Multi-Stream Deep Convolutional Neural Network for PET Preform Surface Defects Detection and Classification

TAOCHUAN ZHANG\(^1\) AND CHUNMEI DUAN\(^2\)
\(^1\)College of Mechanical and Electrical Engineering, Foshan Polytechnic, Foshan 528137, China
\(^2\)School of Electronic Information, Foshan Polytechnic, Foshan 528137, China
Corresponding author: Chunmei Duan (49977405@qq.com)

This work was supported in part by the Special Research Project in Key Fields of Artificial Intelligence in General Colleges and Universities of Guangdong Province under Grant 2019KZDZX1029, in part by the Characteristic Innovation Research Project of University Teachers of Foshan City under Grant 2019XZZ06, in part by the Characteristic Innovation Project in General Colleges and Universities of Guangdong Province under Grant 2019GKTSCX117, and in part by the Special Fund for Science and Technology Innovation Strategy of Guangdong Province, in 2020, under Grant pdjh2020b1221.

ABSTRACT Due to the influence of technology factors, various defects will appear in the production process of PET (Polyethylene Terephthalate) preform, and affect the PET preform quality. In order to meet the requirements of the quality inspection efficiency and accuracy for PET preform, a novel multi-stream deep CNN (Convolutional Neural Network) model is performed to effectively identify and classify PET preform surface defects for the first time in this paper. In recent years, the method based on multi-stream feature fusion has a good application prospect in defect detection. In order to solve the problem of PET preform surface defects detection and classification, we study the deep CNN structure and propose a multi-stream deep network structure, which makes full use of the rich multi-feature information from different network structures. Two different deep CNNs are respectively trained to extract features from the original image, and one of them is trained to extract feature from the corresponding gradient image. We adopt wavelet transform fusion strategy to realize feature fusion and input them into ECOC-SVM (a classified Error Correcting Output Codes model using Support Vector Machines as binary learner) model for detection and classification. Using Bayesian optimization function to optimize the hyperparameters is performed to choose the best performance configuration. Through experimental analyses, the higher detection accuracy of the proposed method is 98.5\%, and it is proved that the proposed model has good convergence, accuracy, stability and generalization ability.

INDEX TERMS PET preform, surface defect, multi-stream network, deep CNN, wavelet transform fusion.

I. INTRODUCTION

Due to the growing demand of the production and consumption for cosmetic bottles, beverage and mineral water bottles, medical plastic packaging and edible oil bottles formed, PET preform with the advantages of non-toxic, uniform quality distribution and good insulation has been widely used in the field of daily chemical products packaging. Therefore, the quality of PET preform is very important significance and can greatly affect the product performance [1]. However, the bubble defect is randomly formed on the preform due to the air inside the injection molding machine, and the black spot defect is formed because of the black particles produced on the bottle blank by the degraded solution, and the irregular scratches are produced on the side wall of the PET preform because of the damage and the collision of the mold surface. The above three types of defects are very common. In addition, the PET preform surface defects also include wire drawing, water mark, dent and other defects.

In many manufacturing industries, most enterprises mainly rely on manual visual inspection. Due to reliance on human experience, the stability and accuracy of detection cannot be guaranteed. In addition, the site has high ambient temperature and harmful dusts, which are harmful to the occupational...
II. RELATED WORKS

Feature extraction is very important in the process of visual detection, which directly affects the detection accuracy and even leads to detection failure. Many scholars have studied feature extraction methods applied to the detection and classification for different materials. G. Kalairasai et al. [4] proposed a hyperspectral image classification method based on Multi-Layer Perceptron Classification technique in order to improve the classification accuracy and minimize the classification time. After preprocessing hyperspectral image, the texture, color and target features of hyperspectral image were extracted by using the Gaussian distribution scale invariant feature transform in the second hidden layer. In the third hidden layer, the Euclidean distance between the extracted feature and the tested feature was measured. Finally, feature matching was performed in the output layer to realize hyperspectral image classification. The model extracted different features through three layers to realize classification based on feature matching and achieved the expected effect. At the same time, the effects of various parameters on the experiment were studied, but there was no specific comparison with other methods.

R.A. Saputra et al. [5] proposed a method of feature extraction for text analysis based on gray level co-occurrence matrix. The feature values of this method included contrast, energy, homogeneity, entropy and correlation. And K-nearest neighbor (KNN) algorithm was used to classify rice leaf diseases. In this paper, different parameter values were analyzed, and the relatively best classification accuracy was obtained, but the accuracy was generally low. Other methods needed to be studied to obtain higher classification accuracy.

T. Wang et al. [6] proposed a convolutional neural network to automatically extract features for distinguishing between the defect free and defective image to achieve product quality control, which could extract powerful features with less prior knowledge about the images for defect detection. The model could distinguish defect free image and defect image very well. However, the classification of defect categories was not implemented.

Although there are few articles on the surface defects detection and classification for PET preform based on deep learning, there are many researches on other materials based on deep learning or quality inspection of PET preform based on non-deep learning method. CHEN Hong-cai et al. [7] proposed a deep CNN detection model in order to detect the appearance defects of medical glass bottles accurately and quickly, and the multi-scale feature information of glass bottles was extracted by connecting and standardizing the shallow and deep feature vectors of the neural network structure. The defect detection method of medicinal glass bottle could accurately classify different defects. But the average precision was slightly lower.

A. Laucke et al. [8] proposed an algorithm to detect defective products for PET preform based on image segmentation, digital production, corrosion smoothing and the most

The main difference between the multi-stream network model for PET preform proposed in this paper and the previous two-stream network model for aluminum profile is that the backbone network and the number of streams are different.

The contributions of the surface defects detection and classification for PET preform based on multi-stream deep CNN model can be summarized as follows:

1. Through the multi-stream network model, a multiple feature fusion mechanism is realized by combining the high-level semantic features with intermediate contour features of deep CNN.

2. According to the common evaluation indexes, the quantitative and qualitative analyses of defect detection for PET preform data set are carried out, which verify the effectiveness of this proposed method on improving the accuracy of automatic detection for PET preform surface defects.

This work is organized as follows: Section 2 introduces the related works about deep learning and multi-feature fusion on the surface defects detection of various materials. Section 3 describes the collection and enhancement of PET preform data sets. Section 4 provides a proposed detailed method including multi-stream network model structure, feature extraction and multi-feature fusion. Section 5 is about the experimental analysis and result discussion. Section 6 presents the limitations of the method and the future research direction. Section 7 presents the conclusions.
effective filter. The authors made experimental analyses on the detection of color uniformity, moisture rings, impurities and neck defects for PET workpiece respectively. However, the classification precision of different defect types had not been studied.

Although the above methods have achieved good results in other materials, there are still some problems to be solved on detection and classification of PET preform surface defects. The single feature expression of surface defects is not obvious, which leads to low recognition rate. Considering the above problem, the multi-stream deep CNN model can be adopted to overcome the difficulties of different feature representations by combining shallow and deep features which include the edge local feature details and the global contour feature for PET preform.

There are a lot of researches on multi-stream or two-stream network for detection and classification of various materials. S. Hao et al. [9] proposed a hyperspectral image classification method which consisted of two-stream architecture and fusion scheme. One stream encoded the spectral value of each input pixel with a stacked denoising automatic coder, and the other stream encoded the corresponding image patch which was processed by deep CNNs. The prediction probabilities of the two streams were fused by adaptive class weights, which could be obtained by the fully connected layer. This paper introduced a novel two-stream architecture based on SdAE and CNNs, which were time consuming. Therefore, the proposed two-stream architecture needed further research to explore a more effective calculation scheme.

Y. Yu et al. [10] proposed a feature representation method and classifier which consisted of two-stream deep network for aerial scene classification, which takes pretrained CNN as the feature extraction network to learn deep features from the original RGB stream and its corresponding saliency stream. Q. Yan et al. [11] proposed a two-stream convolution network to learn more effective feature representations for predicting the image quality, which included two subnetworks with images and gradient images as input respectively. The model adopted two stream network structures to carry out feature extraction and feature fusion for different forms of images. And the effectiveness and superiority of the model were proved comparing with other methods. However, there were no in-depth studies of feature fusion methods in more detail, and no comparisons of multiple feature fusion methods.

F. Ren et al. [12] proposed a multi-stream neural network based on background knowledge to make up for the information imbalance or shortage caused by the limitation of training corpus. The base stream retained the original sequence information, and the other stream was based on background knowledge. The features based on background knowledge could make up for the lack of information, so that it was necessary to extract more useful background to expand the knowledge of more comprehensive external corpus. At the same time, different feature fusion strategies could also be further studied.

C. Ge et al. [13] proposed a deep learning method based on multi-scale feature learning of segmented tissue regions to solve the detection problem of Alzheimer’s disease, which used several parallel three-dimensional multi-scale convolution networks. X. Liu et al. [14] proposed a new pixel level classification method and depth multi-level fusion network, which was to classify multi-sensor data according to the characteristics of multi-source images. For the above two literature analyses, the data set can be expanded by various methods to have more real or richer data for training, so as to reduce the phenomenon of over fitting and solve the problem of small samples.

According to the above analyses, the multi-stream deep neural network structure has many effects on image detection and classification.

### III. DATA ACQUISITION AND PREPROCESSING

#### A. DATA ENHANCEMENT

In order to prevent low detection accuracy and low fitting degree of detection model caused by insufficient training samples, the data is enhanced which effectively solves the problem of insufficient training samples [15]. The original training samples of PET preform are adjusted by brightness and rotated to enlarge the data. Brightness adjustment is to adjust the gray level of the image by nonlinear mapping in the gray range before and after the specified transformation. The rotation operation is to rotate the image randomly at any angle [16], [17]. Figure 1 shows the data enhancement diagrams of PET preform image.

![Enhancement Diagrams](image)

**FIGURE 1.** The enhancement diagrams of PET preform image.

#### B. GRADIENT IMAGE OF ORIGINAL IMAGE

The gradient image of PET preform original RGB image can better represent the edge information of defect area and enhance the prominent change between defect area and uniform background. Moreover, using gradient image and original image for feature fusion can improve the detection performance [18]. The defect area of PET preform image usually has the characteristics of large gray value change rate,
so the gradient algorithm can eliminate most of the redundant background information and retain the edge information with high significance [19]. Sobel operator with a kernel of size 3 × 3 is used in this paper, which slides on the image in two directions respectively [20]. The original RGB image and the corresponding gradient image are shown in Figure 2.

Using Sobel operator, the original RGB image of PET preform represented by \( f(x, y) \) is calculated to obtain the corresponding gradient image represented by \( G(X, Y) \). The calculation formula is shown in formula 1.

\[
G(X, Y) = [G_x, G_y]
\]

therein, \( f(x, y) \) is the image gray value in the position of \((x, y)\), \( G_x \) and \( G_y \) are the convolution results of the original image with Sobel operators on horizontal and vertical directions respectively, and \( G(X, Y) \) is the final gradient image [21]. The calculation formulas of gradient amplitude and direction are shown in formula 2 and 3 respectively.

\[
|G(x, y)| = \sqrt{G_x^2 + G_y^2}
\]

\[
\theta = \arctan\left(\frac{G_x}{G_y}\right)
\]

IV. PROPOSED METHOD
A. MULTI-STREAM DEEP CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE
The typical deep CNN usually contains many layers, which can automatically extract useful features and classify them. However, the network cannot achieve satisfactory prediction performance for PET preform. In order to solve this problem, we propose a multi-stream network model with ECOC-SVM classifier based on deep CNN.

Firstly, the structure uses two different pretrained deep CNN as feature extractors to learn shallow and deep features of the original RGB image respectively, and fuses these two sets of features. And then, one of the trained deep CNNs is used to extract the features of the gradient image corresponding to the original image. The extracted feature is fused with the previous fused feature again to get the final feature fusion. Finally, the feature is input into the multi-class classifier to get the classification results. ECOC-SVM classifier is used to replace the softmax classification layer of CNN to solve the problem of insufficient generalization ability of image classification model. In this paper, the validation experiment will be carried out on the PET preform data set to verify the effect of image classification using ECOC-SVM classifier instead of softmax classifier based on multi-stream network. The multi-stream network structure is shown in Figure 3.

We should select the appropriate deep CNN for PET preform surface defects detection and classification to ensure an acceptable accuracy of the network. The VGG16 and GoogLeNet structures are selected as the trained deep CNNs. The original image features are extracted from inception_4c-output layer of GoogLeNet and conv5_3 layer of VGG16 respectively, and fused according to the wavelet transform fusion strategy. And then, the conv5_3 layer of VGG16 is used to extract feature from the corresponding gradient image of the original image. Finally, the feature is fused with the previous fused feature by wavelet transform fusion strategy.

The VGG16 structure increases the depth of CNN to 16 layers including 13 convolution layers and 3 fully connected layers, which has strong feature extraction ability. It consists of five groups of convolution operations [22]. Each group has two or three convolution layers, which adopt convolution kernel of size 3 × 3 and realize consecutive convolutions at two to four times so as to achieve the same receptive field effect of a large convolution core. After convolution, the maximum pooling operation is implemented which adopts a kernel of size 2 × 2 and stride value of 2 [23].

The GoogLeNet structure designs the network structure of automatic selection of Inception blocks formed by stacking them [24]. The Inception blocks are constructed by the parallel combination of the convolution layers of three convolution kernels with different sizes and the maximum pooling layer to increase the network adaptability to different scales [25]. Because the salient part of the image may have a great size change, the convolution operation must choose the correct kernel size. For example, the larger convolution kernels should be used for more global information, and the smaller kernels should be used for local information. It is better to run multiple sizes of convolution kernels at the same level to make the network wider rather than deeper. The structure of the Inception block is shown in Figure 4.

The input image is processed by convolution operation of the Inception block structure with a kernel of size 1 × 1, which can be used for feature extraction and dimension reduction of output channel. And, three different filters with a kernel of size 1 × 1, 3 × 3 and 5 × 5 and max pooling operation are adopted. And, the four branches of Inception through an aggregation operation are merged [26], [27].

GoogLeNet performs convolution operations firstly and includes three groups of Inception subnetworks with a total of nine Inception blocks [28].

In order to make the proposed network structure more suitable for PET preform surface defects detection and classification, the deep features of the original image are obtained through VGG16 structure, and the middle features of the original image are obtained through GoogLeNet.
structure, and then the features are fused to take into account the middle-level feature and the semantic feature information of the defect target object.

Another branch of the multi-stream structure is to extract deep features from the gradient image through VGG16 structure. The gradient image, having better robustness under different illumination conditions, is selected as another subnetwork input for feature extraction at the same time. And, this feature reflects the high-frequency information of an image significantly and contains the most important detail features, so the defects of the PET preform are identified effectively and easily. The classifier adopts ECOC-SVM classifier having a certain tolerance and correction ability to errors.

**B. FEATURE EXTRACTION**

The feature extraction processes of the three subnetworks are from the shallow feature to the middle feature to the deep feature, which can extract the shallow feature such as color, edge and texture, the middle feature such as shape and contour, and the semantic feature information of the defect object.

Two of the extracted features come from the original image and the corresponding gradient image by thirteen convolution layers and four maximum pool layers. The convolution layer has convolution kernel with size of $3 \times 3$ and stride of 1. The size of pooling unit is $2 \times 2$. The third feature is extracted from the original image by two convolution layers and five Inception blocks. The first convolution layer uses convolution kernel with size of $7 \times 7$ and maximum pool layer with size of $3 \times 3$. The second convolution layer uses convolution kernel with size of $3 \times 3$. The ReLU (Rectified Linear Unit) operation is performed after each convolution. Different sizes of convolution kernels are used to obtain different sizes of receptive fields in the Inception block, and the fusion of different scale features can be obtained by splicing them.

Taking the PET preform image with bubble defect as an example, the RGB image and the corresponding gradient image are respectively input to the different subnetworks to obtain the feature map of each convolution layer. The extracted features are shown in Figure 5.

It can be seen that the shallow and deep features extracted by the same CNN model from RGB image and gradient image are not completely consistent. There are also differences on extracting features from the same data source between VGG16 and GoogLeNet.

**C. FEATURE FUSION**

Selecting different feature fusion methods will affect the detection and classification accuracy, such as Sum fusion, Mean fusion, Maximum fusion and Wavelet transform fusion. Considering that a single feature cannot fully describe the
image, three different subnetworks are trained to extract different features under different image inputs. And, the output features are fused to be fed into the ECOC-SVM classifier for getting the final classification results [29], [30].

We compared different fusion strategies on the accuracy to achieve the better defect detection accuracy for PET preform.

The Sum fusion strategy is to compute the values sum of the same location of three feature maps. The calculation formula is shown in formula 4.

\[ \text{feature}_{i,j}^{\text{sum}} = \text{feature}_{i,j}^{a} + \text{feature}_{i,j}^{b} + \text{feature}_{i,j}^{c} \] (4)

therein, \( 1 < i < \text{Height} \), \( 1 < j < \text{Width} \), and \( \text{feature}_{i,j}^{a} \) is a feature value of feature map of the original RGB image at point \((i,j)\) by VGG16, and \( \text{feature}_{i,j}^{b} \) is that of the original RGB image by GoogLeNet, \( \text{feature}_{i,j}^{c} \) is that of the corresponding gradient image by VGG16 [31]. The image fused according to the Sum fusion strategy is shown in Figure 6.

The Mean fusion strategy takes the mean value at the same position of three feature maps. The calculation formula is shown in formula 5 [31].

\[ \text{feature}_{i,j}^{\text{mean}} = (\text{feature}_{i,j}^{a} + \text{feature}_{i,j}^{b} + \text{feature}_{i,j}^{c})/3 \] (5)

The image fused according to the Mean fusion strategy is shown in Figure 7.

The Maximum fusion strategy takes the maximum value at the same position of three feature maps. The calculation formula is shown in formula 6 [31].

\[ \text{feature}_{i,j}^{\text{maximum}} = \max \{ \text{feature}_{i,j}^{a}, \text{feature}_{i,j}^{b}, \text{feature}_{i,j}^{c} \} \] (6)

The image fused according to the Maximum fusion strategy is shown in Figure 8.

The multi-scale feature fusion algorithm based on wavelet transform is to decompose the image into low frequency components containing the main contour of the image and high frequency components containing details of image. And then, the minimum detail fusion rule is used to fuse the high
frequency part, and the maximum approximation fusion rule is used to fuse the low frequency part. Finally, the fusion result is obtained by inverse wavelet transform [32], [33].

The wavelet decomposition level should not be too large, and 5-level wavelet decomposition is adopted for the sake of trying to ensure that the high-frequency part contains a lot of edge and texture information, and the low-frequency part considers some structural information because the low-frequency part determines the overall quality of the fusion result to a large extent [34], [35]. The fusion structure of wavelet transform is shown in Figure 9.

![Wavelet transform fusion strategy](image)

**FIGURE 9. Wavelet transform fusion strategy.**

For multi-stream network model, the steps of wavelet transform feature fusion are as follows: Firstly, the features of RGB image by different network are fused according to Wavelet transform fusion strategy to get fusion feature image. And then, the previous fusion feature image and the feature from gradient image are fused to get the final feature fusion image. The image fused according to Wavelet transform fusion strategy is shown in Figure 10.

![Feature fusion based on Wavelet transform fusion strategy](image)

**FIGURE 10. Feature fusion based on Wavelet transform fusion strategy.**

From the analyses of the feature images, it can be seen that the fusion features obtained by different feature fusion strategies are not completely consistent.

The features are extracted by convolution layer named conv5_3 of VGG16 structure for the original RGB image and the corresponding gradient image, and the feature is extracted by inception_4c-output of GoogLeNet structure for the original RGB image at the same time. And the three features are fused according to the fusion strategy [36], [37]. The feature fusion model is shown in Figure 11.

The experimental part of this paper compares the influence of different fusion strategies on the detection and classification accuracy. By comparison, we choose the Wavelet transform fusion strategy which can obtain better accuracy. The features extracted from different network streams play different roles in defect detection, so we should make full use of the advantages of each stream.

V. EXPERIMENTS AND DISCUSSION

The rationality, effectiveness and classification performance of the proposed multi-stream deep CNN model are verified by PET preform surface defect dataset with four different classification categories. The method proposed in this paper is analyzed experimentally. We compare our method with the single-stream network based on VGG16 and GoogLeNet respectively. The influence on the surface defect classification performance for PET preform was studied through experiments using different feature fusion strategies.

In this paper, the supervised training method is adopted. The input images are the original images of PET preform and the corresponding gradient images. They all have corresponding “category labels”. There are four categories of category labels, namely “black spot”, “bubble”, “scratch” and “normal sample”. A variety of evaluation indicators are used including Accuracy, Precision, Recall, Specificity and F1-score for the analyses of experimental results.

A. DATA ACQUISITION

At present, there is no open-data set for PET preform in the world. The project team cooperates with the enterprise which provide PET preform samples, which contain bubbles, black spots, scratches defects samples and normal samples. Therefore, we design an image acquisition platform for PET preform to collect image samples and enhance dataset. The platform consists of industrial camera, plane light source and motion control system. The physical drawing of the platform is shown in Figure 12.

The Basler industrial camera used on the platform is equipped with Semiconductor MT9J003 CMOS photosensitive chip 10 frames per second at 10 MP resolution. The PLC is domestic Mitsubishi PLC-FX-3U, which can communicate with MATLAB through serial port. And each endpoint of PLC is controlled through MATLAB. The light source is an LED plane light source, and the brightness of the light source can be adjusted by the regulator. The acquisition platform is enclosed into a closed box through six transparent plates to provide a non-interference environment for PET preform photographing and reduce external light interference.

We conduct our experiments by a training system with double Intel CPU of 2.4 GHz, 128 GB memory and an NVIDIA Tesla M40 24GB GPU to accomplish the calculations of the proposed multi-stream network model.
In the production line of PET preform, the samples are collected to make the PET preform data set. And the defect sample image is captured by the image acquisition platform. The collected samples consist of four types of the PET preform, including three defect categories and one category without defect. The three main defect categories are black spot, scratch and bubble. There is only one defect on each sample image.

The data set consists of 2298 RGB images with size of 2700 × 2700. There are 1898 images in the training set and the verification set which are randomly shuffled and split into 70% training samples and 30% verifying samples for each category of the dataset, and there are 400 images in the test set. There is no overlap among the training dataset, the verifying dataset and the test dataset. The specific number and distribution of each PET preform surface defect are shown in Table 1.

There are 6094 images which includes 5694 images in the training and verification set. The distribution of the enhanced PET preform dataset is shown in Table 1. The collected samples contain four categories, as shown in Figure 13.

Many evaluation indexes are used for experimental results analysis, including Accuracy, Precision, Recall, Specificity and F1-Score. Accuracy is generally used to evaluate the global accuracy of the model, which cannot comprehensively evaluate the model performance. For a certain defect category, Precision is the proportion of real defects among all identified defects, and Recall is the proportion of the correctly identified defects among all real defects, and Specificity is the proportion of negative cases identified in all negative cases. The calculation formulas are as follows:

\[
\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{7}
\]

\[
\text{Precision} = \frac{T_p}{T_p + F_p} \tag{8}
\]
Recall = \frac{T_p}{T_p + F_n} \quad (9)

Specificity = \frac{T_n}{T_n + F_p} \quad (10)

therein, \(T_p\) is the number of positive samples correctly identified as positive samples; \(F_p\) is the number of negative samples incorrectly identified as positive samples; \(T_n\) is the number of negative samples correctly identified as negative samples; \(F_n\) is the number of positive samples incorrectly identified as negative samples.

\(F1 - Score\) is the harmonic average of precision and recall, and the calculation formula is shown in formula 11.

\[
F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}
\]

\[ (11) \]

B. DEFECT DETECTION ANALYSES OF SINGLE-STREAM MODEL BEFORE AND AFTER DATA ENHANCEMENT

For the PET preform surface defects, the detect accuracy of the single-stream models using softmax classifier or SVM multi-classifier is compared on the original images and the enhanced images of PET preform. We propose a learning rate of 0.0003, validation frequency of 30, batch size of 64 and epochs of 15. Other parameters are set to the default values. The comparison results are shown in Table 2. The single-stream network based on VGG16 is called S-VGG16 for short, and the single-stream network based on GoogLeNet is called S-GoogLeNet for short.

Table 2 shows the quantitative analyses of detection accuracy of various single-stream models. From the analyses of single-stream network model, the accuracy of single-stream network using softmax classifier is higher than that of SVM multi-classifier, and the accuracy of single-stream network after data enhancement is slightly higher than that of original data set. However, the overall test accuracy did not meet the expected requirements.

C. COMPARISON OF DEFECT DETECTION BETWEEN MULTI-STREAM NETWORK AND SINGLE-STREAM NETWORK BEFORE AND AFTER DATA ENHANCEMENT

The single-stream network models in this paper are transfer learning networks based on VGG16 and GoogLeNet, which only slightly adjust the parameters and classifier. However, the multi-stream network model uses three independent networks to extract the features of different images, and fuses the extracted features to put them into the classifier for recognition, which is more complex corresponding to the single-stream network.

We compare the difference on accuracy of the single-stream networks and the multi-stream network using ECOC-SVM model (MSN). The comparison results are shown in Table 2.
FIGURE 15. Comparisons on evaluation metrics for each category with different feature fusion strategies.
TABLE 2. Verify and Test accuracy comparison on the single-stream networks and multi-stream network.

| model   | classifier | Verify Accuracy (%) | Test Accuracy (%) | Verify Average accuracy (%) | Test Average accuracy (%) | Verify Accuracy (%) | Test Accuracy (%) | Verify Average accuracy (%) | Test Average accuracy (%) |
|---------|------------|--------------------|-------------------|-----------------------------|---------------------------|--------------------|-------------------|-----------------------------|---------------------------|
| S-VGG16 | Softmax    | 94.9               | 92.5              | 91.4                        | 89.2                      | 98.4               | 93.3              | 97.1                        | 92.6                      |
|         | SVM        | 94.5               | 87.5              | 93.3                        | 84.9                      | 95.7               | 87.5              | 95.3                        | 87.1                      |
| S-GoogleNet | Softmax   | 95.1               | 91.0              | 95.2                        | 87.1                      | 95.7               | 93.0              | 97.4                        | 88.9                      |
|         | SVM        | 89.7               | 79.8              | 91.7                        | 77.8                      | 93.2               | 80.5              | 92.8                        | 79.8                      |
|         | ECOC-SVM   | 100                | 95.0              | 99.2                        | 88.8                      | 99.1               | 98.5              | 99.6                        | 93.1                      |

TABLE 3. Comparison of different feature fusion strategies for the MULTI-stream network.

| Data set | Max fusion | Sum fusion | Mean fusion | Wavelet transform fusion |
|----------|------------|------------|-------------|--------------------------|
|          | ACC (%) | average -ACC (%) | EG-C-Error | ACC (%) | average -ACC (%) | EG-C-Error | ACC (%) | average -ACC (%) | EG-C-Error |
| before data enhancement | 85.8 | 84.7 | 0.0384 | 82.8 | 69.1 | 0.2745 | 87.3 | 85.9 | 0.0309 | 95.0 | 88.8 | 0.0007 |
| after data enhancement | 86.5 | 86.2 | 0.0198 | 73.8 | 65.5 | 0.4378 | 88.3 | 87.0 | 0.0908 | 98.5 | 93.1 | 0.0002 |

It can be seen from table 2 that the test accuracy of the better training model of the multi-stream network model is 98.5% and the average accuracy is 93.1% on the enhanced data. The performances of the adopted method on enhanced data are better than that of the single-stream network with softmax classifier, which improve accuracy by 5.5%. It can be concluded that the multi-feature fusion strategy has better detection accuracy than the single-feature method for the detection of PET preform surface defects.

The training time of the multi-stream network model is about 8000s, and the single-stream network model with ECOC-SVM is about 2500s, and the single-stream model with softmax classifier is more than 10000s. In terms of time comparison, the single-stream network model with ECOC-SVM is the lowest, but its accuracy is relatively low. Therefore, the multi-stream network model is more suitable in general. The time to load the trained model is about 18 seconds, which also depends on the number of parameters of the saved pre-training model. The total detection time of model for 400 images is about 272s, including image reading, feature extraction and fusion, and defect identification time. Feature extraction and fusion take up most of the time.

D. COMPARISON OF DIFFERENT FEATURE FUSION STRATEGIES

In order to evaluate the effects of different feature fusion strategies on accuracy (ACC), average accuracy (average-ACC) and the estimated generalized classification error (EGC-Error), several groups of experiments are designed to verify the performance of the proposed model. The specific data is shown in Table 3.

As shown in Table 3, the results show that Wavelet transform fusion has the better performance, and the Sum fusion has worse performance on the surface defects detection for PET preform in this paper through experiments. And the EGC-Error with Wavelet transform fusion is the lowest. Therefore, the features fusion model with Wavelet transform fusion is more suitable for defect detection for PET preform to achieve better results from a comprehensive point of view.

With the optimal network architecture determined, we now compare performance of each category in the proposed method before and after data enhancement. The specific data is shown in Table 4.

It can be seen from table 4 that the classification performance of each category after data enhancement is significantly better than that of the original data except for the use of sum fusion strategies. In order to clearly obtain the number of correct identification and wrong identification of each category, the confusion matrices are used to list each category current classification status of the proposed model using four fusion strategies, which are shown in Figure 14.

It can be seen from figure 14, all correct prediction results are on the diagonal, so it is convenient and intuitive to see where there are errors from the confusion matrix because they are presented outside the diagonal. In this confusion matrix of the proposed model using Wavelet transform fusion, there are actually 100 scratch defect samples, but the model predicts two of them as bubble defect samples and three of them as normal samples; For 100 normal samples, one of them were predicted to be scratch defect sample. From the confusion matrix, we can see that the model has some problems in distinguishing scratch and normal sample, but the effect of distinguishing black spots is good. The classification effects of other feature fusion strategies are obviously lower than that of Wavelet fusion strategy.

In order to compare the feature fusion strategies more clearly, the comparisons on evaluation metrics including Precisions, Recalls, Specificity and F1-Score for each category are shown in Figure 15.
As can be seen in Table 4 and Figure 15, the multi-stream network model using Wavelet transform fusion strategy is higher than other fusion strategies on classification precision and recall for various PET preform categories. The above results show that the proposed fusion scheme of Wavelet transform fusion strategy achieves the better fusion effect.

No matter which fusion strategy is selected, the defect classification precision and other related performance of black spots defect is relatively the highest, and that of scratch defect is relatively low. Scratch defect sample is easy to be judged as bubble sample or normal sample, and bubble defect sample is also easily mistaken for scratch defect sample.

E. COMPARISON WITH OTHER PET PREFORM SURFACE DEFECTS DETECTION METHOD

We provide a comparison with other PET preform defect detection method. A. Laucka et al. [8] design the filter to detect the surface black spots and other defects of PET preform workpiece and mainly judge whether the PET preform is defective. For comparison, we sort out the data of the optimal detection model to obtain the qualified detection efficiency of PET preform, and only judge whether the sample is defective or no. The comparison results are shown in Table 5.

As can be seen in Table 5, we can observe that the detection efficiency of the proposed method is slightly higher. The normal samples detection efficiency of the proposed method is 3.4% higher than that of the comparison method, and the defective samples detection efficiency is slightly higher.

Obviously, the multi-stream network model has made good use of the structure and visual information of multiple streams to obtain effective features. And, the multi-stream network uses different network structures to extract features, and the performance of multi features fusion scheme is better than that of single-stream network with single feature detection. In multi-stream network model, the gradient image plays an important role in improving the ability of image representation and recognition.

VI. FUTURE WORK

In this paper, only three kinds of PET preform defects and normal samples are classified and identified. However, PET preform defects have other defects, such as dent defect, gate peeling defect and so on. Therefore, it is necessary to improve the defect sample data set to detect more defects as much as possible. In the future, more PET preform surface defect samples will be provided for training network.

The proposed multi-stream network model does not achieve the best effect, and cannot classify all defect categories. Therefore, the performance of defect detection and classification still needs to be improved by modifying the neural network structure through in-depth analysis.

| feature fusion strategies | Category       | class-wise Precisions(%) | class-wise Recalls(%) | Specificity(%) | F1-Score |
|---------------------------|----------------|--------------------------|----------------------|---------------|---------|
|                           |                | before | after | before | after | before | after | before | after | before | after |
| Wavelet transform fusion  | scratch        | 84.5   | 99.0  | 98.0   | 95.0  | 94.0   | 99.7  | 0.91   | 0.97  |
|                           | bubble         | 99.0   | 98.0  | 100.0  | 100.0 | 99.7   | 99.3  | 1.00   | 0.99  |
|                           | black spot     | 100.0  | 100.0 | 100.0  | 100.0 | 100.0  | 100.0 | 1.00   | 1.00  |
|                           | normal sample  | 98.8   | 97.1  | 82.0   | 99.0  | 99.7   | 99.0  | 0.90   | 0.98  |
| Mean fusion               | scratch        | 78.4   | 80.5  | 69.0   | 70.0  | 93.7   | 94.3  | 0.73   | 0.75  |
|                           | bubble         | 97.9   | 97.8  | 94.0   | 89.0  | 99.3   | 99.3  | 0.96   | 0.93  |
|                           | black spot     | 100.0  | 100.0 | 100.0  | 100.0 | 100.0  | 100.0 | 1.00   | 1.00  |
|                           | normal sample  | 74.1   | 77.0  | 86.0   | 94.0  | 90.0   | 90.7  | 0.80   | 0.85  |
| Maximum fusion            | scratch        | 72.6   | 74.7  | 69.0   | 71.0  | 91.3   | 92.0  | 0.71   | 0.73  |
|                           | bubble         | 96.8   | 96.7  | 90.0   | 89.0  | 99.0   | 99.0  | 0.93   | 0.93  |
|                           | black spot     | 100.0  | 100.0 | 99.0   | 100.0 | 100.0  | 100.0 | 0.99   | 1.00  |
|                           | normal sample  | 75.2   | 76.1  | 85.0   | 86.0  | 90.7   | 91.0  | 0.80   | 0.81  |
| Sum fusion                | scratch        | 67.3   | 49.4  | 68.0   | 39.0  | 89.0   | 86.7  | 0.68   | 0.44  |
|                           | bubble         | 84.3   | 74.0  | 91.0   | 49.7  | 94.3   | 89.0  | 0.88   | 0.59  |
|                           | black spot     | 100.0  | 100.0 | 97.0   | 93.0  | 100.0  | 100.0 | 0.98   | 0.96  |
|                           | normal sample  | 79.8   | 69.4  | 75.0   | 86.0  | 93.7   | 87.3  | 0.77   | 0.77  |

| PET preform used in experiment | The proposed method | A. Laucka et al. [8] |
|---------------------------------|---------------------|----------------------|
| Normal sample                   | 100                 | 500                  | 114                 | 158                 |
| Defective sample                | 99                  | 297                  | 109                 | 156                 |
| Number of test samples          |                     |                      |                     |
| Number of identified samples    | 1                   | 3                    | 5                   | 2                   |
| Efficiency                      | 99%                 | 99%                  | 95.6%               | 98.7%               |
| Total efficiency                | 99%                 | 99%                  | 99%                 | 97.4%               |
From the experimental test results, normal samples are easy to be mistaken for scratches, because some of the collected normal sample images have narrow shadow areas on the side because of the light. Therefore, it is necessary to improve the lighting conditions of the collection system. And they are collected in the experimental environment. However, if applied to the on-site collection on the production line, the quality of the collected images may be worse, so it is necessary to strengthen the quality research of sample collection.

VII. CONCLUSION

Firstly, a new multi-stream deep CNN model is proposed to effectively integrate multiple features for PET preform surface defects detection and classification in this paper. The model extracts feature from RGB image stream and gradient image stream respectively, and also extracts features from RGB image through two different deep CNNs. This is the first time to apply multi-stream network model on PET preform surface defects detection and classification, and it is also an important innovation of this paper. And then, a large number of samples are obtained by data enhancement method in order to prevent the over fitting phenomenon in training because the training of multi-stream network model depends on the support of a large number of training data. And many experiments are carried out to verify that the effect of the detection result after data enhancement is better than that of the original data. Finally, experiments are carried out to compare single-stream network with multi-stream deep CNN model in order to verify the effectiveness of the multi-stream deep CNN model proposed in this paper.

REFERENCES

[1] M. Malesa and P. Rajkiewicz, “Quality control of PET bottles caps with dedicated image calibration and deep neural networks,” Sensors, vol. 21, no. 2, p. 501, Jan. 2021, doi: 10.3390/s21020501.
[2] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee, and S. J. Lee, “Automated defect inspection system for metal surfaces based on deep learning and data augmentation,” J. Manuf. Syst., vol. 55, pp. 317–324, Apr. 2020, doi: 10.1016/j.jmsy.2020.03.009.
[3] C. Duan and T. Zhang, “Two-stream convolutional neural network based on gradient image for aluminum profile surface defects classification and recognition,” IEEE Access, vol. 8, pp. 172152–172165, 2020, doi: 10.1109/ACCESS.2020.3025165.
[4] G. Kalairasi and S. Maheswari, “Frost filtered scale-invariant feature extraction and multilayer perceptron for hyperspectral image classification,” Tech. Rep., 2020.
[5] R. A. Saputra, Suharyanto, S. Wasiyanti, D. F. Saefudin, A. Supriyatna, and A. Wibowo, “Rice leaf disease image classifications using KNN based on GLCM feature extraction,” J. Phys., Conf. Ser., vol. 21, no. 2, p. 500, Jan. 2021, doi: 10.3390/s21020500.
[6] T. Zhang, Y. Yang, P. Wang, Y. Tian, and J. Yang, “An improved Sobel edge algorithm and FPGA implementation,” Proc., vol. 26, no. 6, pp. 891–897, 2020, doi: 10.1117/12.259891.
[7] H.-C. Chen, Y.-H. Ren, C.-M. Hao, Y. Cheng, and X.-W. Zhang, “Defect detection method for medical glass bottles based on YOLOv3,” Packag. Eng., vol. 41, no. 7, pp. 241–246, 2020, doi: 10.19554/j.cnki.1001-3563.202007.034.
[8] A. Laucka, D. Andriukaitis, V. Markevicius, and M. Zilys, “Research of the defects in PET preform,” in Proc. Int. Conf. Methods Models Automat. Robot., 2016, pp. 1273–1279.
[9] S. Hao, W. Wang, Y. Ye, T. Nie, and L. Bruzzone, “Two-stream deep architecture for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 4, pp. 2349–2361, Apr. 2018, doi: 10.1109/TGRS.2017.2777843.
[10] Y. Yu and F. Liu, “A two-stream deep fusion framework for high-resolution aerial scene classification,” Comput. Intell. Neurosci., vol. 2018, pp. 1–13, Jan. 2018, doi: 10.1155/2018/869367.
[11] Q. Yan, D. Gong, and Y. Zhang, “Two-stream convolutional networks for blind image quality assessment,” IEEE Trans. Image Process., vol. 28, no. 5, pp. 2200–2211, May 2019, doi: 10.1109/TIP.2018.2883741.
[12] F. Ren and J. Deng, “Background knowledge based multi-stream neural network for text classification,” Appl. Sci., vol. 8, no. 12, p. 2472, Dec. 2018, doi: 10.3390/app8122472.
[13] C. Ge, Q. Qu, I. Y.-H. Gu, and A. S. Jakola, “Multi-stream multi-scale deep convolutional networks for Alzheimer’s disease detection using MR images,” Neurocomputing, vol. 350, pp. 60–69, Jul. 2019, doi: 10.1016/j.neucom.2019.04.023.
[14] X. Liu, L. Xiao, L. Li, X. Tang, and Y. Guo, “Deep multi-level fusion network for multi-source image pixel-wise classification,” Knowl.-Based Syst., vol. 221, Jun. 2021, Art. no. 106921, doi: 10.1016/j.knosys.2021.106921.
[15] Z. Mushtaq and S. Su, “Environmental sound classification using a regularized deep convolutional neural network with data augmentation,” Appl. Acoust., vol. 167, pp. 107389–107401, Oct. 2020, doi: 10.1016/j.apacoust.2020.107389.
[16] M. D. Bloice, P. M. Roth, and A. Holzinger, “Biomedical image augmentation using augmentor,” Bioinformatics, vol. 35, no. 21, pp. 4522–4524, Nov. 2019, doi: 10.1093/bioinformatics/btz259.
[17] A. Buslaev, V. I. Iglrovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: Fast and flexible image augmentations,” Information, vol. 11, no. 2, pp. 125–144, 2020, doi: 10.3390/infor11020144.
[18] N. Shogoi, D. K. Mohanta, and P. K. Dutta, “Defect detection of steel surfaces with global adaptive percentile thresholding of gradient image,” J. Inst. Eng., B, vol. 98, no. 6, pp. 557–565, Dec. 2017, doi: 10.1007/s40301-017-0296-2.
[19] C. Yi, C. Zhengdong, F. Xiang, C. Yubao, and L. Zhenyu, “Compressive sensing ghost imaging based on image gradient,” Optik, vol. 182, pp. 1021–1029, Apr. 2019, doi: 10.1016/j.ijleo.2019.01.067.
[20] X. Zhang, Y. Zhang, P. Wang, Y. Tian, and J. Yang, “An improved Sobel edge algorithm and FPGA implementation,” Proc. Comput. Sci., vol. 131, pp. 243–248, Jan. 2018, doi: 10.1016/j.procs.2018.04.209.
[21] Z. Wang, K. Wang, F. Yang, S. Pan, and Y. Han, “Image segmentation of overlapping leaves based on Chan–Vese model and Sobel operator,” Inf. Process. Agricult., vol. 5, no. 1, pp. 1–10, Mar. 2018, doi: 10.1016/j.ipa.2017.09.005.
[22] Z. Omiotek and A. Kotyra, “Flame image processing and classification using a pre-trained VGG16 model in combustion diagnosis,” Sensors, vol. 21, no. 2, pp. 500, Jan. 2021, doi: 10.3390/s21020500.
[23] J. Pardebe, B. Sitohang, S. Akbar, and M. L. Khodra, “Implementation of transfer learning using VGG16 on fruit ripeness detection,” Int. J. Intel. Syst. Appl., vol. 13, no. 2, pp. 56–62, Apr. 2021, doi: 10.5815/ijisa.2021.02.04.
[24] J.-H. Kim, S.-Y. Seo, C.-G. Song, and K.-S. Kim, “Assessment of electrocardiogram rhythms by GoogLeNet deep neural network architecture,” Int. J. Healthcare Eng., vol. 2019, pp. 1–10, Apr. 2019, doi: 10.1155/2019/2826901.
[25] G. Hirano, M. Nemoto, Y. Kimura, Y. Kiyohara, H. Koga, N. Yamazaki, G. Christensen, C. Ingvar, K. Nielsen, A. Nakamura, T. Sota, and T. Nagaoka, “Automatic diagnosis of melanoma using hyperspectral data and GoogLeNet,” Skin Res. Technol., vol. 26, no. 6, pp. 891–897, Nov. 2020, doi: 10.1111/str.12891.
[26] S. Deepak and P. M. Ameer, “Retrieval of brain MRI with tumor using contrastive loss based similarity on GoogLeNet encodings,” Comput. Biol. Med., vol. 125, Oct. 2020, Art. no. 103993, doi: 10.1016/j.compbiomed.2020.103993.
[27] N. Teerakawanich, T. Leelaruji, and A. Pichetjamroen, “Short term prediction of sun coverage using optical flow with GoogLeNet,” Energy Rep., vol. 6, pp. 526–531, Feb. 2020, doi: 10.1016/j.egyr.2019.11.114.
[28] N. Alsharman and I. Jawarneh, “GoogleNet CNN neural network towards chest CT-coronavirus medical image classification,” J. Comput. Sci., vol. 16, no. 5, pp. 620–625, May 2020, doi: 10.3844/jcssp.2020.620.625.

[29] X. Wang, L. Gao, P. Wang, X. Sun, and X. Liu, “Two-stream 3-D convNet fusion for action recognition in videos with arbitrary size and length,” IEEE Trans. Multimedia, vol. 20, no. 3, pp. 634–644, Mar. 2018, doi: 10.1109/TMM.2017.2749159.

[30] Y. Cao, Y. Dong, Y. Cao, J. Yang, and M. Y. Yang, “Two-stream convolutional neural network for non-destructive subsurface defect detection via similarity comparison of lock-in thermography signals,” NDT & E Int., vol. 112, pp. 102246–102259, Jan. 2020, doi: 10.1016/j.ndteint.2020.102246.

[31] W. Ye, J. Cheng, F. Yang, and Y. Xu, “Two-stream convolutional network for improving activity recognition using convolutional long short-term memory networks,” IEEE Access, vol. 7, pp. 67772–67780, 2019, doi: 10.1109/ACCESS.2019.2918808.

[32] J. Dou, Q. Qin, and Z. Tu, “Image fusion based on wavelet transform with genetic algorithms and human visual system,” Multimedia Tools Appl., vol. 78, no. 9, pp. 12491–12517, May 2019, doi: 10.1007/s11042-018-6756-0.

[33] H. Liu, H. Ren, Z. Xiao, and T. Zhao, “Research on signals extrapolation in the time domain based on wavelet analysis,” J. Eng., vol. 2019, no. 20, pp. 6533–6536, Oct. 2019, doi: 10.1049/joe.2019.0405.

[34] G. Lu, W. Zhao, E. Forte, G. Tian, Y. Li, and M. Pipan, “Multi-frequency and multi-attribute GPR data fusion based on 2-D wavelet transform,” Measurement, vol. 166, Dec. 2020, Art. no. 108243, doi: 10.1016/j.measurement.2020.108243.

[35] E. Daniel, “Optimum wavelet-based homomorphic medical image fusion using hybrid genetic–grey wolf optimization algorithm,” IEEE Sensors J., vol. 18, no. 16, pp. 6804–6811, Aug. 2018, doi: 10.1109/JSEN.2018.2822712.

[36] K. Li, C. Zou, S. Bu, Y. Liang, J. Zhang, and M. Gong, “Multimodal feature fusion for geographic image annotation,” Pattern Recognit., vol. 73, pp. 1–14, Jan. 2018, doi: 10.1016/j.patcog.2017.06.036.

[37] C. Khraief, F. Benzarti, and H. Amiri, “Elderly fall detection based on multi-stream deep convolutional networks,” Multimedia Tools Appl., vol. 79, nos. 27–28, pp. 19537–19560, Jul. 2020, doi: 10.1007/s11042-020-08812-x.

TAOCHUAN ZHANG received the Ph.D. degree from Central South University and is currently an associate professor. His researches include signal processing, deep learning and fault diagnosis.

CHUNMEI DUAN received the Ph.D. degree from Central South University. She is a Senior Engineer. Her researches include computer vision, deep learning and defect classification and identification.