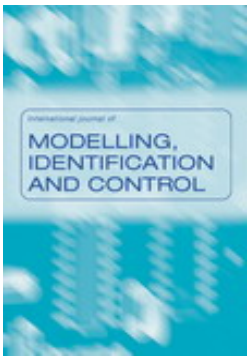


International Journal of Modelling, Identification and Control

ISSN online: 1746-6180 - ISSN print: 1746-6172
https://www.inderscience.com/ijmic

Detection with thermal imaging for packaging bag sealing based on knowledge transfer

Shaoyu Tang, Lisheng Wei, Rui Wang, Pinggai Zhang

DOI: 10.1504/IJMIC.2024.10063841

Article History:

Received:	27 July 2023
Last revised:	04 November 2023
Accepted:	29 February 2024
Published online:	30 September 2024

Detection with thermal imaging for packaging bag sealing based on knowledge transfer

Shaoyu Tang and Lisheng Wei*

School of Electrical Engineering,
Anhui Polytechnic University,
Anhui, 241000, China
Email: 2210320119@stu.ahpu.edu.cn
Email: weilsh@ahpu.edu.cn
*Corresponding author

Rui Wang

School of Computer Science and Technology,
China University of Mining and Technology,
Jiangsu, 221116, China
Email: wrui19961023@163.com

Pinggai Zhang

School of Electronic Engineering,
Chaohu University,
Anhui, 238000, China
Email: zhangpinggai@outlook.com

Abstract: To address the problem that most enterprises still use the manual packaging bag seal detection method with low efficiency and poor stability, we propose an automatic detection method, which is based on knowledge transfer, to detect with thermal imaging the packaging bag sealing. Firstly, the thermal image of packaging bag seal is obtained by a thermal imager, random forest (RF) and support vector machine (SVM) are trained by small sample labels, and the two classifiers are fused to build an expert labelling system for labelling unlabelled samples. Then, the enhanced samples are created by combining the predicted samples and the labelled samples, and input into the fine-tuned VGG16 (visual geometry group) for training and testing. Finally, the experiment shows that the prediction accuracy of this method reaches 96.25%, which verifies the effectiveness and feasibility of the proposed method instead of manual detection method.

Keywords: defect detection; expert labelling system; fine-tuned VGG16; knowledge transfer; thermal imaging.

Reference to this paper should be made as follows: Tang, S., Wei, L., Wang, R. and Zhang, P. (2024) 'Detection with thermal imaging for packaging bag sealing based on knowledge transfer', *Int. J. Modelling, Identification and Control*, Vol. 45, No. 1, pp.58–69.

Biographical notes: Shaoyu Tang received his Bachelor's in Automation from Anhui Polytechnic University. He is currently studying for his Master's degree in the School of Electrical Engineering, Anhui Polytechnic University. His research interests include machine learning and machine vision.

Lisheng Wei received his Bachelor's degree in 2001 from Anhui Polytechnic University, and received his Master's degree in 2004 from China Aerospace Science and Industry Corporation 061 Base, and PhD degree in 2009 from Shanghai University. Currently, he is a Professor and Master's Student Supervisor at Anhui Polytechnic University. His research interests include image recognition and application, embedded instrumentation and system, intelligent network control theory, system and simulation.

Rui Wang received his Master's in Control Science and Engineering from Anhui Polytechnic University. He is currently studying for his PhD in the School of Computer Science and Technology, China University of Mining and Technology. His research interests include machine learning and data mining.

Pinggai Zhang received his PhD in Control Theory and Control Engineering from Shanghai University. He is currently a Lecturer at Chaohu University. His research interests include computer science and automation and control systems.

1 Introduction

Packaging bags are an important part of industrial development, and products in the form of packaging bags are widely used in daily life and industrial production (Mirza et al., 2021). The sealing performance of packaging bags is an important index of packaging quality, which is related to the quality of the packaged products and affects their transportation, storage, and sales (Beshai et al., 2020). The traditional method of detecting defects in the sealing of packaging bags with the human eye cannot continuously and stably complete this highly repetitive and intelligent task. In recent years, with the continuous development of packaging intelligence, computers have been used to simulate human visual mechanisms to obtain and process information and realise target recognition, tracking, and measurement. Such methods have been widely used in various industrial fields (Chai, 2020; Chai et al., 2022). In the packaging field, using machine vision to quickly, accurately, and non-destructively detect sealing defects in packaging bags has become an important problem that is very concerning and needs to be urgently solved (Rodrigues et al., 2021). However, each network requires a large number of datasets to train, and different products require different datasets. The difficulty of obtaining datasets makes the compatibility of detection networks low.

Many scholars have conducted in-depth research on accurate image recognition based on vision. Zhou and Li (2022) used machine learning basis functions to process the texture images of packaging seals and achieved good detection results. Sun et al. (2020) proposed an image detection method of cigarette packaging defects based on support vector machine, which can efficiently identify cigarette packaging defects, has universal and real-time, and can meet the requirements of abnormal situation detection of cigarettes. Zeng (2022) realised the detection of different types of defects in packaging by using improved SVM classifier, which laid a foundation for subsequent research. Wang and Ma (2022) proposed an improved YOLO-v5 (you only look once) method for damage detection of medical waste packaging bags, which has higher detection accuracy and faster reasoning speed in the task of damage detection of medical waste packaging bags. Chen et al. (2020) used the improved VGGNet and the material field model tool in TRIZ innovation theory to classify the defects of carton yolk, achieving the purpose of high precision and meeting the requirements of real-time sorting. Chen et al. (2021) proposed a fast packaging defect detection method based on the improved MobileNet-V2 lightweight network of transfer learning, which effectively detected four surface defects existing in the packaging link of candy production line, and verified the feasibility of automatic defect detection. Sa et al. (2020) proposed a fast, accurate, effective and intelligent packaging inspection system based on ResNet, which combines a highly portable deep learning framework and can be widely used in production processes. Wu et al. (2023) fused the DarkNet-53 backbone network in YOLO-v5 with SE-Net to detect defects in potato chip packaging, improving the recognition speed and accuracy. Vu et al. (2023) proposed a real-time packaging defect detection system based on the YOLO algorithm, which can be integrated into factories and assembly lines to improve production efficiency and save operating costs.

These studies mainly focus on the construction of detection and classification models, and pay less attention to the construction of datasets. However, deep learning requires a large number of labelled datasets for training, and different products require different datasets. Therefore, aiming at the problem of dataset establishment and defect detection, this paper proposes a thermal imaging detection method for packaging sealing pockets based on knowledge transfer. An expert labelling system is constructed by fusing RF and SVM, and the expert labelling system is used to label unlabelled samples, to obtain large datasets quickly. The knowledge learned by the expert labelling system is transferred to the fine-tuned convolutional neural network for training, classification and recognition, so as to realise the fast and accurate detection of thermal imaging packaging bag sealing defects. This method reduces the time of establishing the dataset in the early stage, meets the defect detection of different packaging, and has high recognition accuracy and efficiency. It provides a reference for the large-scale practical application of automatic defect detection systems.

2 Packaging bag sealing defect detection algorithm

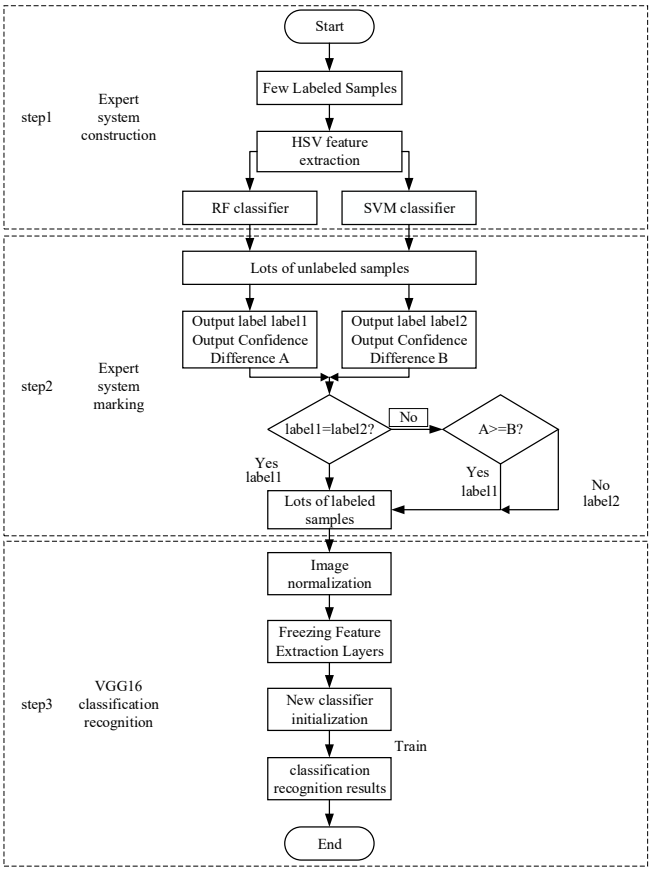
2.1 Algorithm flowchart

In order to improve the speed and accuracy of defect detection in thermal images of packaging bag seals, a method for this based on knowledge transfer was studied, and a flowchart thereof is shown in Figure 1.

It can be seen in Figure 1 that the packaging bag sealing defect detection method based on knowledge transfer was mainly divided into three steps: First, a small number of labelled packaging bag defect image samples were used to train a random forest (RF) and support vector machine, and they were fused to build an expert system. Then, the expert system was used as a marker to mark a large number of unlabelled samples to obtain a large number of labelled samples. Finally, the labelled samples and labelled prediction samples were used to create an enhanced training set, which was input into a convolutional neural network for

training, and the prediction and classification results were obtained.

Figure 1 Schematic diagram of the detection algorithm

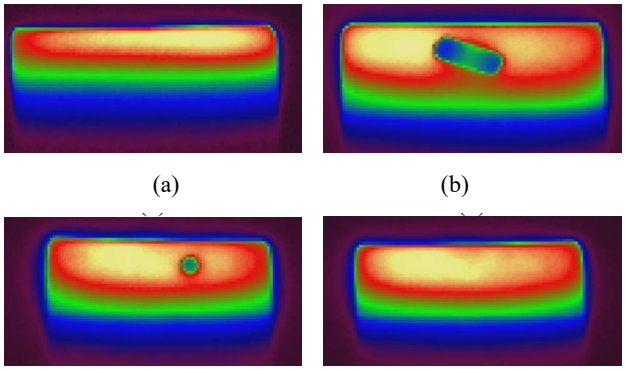


2.2 Feature selection

In order to intuitively judge defect images, a Hikvision thermal imager was used to illuminate common defective packaging bags to form thermal images. For non-defective packaging bags, when the heat flux was injected evenly, it could be evenly diffused on the surfaces of the bags, and the temperature field on the surface was evenly distributed. However, when a packaging bag had heat insulation or heat conduction defects, it locally formed a hot zone or a cold zone. Because of the different colours of the defective parts on the thermal imaging map, the colour features were used to extract the features of the images of packaging bag defects (Hu et al., 2021).

According to the principle of optical imaging, when a lens focuses on an object, its image is clearer in the image plane, while objects located in other positions will appear to be blurred to varying degrees. In actual production, the thermal images of defects in packaging bag seals mainly include half-package, linear-mezzanine, and dotted-mezzanine defects, as shown in Figure 2. Hue, saturation, value (HSV) can perceive colours more accurately than the traditional red, green, blue (RGB) colour space and remain computationally simple. Therefore, HSV features are used in this paper.

Figure 2 Schematic diagram of sealing defect of packaging bags, (a) half-package (b) linear-mezzanine (c) dotted-mezzanine (d) normal (see online version for colours)



The steps of the operation of HSV colour feature extraction were as follows: first, the labelled and unlabelled sample files were read, the RGB colour space of each picture was converted into the HSV colour space by using the translating algorithm, and then each channel was divided into 8 groups to calculate histograms. Then, multi-dimensional arrays are flattened into one-dimensional arrays, which was the HSV feature of the image sample. Finally, the HSV features of each image sample were combined in a large matrix. Through the flow of this algorithm, spatial HSV histogram pixels of different kinds of defect images could be obtained, as shown in Figures 3–5.

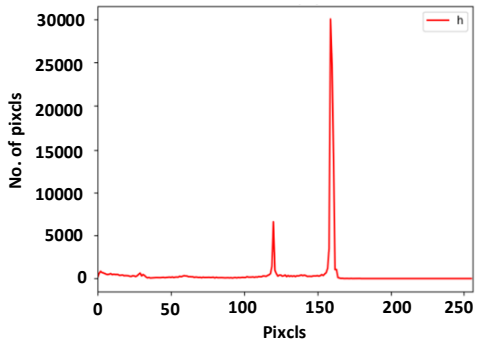
The features of the HSV images were composed of three parts: hue, saturation, and value. Figures 3–5 show the statistics in histograms of pixels of different kinds of defect images. In the above figures, it can be seen that in the HSV space, the main distribution interval of pixels in the H channel was 11, that of the pixels in the S channel was 22, and that of pixels in the V channel was 33. The distributions of the S channel and V channel pixels were obviously different for the different defect types. For example, the peak value in the S channel in the infrared image of a half-package defect appeared near the pixel value of 175, while the peak value in the S channel in the linear-mezzanine image appeared near the pixel value of 200. The obvious differences in different kinds of defect image features verified the effectiveness of HSV features in infrared feature extraction for defect detection, which laid the foundation for building an expert system by fusing a RF and support vector machine under the condition of small samples.

2.3 Construction of the expert system

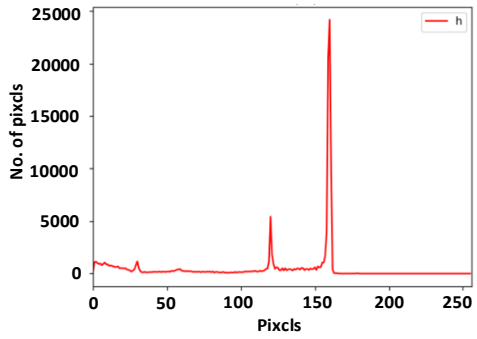
In actual industrial production processes, a large number of image samples are often unlabelled, and the number of defect pictures that have been manually labelled is limited. However, the improvement of the classification performance of a deep neural network often depends on a large number of labelled image samples. Therefore, it is

very important to construct an expert system with excellent classification performance by reasonably utilising the prior information in a small number of labelled defect image samples.

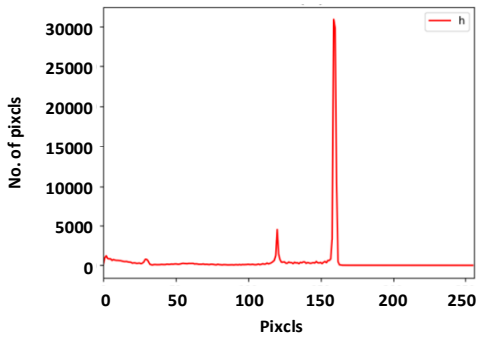
Figure 3 The distribution of pixels for the hue channel, (a) half-package (b) linear-mezzanine (c) dotted-mezzanine (d) normal (see online version for colours)



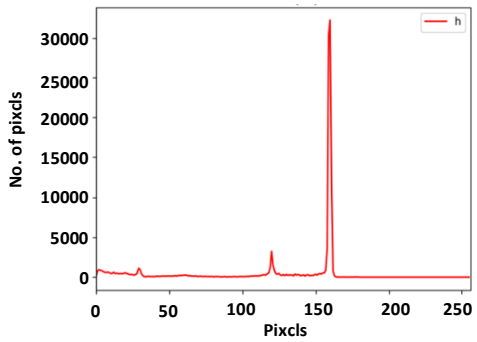
(a)



(b)

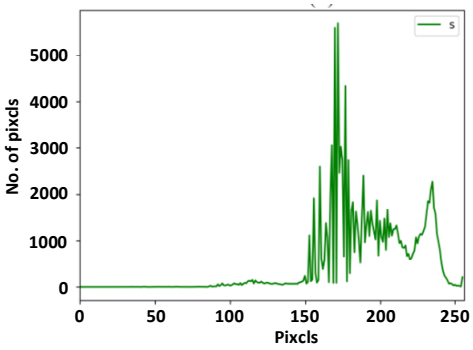


(c)

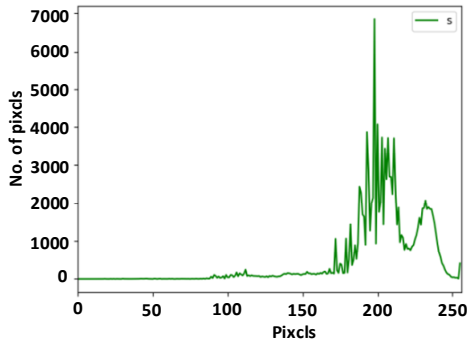


(d)

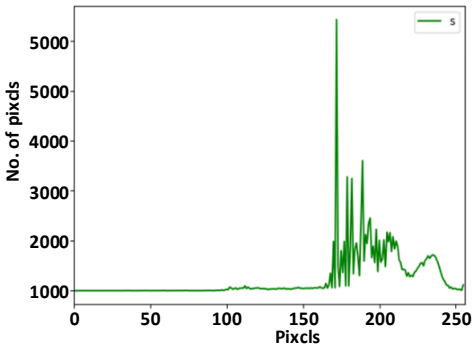
Figure 4 The distribution of pixels for the saturation channel, (a) half-package (b) linear-mezzanine (c) dotted-mezzanine (d) normal (see online version for colours)



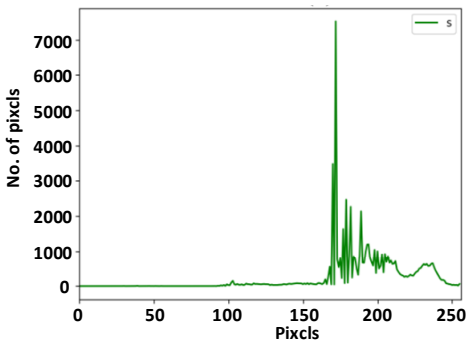
(a)



(b)

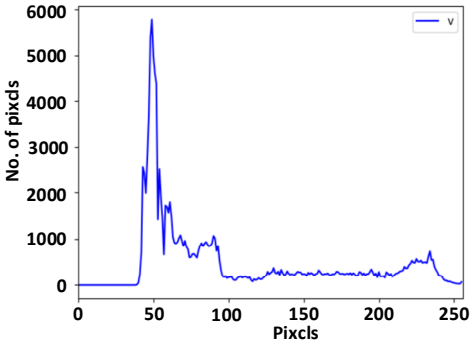


(c)

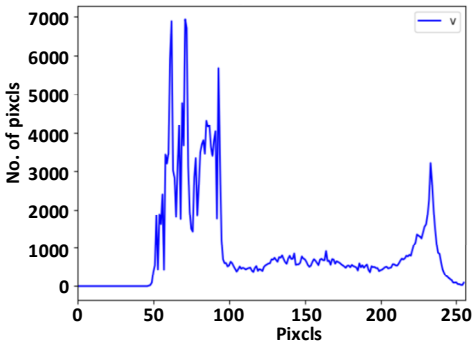


(d)

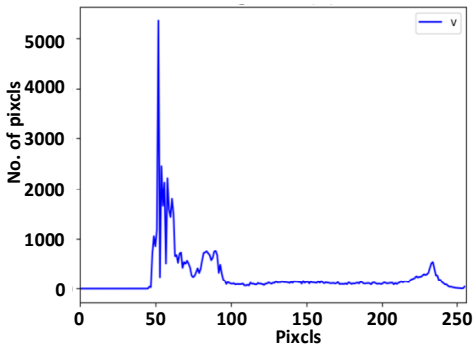
Figure 5 The distribution of pixels for the value channel, (a) half-package (b) linear-mezzanine (c) dotted-mezzanine (d) normal (see online version for colours)



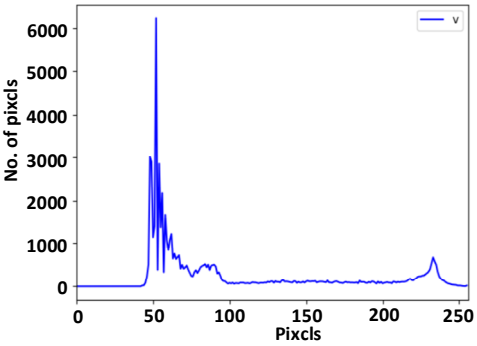
(a)



(b)



(c)



(d)

2.3.1 *Training the RF with small samples*

In order to improve the generalisation ability of the semi-supervised model when labelling small samples, it was necessary to make rational use of the prior information contained in a small number of labelled samples. Therefore, with the help of an existing small-sample classifier used for traditional machine learning, an expert system was constructed before the deep model, and a convolutional network assisted in realising the effective training of a large number of parameters. A RF and support vector machine were selected for the small-sample image classifier.

RF is a machine learning algorithm based on ensemble learning (Li et al., 2021). Each tree in the RF is trained with a subset of the data. The basic idea behind it is combining multiple decision trees when determining the final output, rather than relying on a single decision tree. Voting on the predicted values of each decision tree makes RF a powerful classification algorithm. It is much stronger than a single decision tree, and its generalisation ability is greatly improved. The algorithm's steps are as follows.

Its parameters are set in the following way.

The number of decision trees generated was set to n_estimators = 400, the maximum depth of the tree was set to max_depth = 7, a node divided into several types of samples was set to min i_samples_split = 2, and random seeds were set to random_state = 5.

The HSV colour feature combination training set $L = \{(x_1, y_1), (x_2, y_2), \ldots, (x_j, y_m)\}$ was extracted from m labelled samples.

Firstly, a new training set of m' samples ($m' \leq m$) was formed with the boostrap method for random samples, and n training sets S_n were obtained by repeating n operations. The purpose of this was to ensure that the training decision trees were not exactly the same, but also had certain similarities to improve the generalisation and accuracy.

Secondly, n base classifiers C_n were independently trained on each training set S_n by using decision trees.

Then, for the test sample x, n classification results $C_n(x)$ were obtained with the n classifiers mentioned above.

Finally, for each x, the final prediction result $C'(x)$ was obtained by voting:

$$C'(x) = \arg\max \sum_{i=1}^{k} I\left(C_i(x) = y\right) \tag{1}$$

where $C'(x)$ is the final output, $C_i(x)$ is the result of a single decision tree, I is the sigmoid function, and y is the output variable.

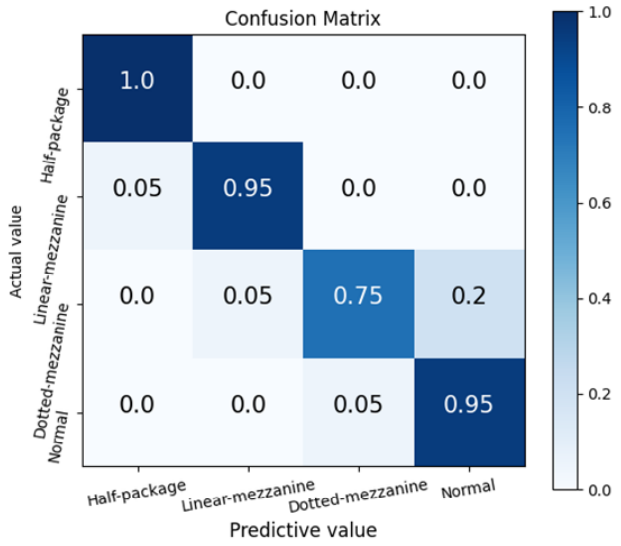
Table 1 The RF's classification and recognition accuracy

Type	*Precision (%)*	*F1-score (%)*	*Accuracy (%)*
Half-package	95.24	97.56	100
Linear-mezzanine	95	95	95
Dotted-mezzanine	93.75	83.33	75
Normal	82.61	88.37	95
Average	91.65	91.07	91.25

From the above steps, the RF classifier model M_{RF} could be obtained. The classification model M_{RF} was called by the

predict_proba function to predict the test set. Each sample confidence level $Q_{RF}{}^{u_i} = [Q_0, Q_1, Q_2, Q_3]$ was output, and $\max(Q_{RF}{}^{u_i})$ was taken to obtain a prediction label.

Figure 6 Confusion matrix for the RF (see online version for colours)



A small number of defect image samples that were manually labelled were divided into a training set and test set with a ratio of 2:1. The total number of labelled samples was 240, and there were four types of defect image samples, which were half-package defect samples, linear-mezzanine defect samples, dotted-mezzanine defect samples, and normal image samples. There were 40 image samples in each class in the training set and 20 image samples in each class in the test set. When the sampling ratio was 4:1, about 20% of the data were not sampled in each bootstrap sampling, and the RF used this part of the data to estimate the internal error. The trained RF classifier was verified on the test set, and its classification and recognition results are shown in Table 1; a confusion matrix of the visual accuracy is shown in Figure 6. The precision was the proportion of all predictions that were positive, the F1-score was the weighted harmonic average of the precision and recall, and the accuracy was the proportion of all predictions that were correct (including positive and negative).

It can be seen in Table 1 that the average accuracy of the RF's prediction was 91.25%. The precision for normal images was at least 82.61%. The accuracy for dotted-mezzanine defects was at least 75%. According to the analysis, defects in the point interlayer were small, and the overlap between the pixel interval in the histogram and that in normal images was relatively large. RF classifiers can be confusing. The pixel intervals in the histograms of the semi-sealed and line interlayer defects were quite different from those of the other two defect images, so the precision reached 95.24% and 95%, respectively.

2.3.2 *Training the small-sample support vector machine*

SVM is a new structured learning method based on structural risk minimisation. It aims to find the maximum margin for solving the normal vector W and displacement b to construct a decision boundary and classify the classifier with the decision boundary. The SVM classifier has the following training steps.

The HSV colour feature combination training set $L = \{(x_1, y_1), (x_2, y_2), \ldots, (x_j, y_m)\}$ is extracted from m labelled samples.

A linear kernel and RBF kernel are generally used, and kernel functions are selected according to Andrew Ng's guidelines:

1 If the number of features is large and is similar to the number of samples, a linear kernel is chosen for the SVM.

2 If the number of features is small and the number of samples is average (neither big nor small), a Gaussian kernel is chosen for the SVM.

3 If the number of features is small and the number of samples is large, it is necessary to manually add some features for the first case to be true.

The kernel function selected according to the above rule was and the penalty coefficient was $C = 2.0$.

Firstly, the appropriate kernel function $K(x, z)$ and penalty coefficient C were selected to construct and solve the optimisation problem:

$$\min_a \frac{1}{2}\sum_{m=1}^{i}\sum_{n=1}^{i} a_m a_n y_m y_n K(x_m, x_n) - \sum_{m=1}^{i} a_m \tag{2}$$

$$s.t \sum_{\substack{m=1 \\ 0 \le a_m \le C, m=1,2,\cdots,j}}^{i} a_m y_m = 0 \tag{3}$$

The optimal solution $a' = (a_1', a_2', \cdots, a_i')$ was obtained. Then, a positive component $0 < a'_n < C$ of a' was selected to calculate the following:

$$b' = y_n - \sum_{m=1}^{i} a_m' y_m K(x_m, x_n) \tag{4}$$

Finally, the decision function was constructed:

$$f(x) = sign\left(\sum_{m=1}^{i} a_m' y_i K(x, x_m) + b'\right) \tag{5}$$

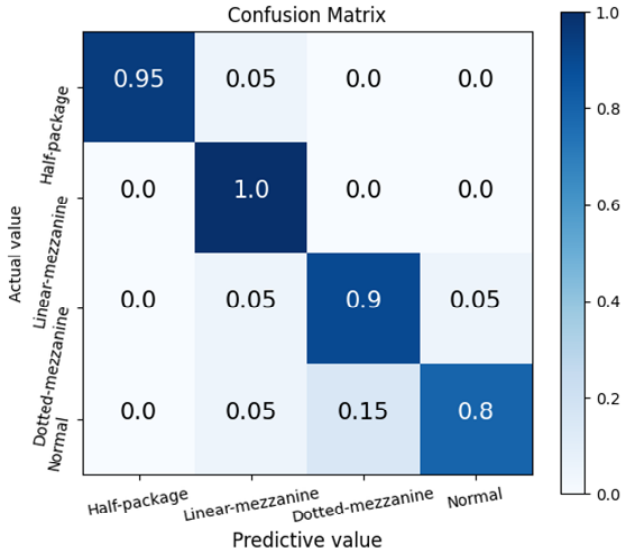
From the above steps, the RF classification model M_{SVM} could be obtained. The classification model M_{SVM} was called by the predict_proba function to predict the test set. Each sample confidence level $Q_{SVM}{}^{u_i} = [Q_0, Q_1, Q_2, Q_3]$ was output, and $\max(Q_{SVM}{}^{u_i})$ was taken to obtain a prediction label.

When this method was used to classify and identify the test set, the experimental training set and the test set were the same as those obtained with the RF classifier. The classification and recognition results are shown in Table 2, and a confusion matrix of the visual accuracy is shown in Figure 7.

Table 2 The SVM classifier's recognition accuracy

Type	*Precision (%)*	*F1-score (%)*	*Accuracy (%)*
Half-package	100	97.44	95
Linear-mezzanine	86.96	90.70	100
Dotted-mezzanine	85.71	87.80	90
Normal	94.12	86.49	80
Average	91.70	91.19	91.25

Figure 7 Confusion matrix for the SVM (see online version for colours)



It can be seen in Table 2 that the average accuracy of prediction with the SVM was 91.25%. The precision for dotted-mezzanine defects was at least 85.71%. At the same time, the F1-scores for the normal and dotted-mezzanine images were lower than those for the other two types. The analysis showed that the dotted-mezzanine defects were small, and the overlap between the pixel intervals in their histograms and those of the normal images was large. SVM classifiers can be confusing. Comparing Figure 6 and Figure 7, it can be seen that both the SVM classifier and the RF classifier showed good classification characteristics for the half-package and linear-mezzanine defect images. However, for the dotted-mezzanine and normal image samples, the classifiers showed different advantages. The accuracy of RF in predicting dotted-mezzanine defects was only 75%, while the SVM classifier had a prediction accuracy of 90% for this class. For images of normal categories, the classification ability of RF was better than that of the SVM, and the classification accuracies were 90% and 80%, respectively.

Because the classification accuracies of the RF and SVM could reach a certain precision and the dominant defect images in the classification were different, the trained RF model and SVM model were combined to build an expert system to prepare for the next marking process.

2.4 Expert system marking

Because different classifiers have different mechanisms, they are good at different fields of image classification. Therefore, these two classifiers with different performances were fused, and the fused model was used as an expert system to mark a large number of unlabelled defect image samples. Samples that were marked by the assignment were called labelled prediction samples. Predicting the label accuracy of the labelled samples is very important when training a neural network. The fused expert system had higher confidence than when labels were assigned by a single classifier. The specific fusion steps were as follows.

The expert system was constructed by using the trained RF classification model *MRF* and the support vector machine classification model *MSVM*, and the unlabelled sample set U was input into it for prediction. If the prediction tags were the same, the following was obtained:

$$y_{RF}{}^{u_i} = y_{SVM}{}^{u_i} \tag{6}$$

where u_i is a subsample of the sample set; $y_{RF}{}^{u_i}$ is the i^{th} sample of the candidate sample set U for predicting tags in the RF classifier (likewise with $y_{SVM}{}^{u_i}$ for the SVM classifier). This formula indicates that the prediction tags in the two classifiers were the same, but if the prediction tags were different, the following could be obtained:

$$y_{RF}{}^{u_i} \neq y_{SVM}{}^{u_i} \tag{7}$$

If the two prediction labels were different, it was necessary to calculate the confidence of both, and the following could be obtained:

$$\begin{cases} A = \max\left(Q_{RF}{}^{u_i}\right) - Second\left(Q_{RF}{}^{u_i}\right) \\ B = \max\left(Q_{SVM}{}^{u_i}\right) - Second\left(Q_{SVM}{}^{u_i}\right) \end{cases} \tag{8}$$

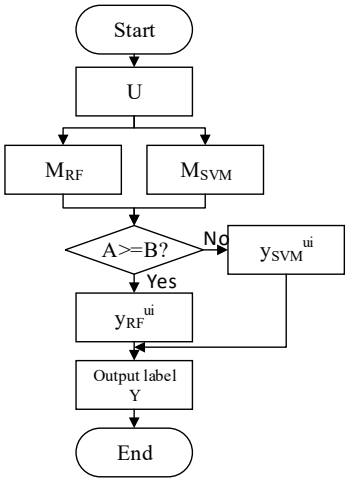
where $\max(Q_{RF}{}^{u_i})$ is the maximum prediction confidence of the RF classifier, $Second(Q_{RF}{}^{u_i})$ is the second largest prediction confidence of the RF classifier, A is the difference between them, and $\max(Q_{RF}{}^{u_i})$, $Second(Q_{SVM}{}^{u_i})$, and B apply in the same way to the SVM classifier. It was necessary to compare the differences in confidence, and the following could be obtained:

$$Y = \begin{cases} y_{RF}{}^{u_i}, & \text{if } A \geq B \\ y_{SVM}{}^{u_i}, & \text{if } A < B \end{cases} \tag{9}$$

where Y is the selection tag, and the difference in the RF classifier's prediction confidence is greater than the difference in the SVM classifier's prediction confidence. The RF prediction tag is selected; otherwise, the SVM prediction tag is selected. Here, the confidence difference is

used for comparison. In order to highlight the difference in predictions, instead of simply comparing the highest confidence, a specific flow of the algorithm was used, as shown in the following flowchart.

Figure 8 Flowchart of the expert system algorithm



The expert system was used to test the classification performance on the test set. The experimental conditions of the training set and the test set were the same as those for the RF classifier. The classification and recognition results are shown in Table 3, and the confusion matrix for the visual accuracy is shown in Figure 9. The expert system was built based on the trained RF and SVM to mark a large number of unlabelled defect samples.

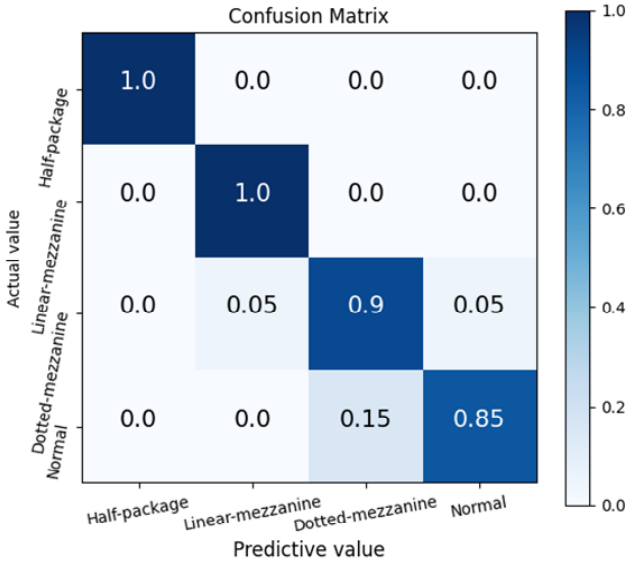
Table 3 The expert system's classification and recognition accuracy

Type	*Precision (%)*	*F1-score (%)*	*Accuracy (%)*
Half-package	100	100	100
Linear-mezzanine	95.24	97.56	100
Dotted-mezzanine	85.71	97.80	90
Normal	94.44	89.47	85
Average	93.85	93.71	93.75

It can be seen in Table 3 that the average accuracy of the expert system was 93.75%, which was 6.25% and 2.5% higher than those of the RF and SVM, respectively. Secondly, as shown in the confusion matrix analysis in Figure 9, the expert system combined the advantages of the RF and SVM to improve the classification accuracy. In addition, due to the greater confidence difference of the support vector machine for normal types, the classification and recognition of the expert marking system for normal classes were biased towards the support vector machine, which had the accuracy of two separate classifiers.

Compared with the RF expert marking system and SVM expert marking system, the accuracy of the RF + SVM expert marking system was improved by 2.5% and 2.5%, respectively, which ensured the accuracy of the subsequent migration of knowledge to the convolutional neural network.

Figure 9 Confusion matrix for the expert system (see online version for colours)



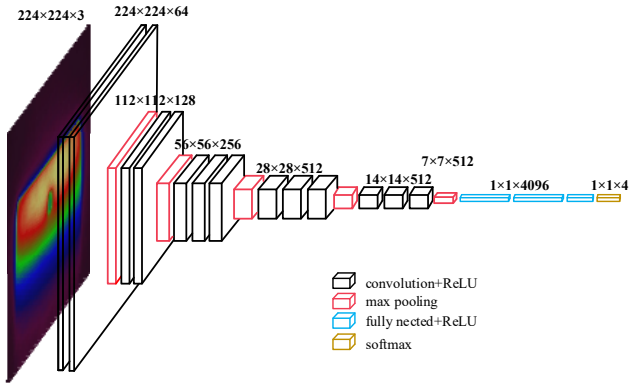
2.5 *Training the convolutional neural network based on knowledge transfer*

Feature extraction with a convolutional neural network CNN) is very powerful, and the CNN is the most commonly used network model for defect classification (Tao et al., 2021; Zhao and Shi, 2021). The training of a deep convolutional neural network often requires a large number of image samples. A large number of labelled prediction samples are obtained by building an expert system based on multi-classifier fusion to build and train the neural network. The VGG16 (visual geometry group) network structure is regular, and its recognition performance can be effectively improved by increasing the network's depth (Xu et al., 2022). And convolution tandem has fewer parameters and more nonlinear transformations than a single convolution layer. And after each pooling, the number of convolution channels will be doubled, so that more features can be retained, more comprehensive and detailed information can be provided, and things can be described more accurately, which improves the recognition rate to a certain extent. Considering the small number of image datasets available for experiments, this study used the transfer learning method to create the trained VGG16 model, and its network structure is shown in Figure 10.

As shown in Figure 10, first, the size of the image was cut to 224 × 224 × 3 (224 was the pixel value of the image, and 3 was the colour of the image), and the cut image was horizontally turned over according to the probability $P = 0.5$, then normalised. Then, SGD was used to optimise the VGG16 network model. Finally, a new fully connected network layer was selected at the feature extraction layer, where the first and second fully connected layers used ReLU as an activation function, and each activation function was followed by a dropout connection to prevent

model over-fitting. In Figure 10, the black box represents the convolution layer for feature extraction. The red box represents the pooling layer, which was used to reduce the dimensions and extract the main features of the image. The blue box represents the fully connected layer, and each node in the fully connected layer was connected with each node in the previous layer to synthesise the output characteristics of the previous layer. The yellow box represents softmax layers, which were used to calculate the probability that each element was taken.

Figure 10 Diagram of the network structure of VGG16 (see online version for colours)

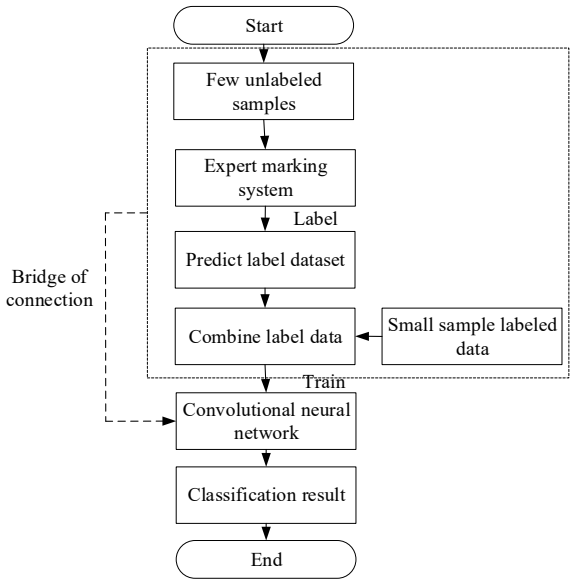


VGG16 was composed of 13 convolution layers, five pooling layers, and 3 fully connected layers; the size of the convolution core was 3 × 3, and the size of the pooling layers was 2 × 2. Because this study involved fine-tuning VGG16, the parameters were fine-tuned only in the fully connected layer as follows: The linear parameters were (25088, 512), (512, 256), and (256, 4), respectively, and four classification results were output; the dropout parameter was set to p = 0.3; the softmax parameter was set to dim = 1.

The convolutional neural network is directly trained with small sets of labelled samples. Convolutional neural networks have some problems, such as low accuracy and slow convergence. Therefore, this study discusses an image classification and recognition method using a convolutional neural network based on the knowledge transfer of classifiers, and the flow used for recognition is shown in Figure 11.

Convolutional neural networks have excellent end-to-end characteristics. The convolution layer and pooling layer can accurately extract effective information from images, and they show good classification performance with a classifier. However, neural networks have a large number of parameters that need to be trained. Therefore, training with a small number of labelled defect image samples will cause the over-fitting phenomenon. With the help of the expert system, knowledge transfer from the fused classifier that was trained with a small number of labelled samples to the deep convolutional neural network was realised. A bridge was established between the traditional machine learning classifier and the convolutional neural network model. After knowledge transfer, the fully connected layer of the convolutional neural network was effectively trained. The fast and accurate detection of defects in thermal images of packaging bag seals was realised.

Figure 11 Flowchart of the knowledge transfer algorithm



3 Experimental verification

To verify the effectiveness of the proposed method, a simulation experiment was carried out by using packaging bags in actual production. To ensure the rigour of the experiment, the test set continues to use the above test set, The test set includes four categories (half-package, linear-mezzanine, dotted-mezzanine defects, and normal) with 20 images in each class. The same test set can enhance the contrast before and after. Thermal images of sealed packaging bags were collected using a Hikvision H10 thermal imager. The thermal images of the packaging bags could be categorised according to four types of defects: half-package, linear-mezzanine, dotted-mezzanine, and normal. The experiment was based on PyTorch 1.7.1 and the Python language, and it was completed in the environment of an Intel Core i7-10750H CPU with 16 GB of RAM and 64-bit Windows 10. The predicted tagged image samples and a small number of tagged samples predicted by the expert system were input into the convolutional neural network with VGG16 for training. And the accuracy rate is used as the evaluation index.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

where *Acc* (accuracy) refers to how many judgments are correct among all the judgments, that is, the positive judgment is positive, and the negative judgment is negative; *TP* (true positive) is the number of positives classified as positive; *TN* (true negative) is the number of negatives that were judged as negative; *FP* (false positive) is the number

of false positives that are classified as positive; *FN* (false negative) the number of positives misjudged as negative.

The results are shown in Table 4.

Table 4 Classification and recognition accuracy of VGG16

Number of samples	*Accuracy (epoch = 25)*	*Loss*
160	78.75%	0.9573
400	90.00%	0.8346
800	93.25%	0.8126
2,000	96.25%	0.7832

As shown in Figure 4, when the number of samples was 160, this meant that the neural network had no training set of labelled prediction samples. The accuracy when there were 25 rounds of training with the original small labelled sample set was only 78.75%. When the number of samples involved in knowledge transfer reached 400, great changes took place, and the accuracy increased by 11.25% to 90%. With the increase in knowledge transfer, the accuracy was 93.25% when the number of samples was 800. When the number of samples was 2,000, the training time was about two hours, the average time taken to identify a single image was 0.1223 seconds, the accuracy was improved to 96.25%, and the loss function value was reduced from the initial 0.9573 to 0.7832. Therefore, it could be ascertained that the number of samples in the training set had a great influence on training. With the increase in knowledge transfer, the accuracy of fine-tuning VGG16 with the test set continuously increased. This showed that the fused knowledge of the RF and SVM with the small sample was successfully transferred to the convolutional neural network, which improved the low classification accuracy and mitigated the over-fitting phenomenon of the neural network when using small labelled sample sets.

In order to more intuitively observe the changes in the loss function and accuracy in the training process of VGG16 under the conditions of a small sample and knowledge transfer, curves of the training process were obtained and are shown in Figures 12–14.

It can be seen in Figures 12–14, that the condition of 160 dataset capacity and 25 rounds of training, the accuracy rate is only 78.75%, and there is no convergence. When the training times are 50 rounds, the training gradually converges, and the accuracy rate is 87.50%. The VGG16 network has converged under the condition of 2,000 dataset capacity and 25 training rounds. The accuracy was improved by 9% to 96.25%. It can be seen that the expert marking system constructed with the RF and SVM successfully migrated the knowledge learned with a small sample of data to a convolutional neural network and improved the network.

In order to further verify the effectiveness of this method, a comparative experiment was carried out together with the methods used in Wang and Li (2021) and Cai et al. (2022), and the recognition accuracy is shown in Table 5.

Figure 12 The training process of the VGG16 network with a sample size of 160 (epoch = 25) (see online version for colours)

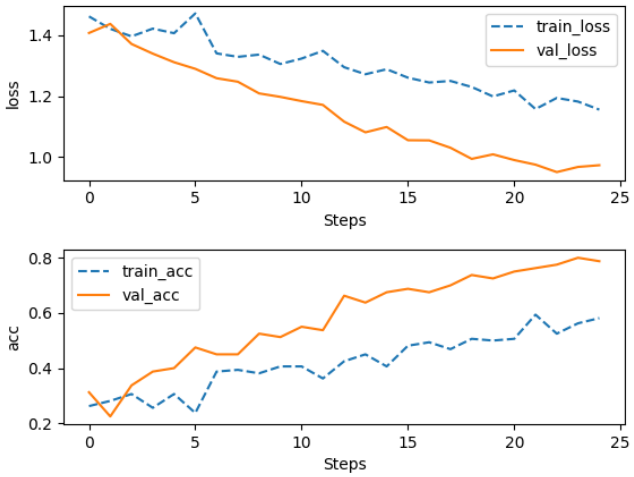


Figure 13 The training process of the VGG16 network with a sample size of 160 (epoch = 50) (see online version for colours)

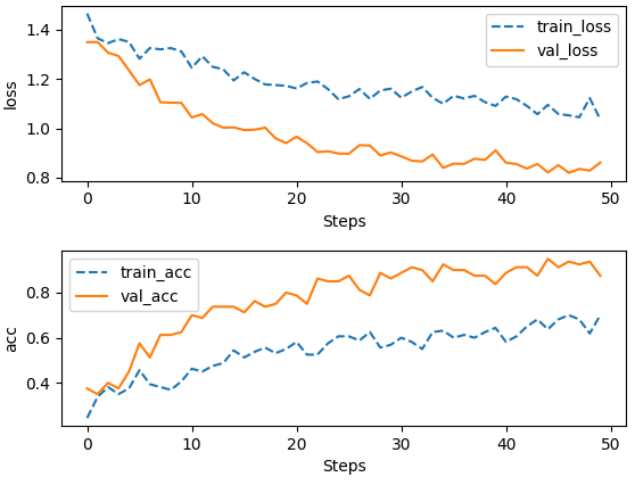
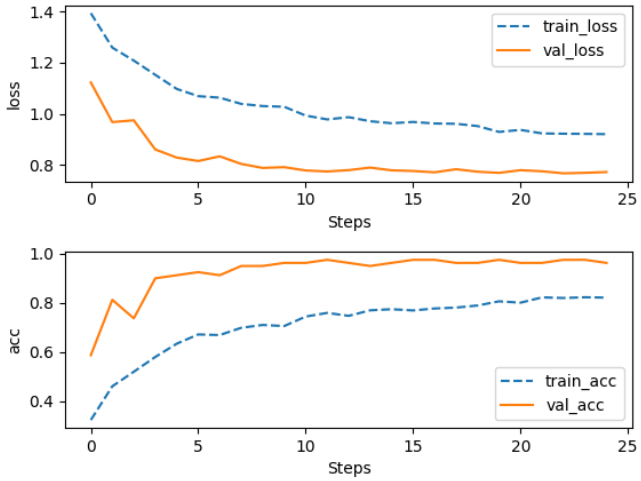


Table 5 Comparison of the accuracy

Method	*Accuracy (%)*
Wang and Li (2021)	93.75
Cai et al. (2022)	78.75
The proposed method	96.25

It can be seen in Table 5 that the knowledge learned by the expert marking system constructed with the RF and SVM with a small sample of data was transferred to VGG16. The accuracy rate was 96.25%, while the accuracy of the classifier in Wang and Li (2021) was 93.75%. Cai et al. (2022) had an accuracy of 78.75% with the original sample. Thus, it was verified that the method in this study is effective and feasible.

Figure 14 The training process of the VGG16 network with a sample size of 2,000 (epoch = 25) (see online version for colours)



4 Conclusions

In order to realise the intelligent production of packaging bags, this paper studies a packaging bag sealing defect detection algorithm based on knowledge transfer. This paper combines traditional machine learning with deep learning, and proposes a convenient way to establish datasets. RF and SVM are trained with supervised training using a small number of labelled samples, and an expert system is constructed by fusing the two classifiers to accurately obtain a large number of prediction samples. At the same time, the expert system can easily expand the team training set and strive to expand different products into a suitable dataset. On this basis, the knowledge was transferred to the deep convolutional neural network (VGG16) for training to realise the fast and accurate detection of packaging bag sealing defects. The accuracy of the RF and SVM classifier is 91.25%, and the accuracy of the expert system is 93.75%, which shows the effectiveness of the expert system. The accuracy of the neural network trained by 2,000 samples is 17.50% higher than that of 160 samples, which shows the necessity of the expert system. This provides a new idea for the rapid establishment of defect detection systems for different products in the future, and can effectively improve the training rate, compatibility and detection accuracy of packaging defect detection systems.

Acknowledgements

We would like to gratefully acknowledge Professor Huosheng Hu from the University of Essex, UK, for his guidance on the English description of this paper.

References

Beshai, H., Sarabha, G.K., Rathi, P., Alam, A.U and Deen, M.J. (2020) 'Freshness monitoring of packaged vegetables', *Applied Sciences*, Vol. 10, No. 21, p.7937.

Cai, J., Du, J.C., Wang, Q and Zhou, H.R. (2022) 'Face emotion recognition based on VGG16 network', *Application of Electronic Technology*, Vol. 48, No. 1, pp.67–70+75.

Chai, T.Y. (2020) 'Development direction of industrial artificial intelligence', *Chinese Journal of Automation*, Vol. 46, No. 10, pp.2005–2012.

Chai, T.Y., Liu, Q., Ding, J.L., Lu S.W., Song, Y.J and Zhang, Y.J. (2022) 'Research prospects of a new model of intelligent optimizing manufacturing in process industry driven by industrial internet', *China Science: Technological Science*, Vol. 52, No. 1, pp.14–25.

Chen, W.G., Lu, J.C and Liao, J.P. (2020) 'TRIZ field model and deep learning in classification of egg yolk crisp packaging patterns', Equipment Manufacturing Technology, Vol.5, pp.87-90+97.

Chen, X.C., Fang, Y.L., Du, S.C., Lv, J and Wang, Y. (2021) 'Rapid packaging defect detection method based on deep learning', Machine Design & Research, Vol.37, No.6, pp.165-169+178.

Hu, J., Jiang, Q., Cong, R., Gao, W. and Shao, F. (2021) 'Two-branch deep neural network for underwater image enhancement in HSV color space', *IEEE Signal Processing Letters*, Vol. 28, No. 1, pp.2152–2156.

Li, L., Tian, X. and Weng, Y.L. (2021) 'Land cover classification based on polarimetric SAR and optical image features', *Journal of Southeast University (Natural Science Edition)*, Vol. 51, No. 3, pp.29–534.

Mirza, A.A., Masoomian, M., Shakooie, M., Khajavi, M.Z and Farhoodi, M. (2021) 'Trends and applications of intelligent packaging in dairy products: a review', *Critical Reviews in Food Science and Nutrition*, Vol. 62, No. 2, pp.383–397.

Rodrigues, C., Souza, V.G.L., Coelhoso, I and Fernando, A.L. (2021) 'Bio-based sensors for smart food packaging-current applications and future trends', *Sensors*, Vol. 21, No. 6, p.2148.

Sa, J.M., Li, Z.H., Yang, Q.J and Chen, X. (2020) 'Packaging defect detection system based on machine vision and deep learning', in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, IEEE, Shanghai, China, pp.404-–408.

Sun, N., Guan, Y.H., Cui, Y.Y., Luo, Y.T. and Huang, G. (2020) 'Appearance defect detection of cigarette packaging based on support vector machine', *Software*, Vol. 41, No. 1, pp.205–210.

Tao, X., Hou, W. and Xu, D. (2021) 'A review of surface defect detection methods based on deep learning', *Chinese Journal of Automation*, Vol. 47, No. 5, pp.1017–1034.

Vu, T.T.H., Pham, D.L and Chang, T.W. (2023) 'A YOLO-based real-time packaging defect detection system', *Procedia Computer Science*, Vol. 217, pp.886–894.

Wang, Y. and Li, Y.H. (2021) 'A semi-supervised classification algorithm based on collaborative training', *Journal of Central China Normal University (Natural Science Edition)*, Vol. 55, No. 6, pp.1020–1029.

Wang, X.L. and Ma, Y.B. (2022) 'Improving Yolov5's damage detection method for medical waste packaging bags', *Computer Engineering and Design*, Vol. 43, No. 12, pp.3513–3520.

Wu, H.R., Chen, X.X and Gao, A. (2023) 'Research on food packaging defect detection algorithm based on deep convolutional neural network', *Intelligent Computer and Applications*, Vol. 13, No. 3, pp.10–15.

Xu, S.J., Hao, M., Meng, Y.B., Liu G.H and Han, J.Q. (2022) 'Crack detection method based on feature-enhanced overall nested network', *Advances in Laser and Optoelectronics*, Vol. 59, No. 10, pp.90–101.

Zeng, X.Y. (2022) *Research on Product Packaging Defect Detection Method based on Machine Learning*, Anhui Polytechnic University, Wuhu, China.

Zhao, B. and Shi, Y.X. (2021) 'Surface defect detection of high-precision optical components based on convolutional neural networks', *Laser Journal*, Vol. 42, No. 11, pp.185–189.

Zhou, G.X. and Li, P. (2022) 'Image detection method for visual defects at the sealing of tobacco packaging based on machine learning', *Bulletin of Science and Technology*, Vol.3 8, No. 3, pp.47–51.