Two-Stream Convolutional Neural Network Based on Gradient Image for Aluminum Profile Surface Defects Classification and Recognition

CHUNMEI DUAN[1,2] AND TAOCHUAN ZHANG[3]
[1]Institute of Light Alloy, Central South University, Changsha 410083, China
[2]School of Electronic Information, Foshan Polytechnic, Foshan 528137, China
[3]College of Mechanical and Electrical Engineering, Foshan Polytechnic, Foshan 528137, China

Corresponding author: Chunmei Duan (49977405@qq.com)

ABSTRACT In this article, a novel two-stream convolutional neural network based on gradient image is performed to effectively classify and identify aluminum profiles defects for the first time. Recent feature fusion methods based on two-stream network prove promising performance for defects classification and recognition. In this article, we use data enhancement methods to obtain a large number of samples to prevent the over fitting phenomenon in deep learning. The image gradient is calculated with the Sobel operator, and normalized to transform the data between zero and one under the same dimension. We design a two-stream convolutional neural network model adopting Wavelet transform fusion strategy to realize feature fusion on the ReLU6 layer, which uses the original RGB image of aluminum profile and the gradient image corresponding to the original RGB image as inputs to extract features through two sub-networks and fuses features on a concatenate layer to be input into SVM classifier for classification and recognition. Using Bayesian Optimization function and computing the cross-validation classification error to optimize the hyperparameters to choose the best performance configuration is performed. A series of experimental data, which include accuracy and estimated generalized classification errors of single-stream and two-stream networks with different feature fusion strategies on different fusion layers, are conducted and show that the current model has good convergence, accuracy, stability and generalization. On this basis, this article also proposes a series of innovative methods for the future research of other defects.

INDEX TERMS Aluminum profile surface defects, two-stream network, gradient image, convolutional neural network, SVM.

I. INTRODUCTION

Due to the growing demand for massive infrastructure investment in the world and rapid development of industrialization, the aluminum profiles with high strength, light weight, corrosion resistance, long service life, rich color and other advantages are mainly used in automobile manufacturing, rail transit, equipment and machinery manufacturing, consumer durables, aerospace industry and other industries [1]. Therefore, the surface quality of aluminum profile is very important significance and can greatly affect the product performance. However, on account of extrusion equipment and extrusion technology in the extrusion process, compression deformation, spray grouts and other operating reasons, the aluminum profile surface will produce abrasion mark, traffic mark, dirty points, convex powder, orange peel and other defects, which affects the quality of aluminum profile [2]. In many manufacturing industries, human inspection is still an important

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

component in the production process, but cannot guarantee the detection stability and accuracy. Therefore, automated defects detection is naturally needed to solve the problem, and then machine vision technology gradually replaces traditional human inspection methods and can satisfy customer requirements because of its highly recognition rate and detection stability [3]. Generally, the aluminum profiles surface defects detection process based machine vision method consists of the following four steps: image capture, image preprocessing, feature extraction and defects recognition. Among them, the feature extraction can directly affect the detection accuracy and even lead to detection failures [4]. The traditional machine vision feature extraction generally includes Histogram of Oriented Gradient(HOG), Scale-Invariant Feature Transform(SIFT), Oriented FAST, Speeded-Up Robust Features (SURF), Oriented FAST and Rotated BRIEF(ORB), Local Binary Pattern(LBP) and Haar-like features(HAAR). And SIFT and HOG are all feature extraction methods based on the histogram of gradient direction in the image. SURF operator is an improvement of SIFT. ORB adopts FAST (Features from Accelerated Segment Test) algorithm to detect feature points, which can be used for real-time feature detection. LBP is a kind of feature operator to describe local texture of image, which has rotation invariance and gray invariance. Haar-like is combined with AdaBoost algorithm to be as the most commonly used face detection algorithm. Haar-like features only use feature template to calculate the depth mode of image color, and the template slides on the image with different sizes and positions to calculate the feature value [5], [6]. Relative to the above feature extraction methods, the features extracted by Deep Learning has better applicability, generalization and expression ability. And convolutional neural network has outstanding advantages on feature extraction and classification performance, which is widely used for image processing, classifying and understanding [7].

Many scholars have studied the effective feature extraction methods in defect detection system. The following is the research on defect detection of aluminum profile by some scholars. Wei and Bi [8] proposed a recognition technology of aluminum profile surface defects with a multiscale defect-detection network based on deep learning. Liu *et al.* [9] proposed aluminum profile texture feature extraction method by means of GLCM algorithm and Gabor wavelet transform methods, and used SVM based on RBF to classify the feature. Neuhauser *et al.* [10] proposed a method of surface defect classification and detection based on a neural network structure adopting data enhancement and migration learning which is the key for network training. Li and Liu [11] studied aluminum plate surface defects recognition method through BP neural network which used three layers neural network structure.

Although there are a few articles on the detection of aluminum profile surface defects, there are many researches on the detection of other materials, such as steel surface defects. Fu *et al.* [12] proposed a deep-learning-based method via a compact yet effective convolutional neural network

model, which emphasized the training of low-level features and incorporated multiple receptive fields. Wang *et al.* [13] proposed a convolutional neural network (CNN) to automatically extract features for distinguish between the defect free and defective image to achieve product quality control. Weimer *et al.* [14] proposed Convolutional Neural Networks (CNN) architecture for industrial inspection by designing different configurations to study the influences of different hyper-parameter on the detection results.

According to the above analysis, although the previous methods have achieved good results, there are still some problems to be solved. Firstly, many models adopt machine learning algorithms for RGB image to detect defects but do not effectively use useful information from other sources or representations. Secondly, the single information source cannot completely and effectively reflect the characteristics of objects. Thirdly, due to surface defects that can occur in any size, shape and orientation, the standard defect feature description obtained by traditional common feature extraction methods in machine vision might lead to low accuracy for classification and recognition.

Considering the above some problems, Convolutional Neural Network can be adopted to overcome the difficulties of different feature representations, and the two-stream convolutional neural network can extract the structural information characteristics and the intensity and color information characteristics of the image respectively. And, in order to appropriately and effectively improve the defect classification accuracy, these methods for combining the local feature details and the global features of the whole picture are adopted on different materials defect detection or other aspects. Hao *et al.* [15] proposed a hyperspectral image classification method based on a two-stream network including two subnetworks with the encoded spectral values of each input pixel by the stacked denoising auto-encoder as one stream and the corresponding image patch processed by deep convolutional neural network as input to the other stream. Yu and Liu [16] designed a two-stream framework which takes convolution neural network as the feature extraction network by the original RGB stream and its corresponding saliency stream for aerial scene classification. Yan *et al.* [17] predicted the image quality via a two-stream convolution network including two subnetworks with images and gradient images as input respectively, which simultaneously paid attention to the extraction of detailed structural features and the information strength.

Although there are a lot of researches on multi-stream for classification and identification of various materials defects, there are few two-stream neural network methods that integrate global and local features on the aspect of defect classification and detection for aluminum profiles. In order to get better accuracy and recognize more types of defects, the architecture of traditional convolutional neural network will be improved. In addition, there are not many samples in the open aluminum profile database, so it is difficult to meet the requirements of training deep neural network which

needs a large number of training samples. And, the storage capacity and computing performance required by the well-known convolutional neural network model are usually very important for real-time defect detection. Therefore, a two-stream convolutional neural network with RGB image and gradient image as input respectively by fusing information from multiple sources is adopted for aluminum profile surface defects classification and recognition, which comprehensively considers combining local structure details with global image features to effectively and easily identify defects for aluminum profile image.

II. DATA PREPROCESSING
A. DATA ENHANCEMENT
In order to prevent the insufficient training samples of aluminum profiles from leading to low detection accuracy, low fitting and low robustness of detection model, the amount of input image samples are increased by properly transforming the input image, which can effectively solve the lack of training samples problem and satisfy requirements for accuracy and stability [18]. Firstly, the size of aluminum profile image is adjusted to 227 × 227. And then, the following geometric transformation methods are selected to amplify the data in this article, such as flip, brightness adjustment, rotation. The flip operation includes vertical flip, horizontal flip, and vertical flip after horizontal flip. Brightness adjustment is to adjust the gray level of the image through non-linear mapping, and specifies the gray range before and after the transformation. The rotation operation, adopting Bilinear Interpolation algorithm, is to rotate a certain angle around the center point of the image, and the image is subsequently cropped to keep the size of the output image consistent with the input image after rotation, because the previous rotation makes output image larger enough to contain the entire rotated image to ensure that the pixel value beyond the image size range is not lost after the rotation of the source image [19], [20]. Figure 1 shows the diagram of data enhancement for aluminum profile image.
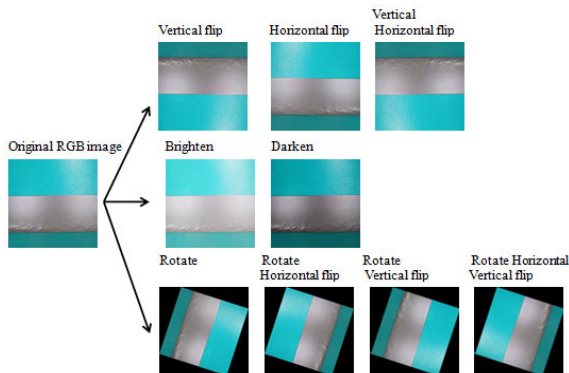

FIGURE 1. The data enhancement diagram for aluminum profile image.

B. GRADIENT IMAGE OF THE ORIGINAL GRB IMAGE
Under normal circumstances, the gradient image can robustly reflect the image structures in details under the variations of the image intensities and colors, and the defect area gradient information is more sensitive to defect classification and recognition [21]. We believe that using the two-stream feature fusion based on the gradient image stream and the original RGB image stream will promote the detection performance.

As an important content of data preprocessing, the image gradient algorithm can enhance the outstanding change between defect area and homogeneous background due to that the gradient direction is reflected in the maximum change rate of image gray level, and image gradient can strengthen the edge information in the image. And the defect areas of aluminum profile usually have the characteristics of large change rate of gray level value, so most of the redundant background information can be eliminated by gradient algorithm, and the edge information with higher significance can be retained. When calculating the gradient image of aluminum profile image, different operators are selected for comparison to ensure that the calculated gradient image is the most representative. Considering the characteristics of different defects, it is better to consider Sobel operator comprehensively based on the experimental comparison of gradient algorithm with different edge detection operator and [22]. Image gradient is generally obtained by derivative of image function, which can express the change rate of image gray. In most cases, we use difference to approximately express derivative for image gradient. The difference form of Sobel operator is shown in Formula 1 [23].

$$\begin{cases} \Delta_x f(x, y) \\ = [f(x+1, y+1) + 2f(x, y+1) + f(x-1, y+1)] \\ -[f(x+1, y-1) + 2f(x, y-1) + f(x-1, y-1)] \\ \Delta_y f(x, y) \\ = [f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)] \\ -[f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)] \end{cases} \quad (1)$$

Image gradient is calculated by Sobel operator with a kernel of size 3 × 3 which slides on the image in two directions, and the template convolves with the image using a convolution kernel of size 3 * 3 with 9 pixels. And then according to the gray weighted difference of the upper, lower, left and right neighboring pixels, the edge is detected by the phenomenon of reaching the extreme value at the edge. There are two Sobel operators, one is to detect the horizontal edge and the other is to detect the vertical edge. The calculation formulas are shown in Formula 2 and 3 [24].

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * f(x, y)$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * f(x, y) \quad (2)$$

$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$

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 Sobel
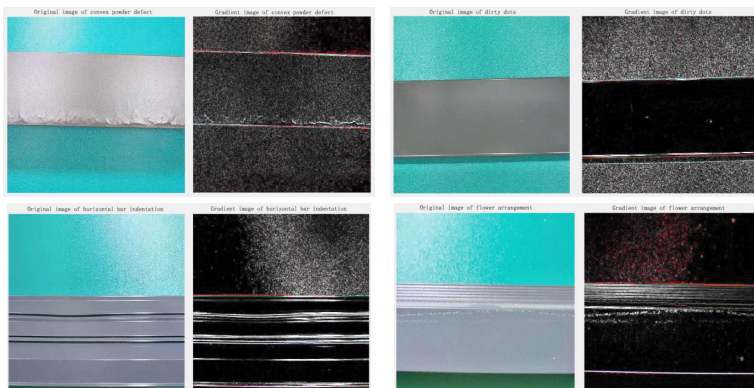
FIGURE 2. The aluminum profiles GRB images with defect and the corresponding gradient images.

operators on horizontal and vertical directions respectively, and G is the final edge amplitude, which is sometimes simplified as shown in formula 4 to reduce the amount of calculation [25].

$$G = |G_x| + |G_y| \tag{4}$$

According to the above calculation method, the original aluminum profile RGB image with different types of defects are calculated to get the gradient image combining X direction and Y direction. After that, the gradient image is normalized to transform the data between zero and one under the same dimension. Figure 2 shows the original RGB images and the corresponding gradient images [26].

III. PROPOSED METHOD

A. TWO-STREAM CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

The traditional convolutional neural network, generally operated on RGB image or Gray image, can extract various effective features map of image by nonlinear transformation from a single image. And the RGB image by itself is important information, including color feature, texture feature, shape feature and spatial relation feature. Compared with the original RGB image, the Gray image does not contain color information, so that the information content is greatly reduced and the image processing calculation is correspondingly reduced. These traditional methods, using RGB or Gray image to train based on a single-stream convolution neural network, do not consider dealing with local structure characteristics. If we want to consider the extraction of local structure detail information, we tend to select gradient image as another network input for feature extraction at the same time, because the image gradient can significantly reflect the high-frequency information of an image and includes the most significant detail features for identifying the defects of aluminum profile image effectively and easily. And the gradient image has better robustness under different illumination conditions. The gradient feature of the image can find the potential relationship between adjacent pixels in the aluminum profile image, which is insensitive to illumination. However, the

experiments as described in Section 4 show that using the gradient image alone is similar to use the original RGB image alone for defect classification accuracy on a single-stream network or worse. In addition, there are also over fitting phenomena in the above two cases [17].

Although the deep learning with a single-stream network based the above works have a certain good performance, there is still a certain distance to achieve satisfactory results. There is also a problem whether it is advantageous to use multi-feature fusion or whether one image suffice. The image gradients keenly reflect the structural components of images, such as image edges. The gradient image can robustly reflect the image structures in details under the variations of the image intensities and colors. The gradient amplitude can reflect the change of texture details, and the histogram of gradient direction can describe the representation and shape of local objects. Considering the above reasons, the two-stream technology is usually used to improve performance by fusing features extracted by two independent convolutional neural networks. And then, some features can be extracted from gradient image which can well reflecting the characteristics of image edge and other details, and reduce the difficulty of extracting some effective features from the RGB images stream. The gradient image stream focuses on the extraction of detail structure features, while the image stream pays more attention to the extraction of intensity information and content features. As far as we know, the multi feature fusion method based on two independent convolutional neural network streams (RGB image stream and gradient image stream) has not yet been applied for aluminum profile defects classification and detection.

Due to the variety of aluminum profile defects, subtle defects, complex shape and the more difficult discovery with the lack of careful observation, the defects are still difficult to distinguish even when the defects are manually marked. In order to improve the accuracy of detection, we design a two-stream convolutional neural network model(TSCNNM) using the original RGB image of aluminum profile and the gradient image corresponding to the original RGB image as input to extract features through two network and fuse
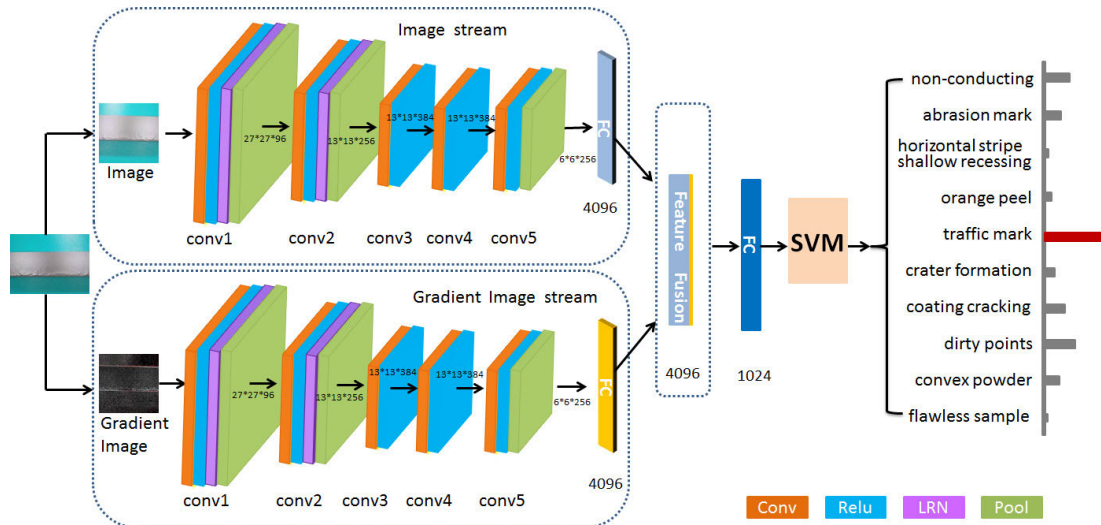
FIGURE 3. The two-stream network architecture.

features on a concatenate layer to be input into classification layer for classification. In the two-stream network used in this article, the gradient image of the original RGB image is trained as the network input together to retain more edge information. The two-stream network architecture as shown in Figure 3 consists of two independent streams taking an original RGB image and a gradient image of the corresponding original GRB image as input respectively. In this article, the domain knowledge of aluminum profile defects is fully considered. The network structure proposed for defect classification consists of two convolutional neural networks, which process the complete RGB defect image and the corresponding gradient image respectively. Then, the output features of the two streams are spliced and fused on a layer. Finally, the classification probability is obtained through the classifier by the output fused features [27], [28].

We should select the appropriate convolutional network for defects classification and identification to ensure an acceptable processing rate and accuracy of the system. The each stream of the two-stream network architecture in this article is based on ALEXNET [29]. In order to meet the network input requirements, all the images are standardized to size 227×227. And then the network inputs the complete original RGB defect image and the corresponding gradient image to the feature extraction network respectively. Each feature extraction network consists of five convolution layers, one full connection layer for extracting more feature maps effectively. The design of convolution layer is identical with that of ALEXNET. The Rectified Linear Unit immediately follows each convolution layer. The response-normalization layers only follow the conv1 and conv2 layers, and the three pooling layers follow the norm1 layer, norm2 layer and the fifth convolution layer respectively. A full connection layer connects the dropout layer to prevent over fitting effectively. And then, the feature vectors of full connection layers with

the ReLU in the two-stream network are spliced on the concatenate layer which connects the output of the two feature extraction network, and the fused feature is input to one full connection layers of 1024 nodes coming after the concatenate layer, and the output is fed into a SVM classifier used to map features to a probability distribution of defect classes for the final decision. The final defect type prediction gives 10 classes [30], [31].

B. FEATURE EXTRACTION

The purpose of convolution operation is to extract different features of input. The early convolution layers can only extract some low-level features such as edge, line and angle. The late convolution layers of network can extract more complex features from low-level features iteratively. In the first convolution layer, ninety-six convolution kernels of size $11 \times 11 \times 3$ are used to slide on images of size $224 \times 224 \times 3$ with 4-pixel intervals to realize convolution operation. And then the convolution feature of this layer is formed, and we make an operation of ReLU and normalization to get the corrective feature map, and the max-pooling operation with a kernel of size 3×3 and a stride of 2 is performed to complete the convolution pooling operation of this layer and provide the output to the next layer. The operations of another convolution layers are basically similar to that of the first convolution layer, except that the convolution kernel is different. The convolution kernel size is $5 \times 5 \times 48$ in the second layer, $3 \times 3 \times 256$ in the third layer, and $3 \times 3 \times 192$ in other layers in turn. In addition, the first, second and fifth convolutions are followed by a pool operation with an overlapping size of 3×3 and a stride of 2 [29].

Taking the aluminum profile image with convex powder defect as an example, the RGB image and the corresponding gradient image are respectively input to the convolution neural network to obtain the feature map of each
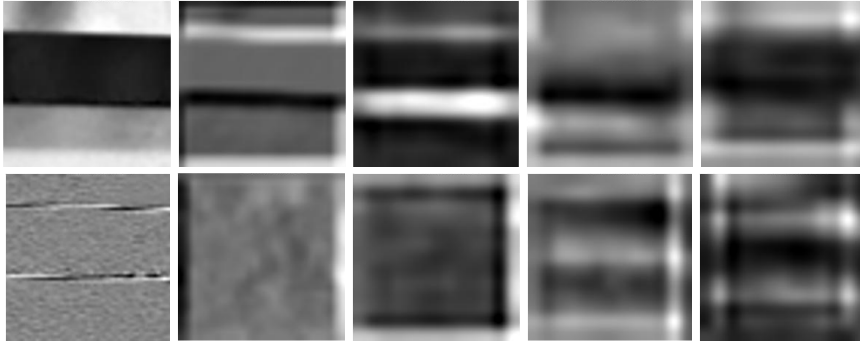
FIGURE 4. Feature maps of the original RGB image with convex powder defect (top) and the corresponding gradient image (bottom).

convolution layer. Figure 4 shows the feature maps of the original GRB image with convex powder defect (top) and that of the corresponding gradient image (bottom) extracted from the five convolution layers. It can be seen that the image features respectively extracted from two sub-network streams are not in complete accord, some of which separate background features and extract contour features, color difference features, and even solid color features.

After using the above method to extract the each layer features of the original RGB image and gradient image respectively, we normalize these two features using standard deviation, and then merge the features by weighting in the feature fusion stage. The main purpose is to transform the data of different levels into the same level, and use the calculated value to measure, so as to ensure the comparability between the data.

C. FEATURE FUSION

There are many feature fusion methods, such as Sum fusion, Maximum fusion, Concatenation fusion, Wavelet transform fusion and so on. Selecting different feature fusion methods or different feature fusion layers will affect the classification and recognition accuracy. Considering that a single image feature cannot fully describe the image, two identical sub-network streams are trained at the same time to extract the different and effective image features with different image inputs in this article. The output features of these two network streams at a certain layer, given different weights, are fused to be fed into the corresponding SVM classifier for training and get the final classification results. In order to improve the accuracy of defect detection for aluminum profile, we compared the effect of different layer fusion selection under different fusion strategies on the accuracy. In the fourth section of the paper, we listed the comparison results. Through the comparison, we selected the appropriate fusion strategy and fusion layer, which can help us achieve better accuracy [32], [33].

We have evaluated different feature fusion strategies to compare the accuracy for aluminum profile defect classification. The calculation formula of the Sum fusion strategy

which is to compute the sum of the values of the same location of two feature maps is shown in formula 5.

$$f_{i,j}^{sum} = f_{i,j}^a + f_{i,j}^b \tag{5}$$

therein, $1<i<H$, $1<j<W$, and $f_{i,j}^a$ is a feature value of feature map of the original RGB image at point (i, j), and $f_{i,j}^b$ is that of the corresponding gradient image [34].

The calculation formula of the Concatenation fusion strategy is shown in formula 6. The spliced feature matrix is a longitudinal splicing matrix, where the matrix $f_{i,j}^b$ is added to the last row of the matrix $f_{i,j}^a$ [34].

$$f_{i,j}^{concatenation} = \left[f_{i,j}^a; f_{i,j}^b \right] \tag{6}$$

The calculation formula of the Maximum fusion strategy which takes the maximum value at the same position of the two feature maps is shown in formula 7 [34].

$$f_{i,j}^{maximum} = \max \left\{ f_{i,j}^a, f_{i,j}^b \right\} \tag{7}$$

The Wavelet transform fusion strategy is that the 5-level wavelet decomposition of two feature images are executed, and the maximum value of approximate signals and the minimum value of detail signals for fusion with 'db2' wavelet are adopted for fusion [35].

The double scale classification formula of two-dimensional image wavelet decomposition is shown in formula 8 [36].

$$\begin{cases} \varphi_{i-1,l}(x) = \sum_{k \in Z} h_0(k - 2l)\varphi_{ik}(x) \\ \psi_{i-1,l}(x) = \sum_{k \in Z} h_1(k - 2l)\varphi_{ik}(x) \end{cases} \tag{8}$$

therein, φ is the scale function, ψ is the wavelet function. The feature maps of the aluminum profile image and the corresponding gradient image are respectively decomposed into sub images which has high and low frequency sub information by two-dimensional Mallat algorithm of wavelet transform.

The low-frequency information contains the main outline of the image and reflect the approximate and average

characteristics of the image. The high-frequency information contains the details of the image output such as bright lines,boundaries and area contours.

The N-level decomposition of aluminum profile feature map can be realized by Mallat algorithm. Each decomposition layer has four sub images, which represent the low-frequency component, horizontal high-frequency component, vertical high-frequency component and diagonal high-frequency component of the decomposed image. Choosing different decomposition levels has different effects. The selection of decomposition level should be moderate, otherwise the low-frequency contour can't be decomposed well or the low-frequency contour will be lost too much. In this article, 5-level decomposition is used, then the fusion is carried out by adopting the appropriate fusion rules under the optimal number of decomposition layers. The selection of fusion rules is an important part in the process of image fusion, which directly determines the quality of the fused image. Many experiments have been done for a variety of fusion rules, and the fusion rules with the highest detection rate are selected. In this article, the high frequency part is fused by the minimum detail fusion rule. The low frequencies are fused by the maximum approximation fusion rule, and the weighted average method is generally used to fuse for low-frequency coefficients [37], [38].

The effect of Wavelet transform fusion is that if the target in the gradient image before fusion is more significant than the same target in the original RGB image, the target in the gradient image after fusion will be retained and the target in the original RGB image will be ignored. In this way, the wavelet transform coefficients of the target in these two images will dominate at different resolution levels, so that the salient objects in the gradient image and the original image are preserved in the final fusion image. The flow chart of decomposition and fusion of Wavelet transform is shown in the figure 5.
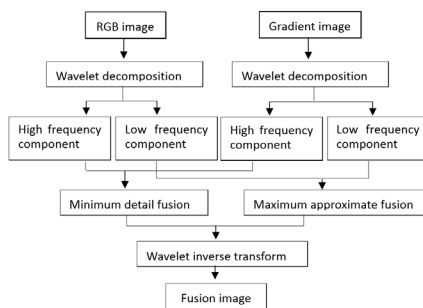

FIGURE 5. Wavelet transform fusion.

At the same time, we've evaluated the fusion on different layers. There are great differences on accuracy for feature fusion on different layers. The features from the conv5 layer or the previous layers also described as feature level fusion are fused before being input into the classification model as shown in figure 6 [39], [40].

The feature fusion for aluminum profile defect detection before the fully connection layer cannot get higher accuracy, even the results are worse than expected results. But the output of feature fusion on the first fully connection layer can get better results to satisfy the defect detection for the aluminum profile data set in this article and is more flexible. This fusion method can be said to be a combination of output classifications from two streams, which structure is as shown in Figure 7 [39], [40].

The two features from the fully connection layer 6 are integrated by Wavelet transform fusion and fed to SVM to be trained to generate the final detection result for the proposed method. The different streams do not play the same part on detecting defect events, so they should have different weights in order to make perfect use of the advantages of each stream. Therefore. The weighted average method is often used in the fusion of wavelet transform.

D. OPTIMIZE CLASSIFIER

In the deep neural network, it is not easy to adjust the hyperparameter combination, because it is very time-consuming to train the deep neural network and impossible to optimize hyperparameters by gradient descent method like general parameters. The time cost of evaluating a set of hyperparameters configurations is very high. The choice of hyperparameters has a great influence on the final effect of the model, and different models will have different optimal hyperparameters combinations. The influence of different hyperparameters on the model performance is very different. Some hyperparameters, such as regularization coefficient, have limited influence on the model performance, while other hyperparameters, such as learning rate, have great influence on the model performance. Optimizing the hyperparameters by appropriate optimization methods to choose the best performance configuration is very critical [41].

In this article, the classifier uses Support Vector Machine (SVM) model. The optimizer uses Bayesian Optimization which is an adaptive hyperparameter search method and can predicts the next combination with the maximum benefit according to the currently tested hyperparameter combination to select the model and its hyperparameter values [42]. And then we compute the cross-validation classification error for each model and use standard categorical cross entropy loss to optimize our two-stream network. After the optimization, the whole training data set is trained to obtain the optimized model. The optimized model can classify the test data optimally so as to check model performance [43].

IV. EXPERIMENTS AND DISCUSSION

The reasonableness, validity, robustness and classification performance of the two-stream convolution neural network model proposed in this article are verified on the aluminum profile defect data set representing 10 different classification categories with visual defects occurring on the surface. Firstly, ALEXNET is used as the basic network to realize the two-stream convolution neural network model
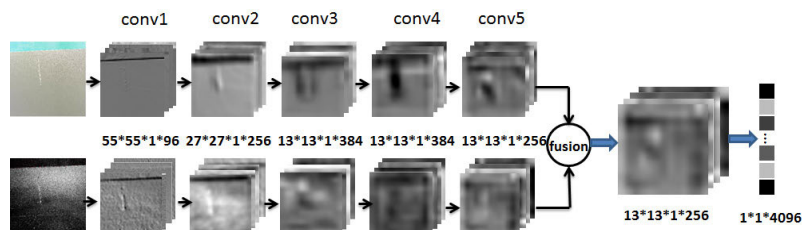
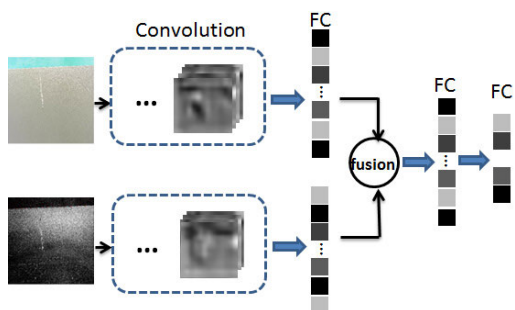**FIGURE 6.** Convolutional layer early feature fusion strategy.

FIGURE 7. Fully connection layer late fusion strategy.

TABLE 1. Classification statistics of aluminum profile data set.

Category	Number of raw data			Number of enhanced data		
	train	verify	test	train	verify	test
non-conducting	24	10	5	286	122	35
abrasion mark	86	37	10	1033	443	70
horizontal stripe shallow recessing	29	13	5	353	151	35
orange peel	56	24	10	672	288	70
traffic mark	47	20	10	563	241	70
crater formation	31	13	10	370	158	70
Coating Cracking	21	9	5	252	108	35
dirty points	139	60	10	1672	716	70
convex powder	27	11	5	319	137	35
Flawless sample	699	299	20	699	299	55

in the experiment. In this case, the two-stream network should be pre-trained. Additionally, during the training stage, the genuine training RGB image samples and the corresponding gradient image samples are input to each stream of the two-steam network respectively, and a concatenate layer is used to fuse GRB image and gradient image features. And then we perform experimental analysis for the proposed method in this article. We compare our method with single stream network structure which input is original RGB image or gradient image respectively. We also study the feature fusion influence on the performance of defect classification for aluminum profile with different feature fusion strategies on different layers of ALEXNET network. Finally, we compare the optimized network with the un-optimized network. In order to prove the robustness of our network, we take random sorting samples for five times to get the value in the process of network training and testing procedure.

A. EXPERIMENTAL DEVELOPMENT ENVIRONMENT AND DATA

We conducted our experiments by a training system with twelve cores of 2.5 GHz double CPU, 128 GB memory and an NVIDIA Tesla M40 24GB GPU to accomplish the calculations of our proposed two-stream network system using Matlab. In our experiments, the data used in this article are mainly from [44]. We selected nine sample types with single defect and one defect-free sample type. The selected samples include a total of 1745 aluminum profile images and be manually divided into 10 different categories from dataset. The ten categories are non-conducting (n-c), abrasion mark(AM), horizontal stripe shallow recessing (HSSR),

orange peel (OP), traffic mark (TM), crater formation (CF), coating cracking (CC), dirty points(DP), convex powder(CP), flawless sample(FS). All data with the original resolution of 2560×1920, which are cropped to 227×227 pixels to contain aluminum profile pixels and subtract unnecessary background pixels, are randomly divided into training, validation, and testing samples, and the specific number of each image type is shown in Table 1. The samples in the data set are amplified by the way of data enhancement in this article, and the data after amplification is shown in Table 1. The number of the corresponding gradient images is consistent with the number of original images.

The whole data used in the experiment is divided into two parts. The first complete set of images now contains 1655 original samples and 8882 samples after data enhancement which are randomly shuffled and split into 70% training samples and 30% verifying samples for each category of the dataset for each evaluation. And then the rest part of the data which includes 90 original samples and 545 samples after data enhancement was tested with the trained classifier.

In the training of the network, we proposed a learning rate of 0.001, batch size of 10, validation frequency of 30, momentum of 0.9 and epochs of 8 (epoch means the number of the network is trained). Other parameters are established as the default values.

B. COMPARISON OF THE PROPOSED TWO-STREAM NETWORK AND THE SINGLE-STREAM NETWORK

In this section, we compare the two-stream network with the single-stream network for proving the reasonableness of the

TABLE 2. Test accuracy and estimated generalized classification error comparison on the two-stream and single-stream networks.

| Model	Accuracy							Estimated Generalized Classification Error						
	conv5	relu5	pool5	fc6	relu6	drop6	fc7	conv5	relu5	pool5	fc6	relu6	drop6	fc7
two-stream network (proposed)	77.78	84.44	88.44	92.22	94.44	92.22	86.67	0.014	0.009	0.009	0.009	0.010	0.008	0.015
single-stream network (RGB image)	64.44	82.22	84.44	87.78	87.78	88.89	84.44	0.130	0.089	0.072	0.063	0.077	0.073	0.073
single-stream network (gradient image)	72.22	78.89	83.33	84.44	86.67	86.67	82.22	0.132	0.079	0.086	0.085	0.089	0.088	0.100
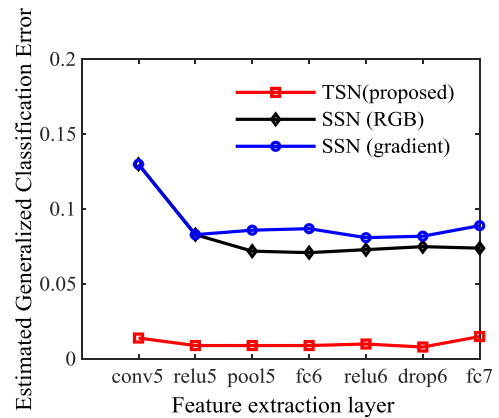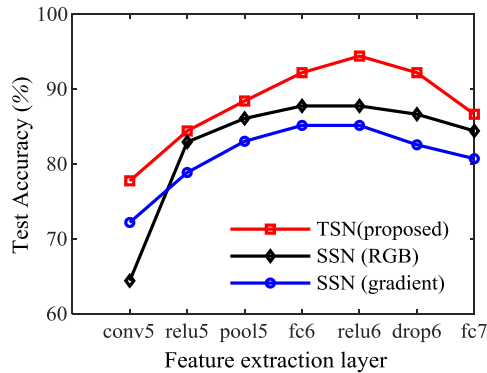
FIGURE 8. Comparison of the test accuracy and EGC-Error for single-stream and two-stream network on different layer.

network designed in this article by the experiment analysis. And then we design experiments to fuse the features of different layers of the network to be input to SVM for classification. As for the aluminum profile defect classification task, a variety of evaluation metrics are used, including accuracy, Receiver Operating Characteristic, Precision, F1-score and so on [45].

Firstly, we compare the verification and test results of aluminum profile defect classification and detection between two-stream and single-stream networks. We use accuracy and estimated generalized classification error (EGC-Error) to measure the difference between the two networks. The specific data is shown in Table 2. We conducted all defect classification and detection experiments on the same data set in this article, and then the data set followed the same training and validation segmentation settings and unified test data.

We compare the difference on accuracy of the single-stream network and the two-stream network which features from different layers are fed to SVM classifier for classification. Table 2 lists test accuracy and the estimated generalized classification error (EGC-Error) of the single-stream network and two-stream network on the aluminum profile image data set. It can be seen from table 2 that using two-stream network can achieve better accuracy, no matter which layer is used for feature fusion with wavelet transform. The comparison trend chart of the test accuracy and the estimated generalized classification error on three different networks on different layers is shown in Figure 8.

Figure 8 shows that the test accuracy and the estimated generalized classification error trend curves by three networks, one of which is two-stream network and the others are single-stream networks for original RGB images and gradient images respectively. The red curve denotes the two-stream network on test set and the black and blue curve denote single-stream networks on testing set. So the conclusion from Table 2 and Figure 8 is that the proposed multi feature fusion strategy achieves better detection accuracy than that using single feature method in this case for aluminum profile defect detection on data set in this article.

C. COMPARISON OF DIFFERENT FEATURE FUSION STRATEGIES ON DIFFERENT LAYER

In order to evaluate the effects of different feature fusion strategies on different layers, several groups of experiments were designed. The specific data is shown in Table 3. We compare the difference on the accuracy with four feature fusion strategies, including Sum fusion, Maximum fusion, Concatenation fusion and Wavelet transform fusion. And in order to find the most suitable layer location for defect classification and detection in this data set, we study the influence on the accuracy with selecting different fusion layers. The location of the feature fusion layers includes conv5, releu5, pool5, fc6, releu6, drop6, and fc7 layers.

The accuracy comparison figure with different feature fusion strategies on different layers for test is shown in Figure 9(left), and the estimated generalized classification

TABLE 3. Comparison of different feature fusion strategies on different layer for the two-stream network.

| Fusion layers	Max fusion			Sum fusion			Concatenation fusion			Wavelet transform fusion		
	V-A	T-A	EGC-Error	V-A	T-A	EGC-Error	V-A	T-A	EGC-Error	V-A	T-A	EGC-Error
fc7	93.75	87.78	0.066	93.75	87.78	0.069	90.73	86.67	0.073	96.17	86.67	0.015
drop6	94.56	91.11	0.067	94.96	88.89	0.062	92.44	87.78	0.073	95.16	92.22	0.008
relu6	93.95	92.22	0.057	94.35	91.11	0.060	93.35	86.67	0.069	96.37	94.44	0.010
fc6	93.95	93.33	0.071	94.35	91.11	0.065	93.04	87.28	0.082	98.39	92.22	0.009
pool5	93.15	90.00	0.062	94.35	85.56	0.066	92.84	84.44	0.070	96.81	88.44	0.009
relu5	92.14	85.56	0.077	92.14	85.56	0.079	91.13	85.56	0.076	96.77	84.44	0.009
conv5	91.56	83.33	0.101	74.44	48.89	0.451	86.59	80.00	0.140	95.36	77.78	0.014
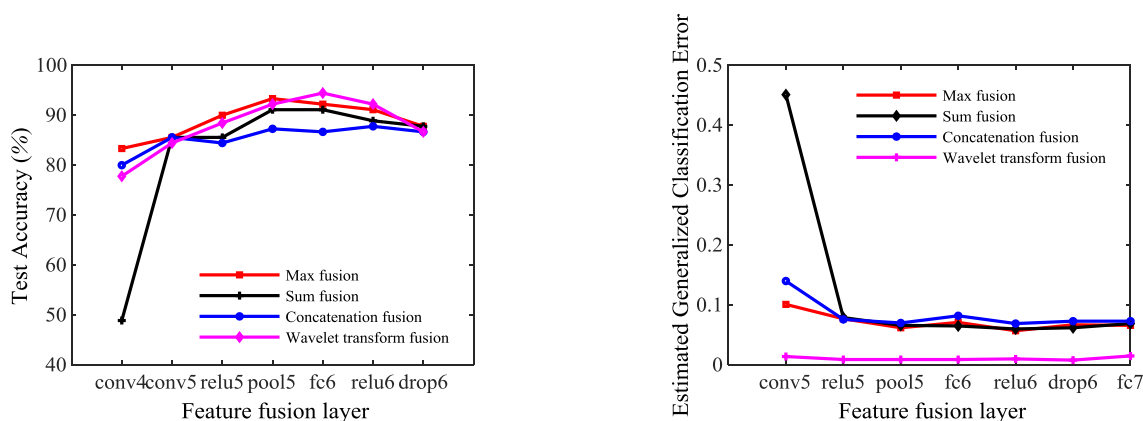
FIGURE 9. Comparison of the test accuracy and EGC-Error with different feature fusion strategies on different layers.

error comparison with different feature fusion strategies on different layers for test is shown in Figure 9(right).

As shown in Table 3 and figure 9, the results show that Maximum fusion and Wavelet transform fusion have the better performance, and the Concatenation fusion has the worst performance in this article through experiments. And the accuracy of the classification results by convolution layer feature fusion is obviously lower than that of FC6 layer in this case. And then the total training verification and test time on the whole data set for Maximum fusion strategy, Sum fusion, Concatenation fusion and Wavelet transform fusion are is 674.872s, 1380.963s, 1756.028s, 690.016s respectively, so the Maximum fusion and Wavelet transform fusion strategy has less calculation than others. However, the estimated generalized classification error with Wavelet transform fusion is the lowest. Therefore, from the analysis of experimental data, the features fusion model on the ReLU6 layer with Wavelet transform fusion is more suitable for defect detection of aluminum profile based on the data set in this article to achieve better results from a comprehensive point of view. The accuracy of the adopted method for classification verification is 96.37%, and for classification test is 94.44%.

The estimated generalized classification error is 0.010. The curve of the accuracy and loss value of the adopted model on the training set and verification set with the number of iterations increasing is shown in Figure 10.

The accuracy curve increases and the Loss curve decreases smoothly with the increases of the training iteration numbers, and the curves are relatively balanced and have little fluctuation. As shown in Figure 9 and 10, the proposed method can obtain a satisfactory result which has good convergence effect in 8 epochs and 384 iterations.

With the optimal network architecture determined, we now compare positive predictive values and true positive rates for each class in the proposed method, and the specific data is shown in Table 4. The comparisons on evaluation metrics including precisions, recalls, specificity and F1-Score for each aluminum profile defect class are shown in Figure 11.

As can be seen in Table 4 and Figure 11, abrasion mark is easily mistaken for convex powder, flawless sample, traffic mark and dirty points. And the small defect of traffic mark and abrasion mark or flawless sample with a defective background is easily mistaken for a dirty spot.
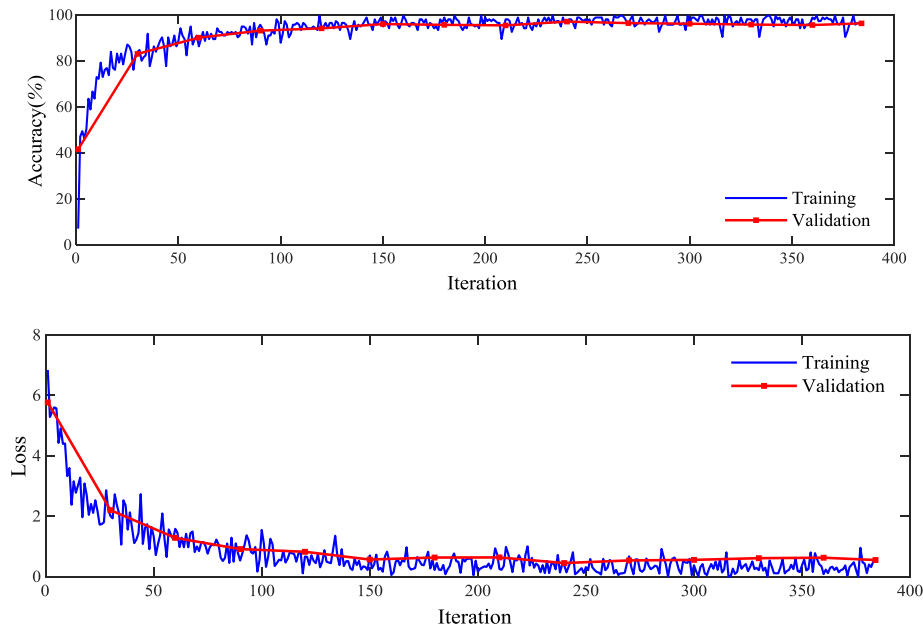
FIGURE 10. The accuracy and loss curve during network training process with Wavelet transform fusion strategy on relu6 layer.

TABLE 4. Comparison on evaluation metrics for each class.

Category	class-wise Precisions(%)		class-wise Recalls(%)		Specificity(%)		F1-Score	
	Verify	Test	Verify	Test	Verify	Test	Verify	Test
non-conducting	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
abrasion mark	81.1	90.0	93.8	90.0	98.5	98.8	87.0	90.0
horizontal stripe shallow recessing	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
orange peel	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
traffic mark	85.0	80.0	100.0	100.0	99.4	97.6	91.9	88.9
crater formation	76.9	100.0	100.0	100.0	99.4	100.0	87.0	100.0
Coating Cracking	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
dirty points	96.7	90.0	95.1	90.0	99.5	98.8	95.9	90.0
convex powder	90.9	80.0	100.0	100.0	99.8	98.8	95.2	88.9
Flawless sample	99.3	100.0	95.8	87.0	98.9	100.0	97.5	93.0

Obviously, according to the above results, the two-stream network has made good use of the structure and visual information of multiple streams to obtain effective features. The performance of a high-level feature fusion scheme for the two-stream network is better than that of the low-level feature fusion scheme or the single-stream network. Gradient image plays an important role in improving the ability of image representation and recognition. The proposed fusion scheme of Wavelet transform fusion strategy achieves the better fusion effect.

The above results show that the accuracy of most early feature fusion schemes is less than 83%. In contrast, the accuracy of later feature fusion has been improved by about 10%. The feature fusions on FC6, ReLU6 and Drop6 layers achieve the better detection accuracy of 92%. The performance of the adopted method is better than that of the single stream network, which improves accuracy by about 3% and about 7% respectively. Experimental results show that the proposed network structure can effectively achieve better detection accuracy.

D. OPTIMIZATION INFLUENCE

We often need to optimize the system parameters to achieve desired detection results. This section shows how to use hyperparameters optimization to optimize a classification model with training predictor and response data. Experimentally, the optimum model and its hyperparameter values are obtained by using Bayesian Optimization function and computing the cross-validation classification error, which are expected to classify new data optimally. Due to the over fitting phenomenon caused by hyperparameters adjustment, we often compare the trained optimizable model which does not
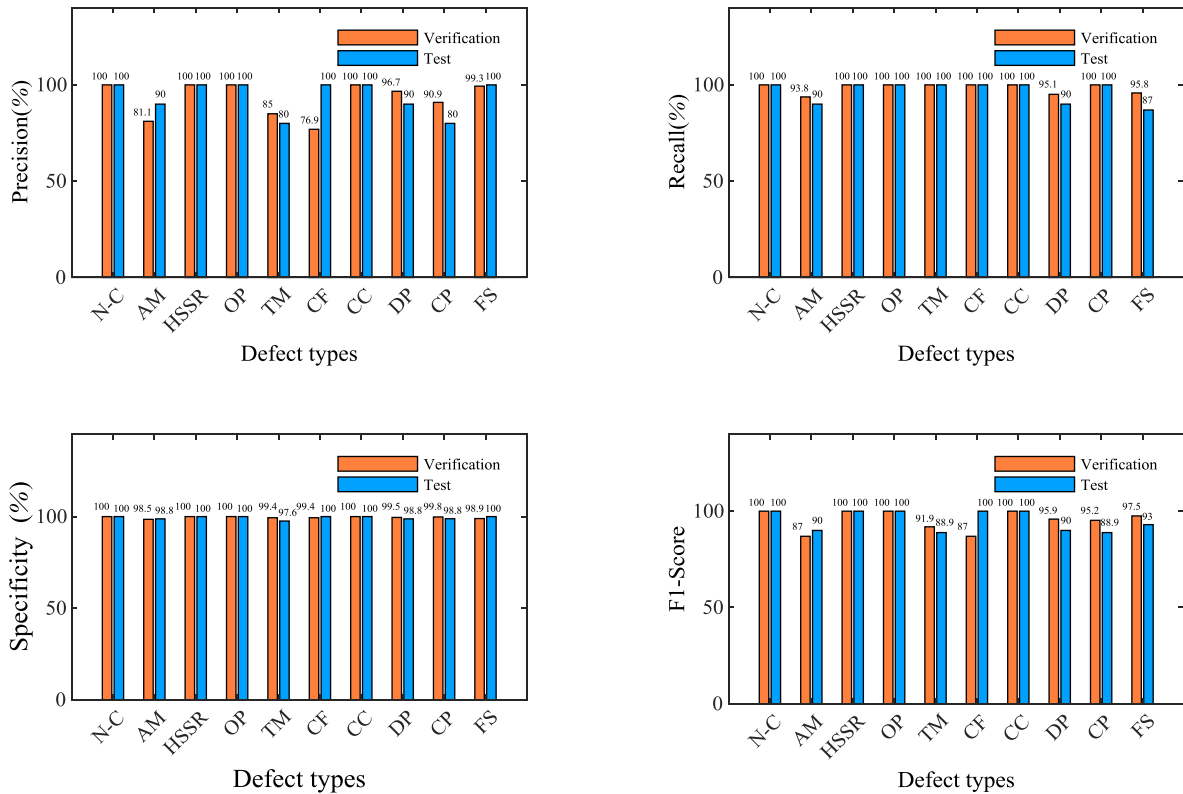
FIGURE 11. Comparisons on evaluation metrics for ten aluminum profile defect types.

always have a higher accuracy to the previously trained models and run the optimization for longer to try to get better results.

In the optimization process, the accuracy and generalized classification errors are compared before and after optimization. And the higher accuracy and the lower error value indicate that the generalization of the classifier is better. The comparison of the accuracy and the generalized classification error is shown in Table 5. And then, the min observed objective is compared to estimated min objective, which is shown in Figure 12. And the total function evaluations is set as 30.

TABLE 5. Comparison of accuracy and estimated generalized classification error before and after optimization.

method	before optimization			after optimization		
	V-A	T-A	EGC-Error	V-A	T-A	EGC-Error
the proposed method	96.37	94.44	0.010	96.98	95.56	0.004

The curve trend of the minimum value of the estimated objective is nearly consistent with that of the minimum value of the observed objective. And the minimum estimated objective value is 0.001730 comparing with the minimum the observed objective value which is 0.001725. And the total optimization time is 2038.486 seconds.
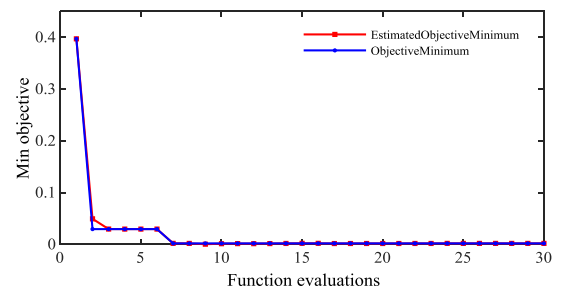

FIGURE 12. Comparison of the observed min objective with estimated min objective.

V. FUTURE WORK

As the proposed neural network system has not yet achieved the best effect, it is still necessary to improve the performance of defect detection. Firstly, due to insufficient samples of training a deep neural network, more various aluminum profile surface defects samples will be provided for training the network in the future. Secondly, in order to improve detection accuracy for some defect classes with poor performance, we will improve the neural network structure by conducting deeper analysis. We can consider adding a network stream based on the method in this article to realize multi-stream network. We can generate ROI around defects in aluminum section area to be as the input of the third neural network stream for multi feature fusion to obtain the better extracted features by region segmentation method or RPN (Region

FIGURE 13. Flawless sample with abrasion mark defect in background area.

FIGURE 14. ROI of Flawless sample with the red rectangle.

Proposal Network). For example, the defect-free aluminum profile image shown in Figure 13 has an abrasion mark defect indicated with the yellow rectangle in its background area, which is likely to be predicted as an abrasion mark sample during the training. Therefore, by extracting the smallest ROI which can cover the whole aluminum profile indicated with the red rectangle as shown in Figure 14, the defect location of the image can be analyzed intuitively, so the purpose of this method is to reduce the influence of similar defects in the background area on the detection accuracy.

Finally, the proposed network can also deal with the defect detection of other materials, such as steel surface defects, infrastructure surface cracks, rail surface defects and others.

VI. CONCLUSION
Firstly, we propose a novel two-stream convolutional neural network to extract features from RGB image stream and gradient image stream to effectively fuse multiple features for defect classification and detection of aluminum profiles, which is an important innovation of this article. And each network stream has a feature learning process from low level to high level. Secondly, we mainly introduce the image enhancement method and the method of obtaining gradient image for image preprocessing. Because deep learning mostly relies on the support of a large number of training data, the data enhancement methods are used to obtain a large number of samples in order to prevent the phenomenon of over fitting in deep learning. The image gradient is calculated by Sobel operator and normalized to transform the data between zero and one under the same dimension. Thirdly, a large number of experiments are carried out between the single-stream network and the two-stream network with different feature fusion strategies on different fusion layers using the original RGB image and the corresponding gradient image of

aluminum profiles to prove the effectiveness of the proposed two-stream network.

REFERENCES

[1] Z. Liu, L. Li, S. Li, J. Yi, and G. Wang, "Simulation analysis of porthole die extrusion process and die structure modifications for an aluminum profile with high length-width ratio and small cavity," *Materials*, vol. 11, no. 9, pp. 1517–1536, 2018, doi: 10.3390/ma11091517.

[2] A. Chondronasios, I. Popov, and I. Jordanov, "Feature selection for surface defect classification of extruded aluminum profiles," *Int. J. Adv. Manuf. Technol.*, vol. 83, nos. 1–4, pp. 33–41, Mar. 2016, doi: 10.1007/s00170-015-7514-3.

[3] 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.

[4] B. Zhang, M. Liu, Y. Tian, G. Wu, X. Yang, S. Shi, and J. Li, "Defect inspection system of nuclear fuel pellet end faces based on machine vision," *J. Nucl. Sci. Technol.*, vol. 57, no. 6, pp. 617–623, Jun. 2020, doi: 10.1080/00223131.2019.1708827.

[5] S. Gupta, K. Thakur, and M. Kumar, "2D-human face recognition using SIFT and SURF descriptors of face's feature regions," *Vis. Comput.*, pp. 1–10, Feb. 2020, doi: 10.1007/s00371-020-01814-8.

[6] Y. Wang, M. Li, C. Zhang, H. Chen, and Y. Lu, "Weighted-fusion feature of MB-LBPUH and HOG for facial expression recognition," *Soft Comput.*, vol. 24, no. 8, pp. 5859–5875, Apr. 2020, doi: 10.1007/s00500-019-04380-x.

[7] W. Xu, H. Parvin, and H. Izadparast, "Deep learning neural network for unconventional images classification," *Neural Process. Lett.*, vol. 52, no. 1, pp. 169–185, Aug. 2020, doi: 10.1007/s11063-020-10238-3.

[8] R. Wei and Y. Bi, "Research on recognition technology of aluminum profile surface defects based on deep learning," *Materials*, vol. 12, no. 10, pp. 1681–1694, 2019, doi: 10.3390/ma12101681.

[9] Z. H. Liu, H. B. Shi, and X. F. Zhou, "Aluminum profile type recognition based on texture features," *Appl. Mech. Mater.*, vols. 556–562, pp. 2846–2851, May 2014, doi: 10.4028/www.scientific.net/AMM.556-562.2846.

[10] F. M. Neuhauser, G. Bachmann, and P. Hora, "Surface defect classification and detection on extruded aluminum profiles using convolutional neural networks," *Int. J. Mater. Forming*, vol. 13, no. 4, pp. 591–603, Jul. 2020, doi: 10.1007/s12289-019-01496-1.

[11] Q. H. Li and D. Liu, "Aluminum plate surface defects classification based on the BP neural network," *Appl. Mech. Mater.*, vol. 734, pp. 543–547, Feb. 2015, doi: 10.4028/www.scientific.net/AMM.734.543.

[12] G. Fu, P. Sun, W. Zhu, J. Yang, Y. Cao, M. Y. Yang, and Y. Cao, "A deep-learning-based approach for fast and robust steel surface defects classification," *Opt. Lasers Eng.*, vol. 121, pp. 397–405, Oct. 2019, doi: 10.1016/j.optlaseng.2019.05.005.

[13] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *Int. J. Adv. Manuf. Technol.*, vol. 94, nos. 9–12, pp. 3465–3471, Feb. 2018, doi: 10.1007/s00170-017-0882-0.

[14] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Ann.*, vol. 65, no. 1, pp. 417–420, 2016, doi: 10.1016/j.cirp.2016.04.072.

[15] 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.2017.2778343.

[16] 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: 10.1155/2018/8639367.

[17] 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.

[18] 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.

[19] 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.

[20] A. Buslaev, V. I. Iglovikov, 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/info11020125.

[21] N. Neogi, D. K. Mohanta, and P. K. Dutta, "Defect detection of steel surfaces with global adaptive percentile thresholding of gradient image," *J. Inst. Eng. (India), Ser. B*, vol. 98, no. 6, pp. 557–565, Dec. 2017, doi: 10.1007/s40031-017-0296-2.

[22] 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.

[23] K. Zhang, Y. Zhang, P. Wang, Y. Tian, and J. Yang, "An improved sobel edge algorithm and FPGA implementation," *Procedia Comput. Sci.*, vol. 131, pp. 243–248, Jan. 2018, doi: 10.1016/j.procs.2018.04.209.

[24] J. Jiang, Z. Jin, B. Wang, L. Ma, and Y. Cui, "A Sobel operator combined with patch statistics Algorithm for fabric defect detection," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 2, pp. 687–701, 2020, doi: 10.3837/tiis.2020.02.012.

[25] 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.inpa.2017.09.005.

[26] R.-G. Zhou and D.-Q. Liu, "Quantum image edge extraction based on improved sobel operator," *Int. J. Theor. Phys.*, vol. 58, no. 9, pp. 2969–2985, Sep. 2019, doi: 10.1007/s10773-019-04177-6.

[27] Y. Fang, Q. Wu, and W. Kang, "A novel finger vein verification system based on two-stream convolutional network learning," *Neurocomputing*, vol. 290, pp. 100–107, May 2018, doi: 10.1016/j.neucom.2018.02.042.

[28] Y. Hu, A. Soltoggio, R. Lock, and S. Carter, "A fully convolutional two-stream fusion network for interactive image segmentation," *Neural Netw.*, vol. 109, pp. 31–42, Jan. 2019, doi: 10.1016/j.neunet.2018.10.009.

[29] A. KrizhevskyI Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[30] T. Shanthi and R. S. Sabeenian, "Modified alexnet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56–64, Jun. 2019, doi: 10.1016/j.compeleceng.2019.03.004.

[31] X. Han, Y. Zhong, L. Cao, and L. Zhang, "Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," *Remote Sens.-Basel*, vol. 9, no. 8, pp. 848–869, 2017, doi: 10.3390/rs9080848.

[32] 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.

[33] 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, 2020, doi: 10.1016/j.ndteint.2020.102246.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

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

[41] R. Z. Cabada, H. R. Rangel, M. L. B. Estrada, and H. M. C. Lopez, "Hyperparameter optimization in CNN for learning-centered emotion recognition for intelligent tutoring systems," *Soft Comput.*, vol. 24, no. 10, pp. 7593–7602, May 2020, doi: 10.1007/s00500-019-04387-4.

[42] D. Sun, H. Wen, D. Wang, and J. Xu, "A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm," *Geomorphology*, vol. 362, pp. 107201–107249, Apr. 2020, doi: 10.1016/j.geomorph.2020.107201.

[43] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage*, vol. 180, pp. 68–77, Oct. 2018, doi: 10.1016/j.neuroimage.2017.06.061.

[44] *Guangdong Industrial Intelligence Big Data Innovation Competition*. Accessed: Sep. 17, 2019. [Online]. Available: https://tianchi.aliyun.com/competition/entrance/231682/information?lang=en-us

[45] W. Nie, L. Qu, M. Ren, Q. Liang, Y. Su, Y. Li, and H. Jin, "Two-stream network based on visual saliency sharing for 3D model recognition," *IEEE Access*, vol. 8, pp. 5979–5989, 2020, doi: 10.1109/ACCESS.2019.2963511.

CHUNMEI DUAN is currently pursuing the Ph.D. degree with the Institute of Light Alloy, Central South University, Changsha, China. She is also a Senior Engineer. Her research interests include computer vision, deep learning, and defect classification and identification.

TAOCHUAN ZHANG received the Ph.D. degree from Central South University. He is currently an Associate Professor. His research interests include signal processing, deep learning, and fault diagnosis.

• • •