural networks. Applied Mechanics and Materials. 2022;909(1):75-80. https://doi.org/10.4028/p-fy36nq.
6. Chen L, Yao X, Liu K, Tan C, Moon S. Multisensor fusion-based digital twin in additive manufacturing for in-situ quality monitoring and defect correction. Proceedings of the Design Society. 2023;3(1):2755-2764. https://doi.org/10.1017/pds.2023.276.
7. Pratt L, Govender D, Klein R. Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation. Renewable Energy. 2021;178(1):1211-1222. https://doi.org/10.1016/j.renene.2021.06.086.
8. Nagata F, Watanabe K, Habib M K. Design and implementation of convolutional neural network-based SVM technique for manufacturing defect detection. International Journal of Mechatronics and Automation. 2021;8(2):53-61. https://doi.org/10.1504/IJMA.2021.115240.
9. Eshkevari M, Rezaee M J, Zarinbal M, Hamidreza I. Automatic dimensional defect detection for glass vials

based on machine vision: A heuristic segmentation method. Journal of Manufacturing Processes. 2021; 68(2):973-987. https://doi.org/10.1016/j.jmapro.2021.06.018.
10. Iker P, Sanz B, Tellaeche A, Psaila G, Gaviria J, Bringas P. Quality assessment methodology based on machine learning with small datasets: Industrial castings defects. Neurocomputing. 2021;456(7):622-628. https://doi.org/10.1016/j.neucom.2020.08.094.
11. Mohamed A R, Elgamal R A, Elmasry G, Radwan S. Development of a real-time machine vision prototype to detect external defects in some agricultural products. Journal of Soil Sciences and Agricultural Engineering. 2021;11(9):317-325. https://doi.org/10.21608/jssae.2021.178987.
12. Liu J, Guo F, Gao H, Li M, Zhang Y, Zhou H. Defect detection of injection molding products on small datasets using transfer learning. Journal of Manufacturing Processes. 2021;70(7):400-413. https://doi.org/10.1016/j.jmapro.2021.08.034.
13. Zuo F, Liu J, Fu M, Wang L, Zhao Z. An X-Ray-based multiexpert inspection method for automatic welding defect assessment in intelligent pipeline systems. IEEE/ASME Transactions on Mechatronics; 2024; 1(2):1-12. https://doi.org/10.1109/TMECH.2024.3408337.
14. Jiang H, Hu Q, Zhi Z, Gao J, Gao Z, Wang R, et al. Convolution neural network model with improved pooling strategy and feature selection for weld defect recognition. Welding in the World. 2021;65(1):731-744. https://doi.org/10.1007/s40194-020-01027-6.
15. Amini N, Shalbaf A. Automatic classification of severity of COVID-19 patients using texture feature and random forest based on computed tomography images. International Journal of Imaging Systems and Technology. 2022;32(1):102-110. https://doi.org/10.1002/ima.22679.
16. Wei W, Liu C, Wang J. Objective evaluation of optical illusion skirt based on image texture features. International Journal of Clothing Science and Technology. 2021;33(5):843-859. https://doi.org/10.1108/IJCST-10-2020-0163
17. Hui T, Xu Y L, Jarhinbek R. Detail texture detection based on Yolov4-tiny combined with attention mechanism and bicubic interpolation. IET Image Processing. 2021;15(12):2736-2748. https://doi.org/10.1049/ipr2.12228.
18. Pal S, Roy A, Shivakumara P, Pal U. Adapting a swin transformer for license plate number and text detection in drone images. Artificial Intelligence and Applications. 2023;1(3):145-154. https://doi.org/10.47852/bonviewAIA3202549.
19. Vafaeian B, Riahi H T, Amoushahi H, Jomha N M, Adeeb S. A feature-based statistical shape model for geometric analysis of the human talus and development of universal talar prostheses. Journal of Anatomy. 2022; 240(2):305-322. https://doi.org/10.1111/joa.13552.
20. Li Z, Wei X, Hassaballah M, Li Y, Jiang X. A deep learning model for steel surface defect detection. Complex & Intelligent Systems. 2024;10(1):885-897. https://doi.org/10.1007/s40747-023-01180-7.
21. Tang R, Liu Z, Song Y, Duan G, Tan J. Hierarchical multi-scale network for cross-scale visual defect detection. Journal of Intelligent Manufacturing. 2024; 35(3):1141-1157. https://doi.org/10.1007/s10845-023-02097-1.
22. Yang Z, Zhang M, Chen Y, Hu N,Gao L, Liu L, Xi E, Song J. Surface defect detection method for air rudder based on positive samples. Journal of Intelligent Manufacturing. 2024;35(1):95-113. https://doi.org/10.1007/s10845-022-02034-8.
23. Kansal K, Chandra T B, Singh A. ResNet-50 vs. EfficientNet-B0: Multi-centric classification of various lung abnormalities using deep learning "session id: ICMLDsE.004". Procedia Computer Science. 2024; 235(1):70-80. https://doi.org/10.1016/j.procs.2024.04.007.
24. Russel NS, Selvaraj A. Wavelet scattering transform and deep features for automated classification and grading of dates fruit. Journal of ambient intelligence and humanized computing. 2024;15(6):2909-2923. https://doi.org/10.1007/s12652-024-04786-y.
25. Jenipher V N, Radhika S. Lung tumor cell classification with lightweight mobileNetV2 and attention-based SCAM enhanced faster R-CNN. Evolving Systems. 2024;15(4):1381-1398. https://doi.org/10.1007/s12530-023-09564-3.
26. Zhang X, Chen G. Detection of dense small rigid targets based on convolutional neural network and synthetic images. Traitement du Signal: signal image parole, 2021;38(1):61-71. https://doi.org/10.18280/ts.380106.
27. Ruckstuhl Y, Janji T, Rasp S. Training a convolutional neural network to conserve mass in data assimilation. Nonlinear Processes in Geophysics. 2021;28(1):111-119. https://doi.org/10.5194/npg-28-111-2021.
28. Bhaskar N, Ganashree. Lung nodule detection from ct scans using gaussian mixture convolutional autoencoder and convolutional neural network. Annals of the Romanian Society for Cell Biology. 2021;25(4):6524-6531.

Su XU, male, born in 1973, master, associate professor, master tutor. His main research interests are industrial vision, artificial intelligence, mechatronics technology, digital design and manufacturing.
e-mail: xusu1973@tom.com

Wenxuan CUI, male, born in 1999, postgraduate student. His research interests are machine vision and deep learning.
e-mail: cuiwenxuan216@163.com

Xianzhang ZHOU, male, born in 1975, holds a master's degree and is an associate professor and supervisor of master's candidates. His main research direction focuses on big data and artificial intelligence.
e-mail: ccqzxz@outlook.com

Qiucheng ZHONG, male, born in 2002, Jiangsu Ocean University student.
e-mail:
zqc19549532@163.com

Yifan WEI, male, born in 2001, Jiangsu Ocean University student.
e-mail:
wyf1506751424@126.com

Yipeng WANG, male, born in 2000, Jiangsu Ocean University student.
e-mail:
15251286169@139.com