IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 3, 2020, accepted September 29, 2020, date of publication October 19, 2020,
date of current version November 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3032164

A Novel Framework for Classifying Leather Surface Defects Based on a Parameter Optimized Residual Network

JIEHANG DENG[1,2], JIAXIN LIU[1], CHANGZHENG WU[1], TAO ZHONG[1], GUOSHENG GU[1], AND BINGO WING-KUEN LING[3], (Senior Member, IEEE)

[1]School of Computers, Guangdong University of Technology, Guangzhou 510006, China
[2]School of Computer Science and Software, Zhaoqing University, Zhaoqing 526061, China
[3]School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China

Corresponding author: Guosheng Gu (gsgu@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61202267; in part by the National Key Research and Development Program of China under Grant 2016YFB0200602; in part by the Science and Technology Program of Guangzhou under Grant 201807010058; in part by the Key Areas Research and Development Program of Guangdong Province under Grant 2019B010139002; and in part by the project of Guangzhou Science and Technology under Grant 201902020007, Grant 202007010004, and Grant 201807010058.

ABSTRACT Existing works on leather surface defects are mainly focused on the defect detection without performing the classification automatically. On the other hand, the manual classification has the disadvantages including a misjudgment, the occurrence of cumbersome labors and a high cost. To address these issues, this paper proposes a framework for classifying the leather surface defects based on a parameter optimized residual network. In this framework, an ultra high definition imaging system is first utilized to obtain the images of the leather surface. Then, two important network parameters are optimized. They are the size of the image data set and the size of the sliding patch window. Here, the size of the image data set is determined by achieving the tradeoffs between the evaluated workload and the classification accuracy. On the other hand, the size of the sliding patch window is obtained by the least squares method. In order to mitigate the limitation of the gradient disappearance in a deep network, a residual module is introduced to the proposed framework. Finally, the defect classification is performed based on the extracted features using multi-layer convolution and pooling operations. Computer numerical simulation results show that the leather surface defects can be classified effectively by the proposed framework. Also, the classification accuracy can reach 94.6%.

INDEX TERMS Automatic classification, leather defects, parameter optimized, residual network, sliding patch window.

I. INTRODUCTION

With the improvement of the living standards, most consumers' concept of leather products is changing from the practicality to the high quality. Also, the consumers have higher requirements for the quality and aesthetics of leather products [1]. The fine leather products must meet the highest quality standards, e.g., the upholstered furniture or interiors of top brand luxury cars [2]. Yi *et al.* [1] summarized the current situation of leather surface defect detection. They found although various technologies of the industry are constantly improving, the current industrial development of the leather products is still relatively low. The reason is that the automation level of the defect detection of leather surface is relatively low. In order to improve the quality of leather products and make more rational use of leather materials, defect detection and classification of leather surface becomes much essential. However, the work of detecting and classifying these surface defects has been basically done manually, which is subjective and low efficiency and low accuracy.

With the development of computer vision, automatic detection and classification of leather surface defects has become a crucial issue in leather industry. In order to address this issue, researchers have proposed various detection

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

and classification algorithms based on computer vision. Tsai and Huang [3] formulated the surface inspection issue as a thresholding problem by eliminating the periodic, repetitive patterns of the statistical textures using Fourier Transform. However, they only decided whether there was a defect on the surface or not, lacking of the decision of what type of the defect was. Tsai and Hsiao [4] proposed a multiresolution method by using wavelet transforms for inspecting the surface defects. However, this approach was a supervised one, the key parameters must be manually predetermined for each texture class. Yang [5] presented a modified FCM (fuzzy c means clustering) based on neighbor information for detecting local defects of leather surface. The feature vectors of the leather texture were extracted by a co-occurrence matrix. Then the leather defects can be detected by clustering the feature vectors using fuzzy clustering. However, they only decided whether there was a defect on the surface or not, lacking of the decision of what type of the defect was, also. Chen Hong [6] obtained the defect regions of leather surface by combing a co-occurrence matrix and improved Fuzzy C-means based on texture based analysis. However, the proposed method was only validated by experiments qualitatively.

Most importantly, the leather surface is detected only to determine whether it is defective or not by the approaches above. However, in the leather industry, the leather surface is detected not only to determine whether it is defective or not, but also to obtain the defect types precisely. Jian *et al.* [7] performed the leather defect detection and classification by combining a feedforward neural network and an improved decision tree. However, Li's approach needed to choose many image features to classify the local leather defects empirically. Similarly, Filho *et al.* [8] evaluated the leather quality based on an automated system with more than eight handcraft features and performed the SVM classifier to classify the leather into seven quality levels. However, the selection of the handcraft features is in a subjective manner. However, in industrial applications, leather defect detection is usually carried out on the entire piece of leather instead of performing the detection locally. Therefore, how to classify the surface defects on the whole piece of leather automatically and objectively is still a crucial issue in leather industry.

In recent years, applying deep learning networks in the automatic classification has been employed in various application areas. In the field of road transportation, it was of great significance in monitoring the traffic flow. The deep convolution neural network was used to recognize the vehicle types successfully [9], [10]. In the medical field, the pathological features were extracted from the pathological images for performing the diagnosis of different stages of diseases by a deep convolution network. This could be applied to prevent and effectively diagnose the infections of various diseases [11]–[15]. In the agriculture, the convolution network was used to classify the patches of images for the diagnosis of the growing stages of the diseases [16], [17]. In summary, the deep learning is an effective way to provide the automatic and objective image classification.

Therefore, in order to classify the surface defects on the whole piece of the leather automatically and objectively, a novel framework is proposed in this paper. The framework is based on a parameter optimized residual network. The defects of the leathers will be automatically divided into five categories: the scratches, the rotten surfaces, the holes, the needle eyes and no defects. The classification results are able to effectively improve the subsequent leather processing quality.

On the other hand, a sliding patch window is employed to obtain more sample images in our proposed framework. The sliding window is the window cropping the collected ultra-high definition leather sample image with a specific pixel size according to certain moving rules. The defects of the leather are inside the small regions on the leather sample images. A too small sliding patch window cannot fully contain the defect area, which will make the corresponding defect feature extraction inaccurate in the network training. On the contrary, a too large sliding patch window may weaken the effectiveness of the extracted defect features. Thus, how to determine the appropriate size of the sliding patch window is a challenge issue. In order to solve this problem, the comparative computer numerical simulations of different sizes of the sliding patch windows are conducted. Based on the computer numerical simulation results, the final size of the sliding patch window is determined by the least squares method.

In the deep learning network, the small training data sets may result in overfitting. On the contrary, the large-scale training data sets can improve the network training precision, but it will greatly increase the computer numerical simulation workload. Therefore, the appropriate size of the data set is determined by examining the computer numerical simulation results on different scales of data sets in this paper.

A. MOTIVATION AND CONTRIBUTION

The ResNet is an important tool for classification. Chen and Zhang [18] proposed a hybrid model for compressive image recovery by integrating the ResNet. Duan *et al.* [19] monitored the machine health by employing the ResNet. Ou *et al.* [20] detected the moving objects via the ResNet. There are a lot of other authors used similar approaches in different applications including industry. Specifically, Korbar *et al.* [12] presented an effective deep residual network for performing the classification of the colorectal polyps on the whole slide images. In the images of the colorectal polyps, the colors, the structures and the textures of both the normal and the abnormal tissues were appeared with the similar degrees. The sizes and the shapes of the colorectal polyps were different. Also, the same kinds of colorectal polyps were with different sizes and shapes. In this case, the colorectal polyps were difficult to be classified because of the interference from the normal tissue with various sizes and shapes. Despite of these difficulties, various colorectal

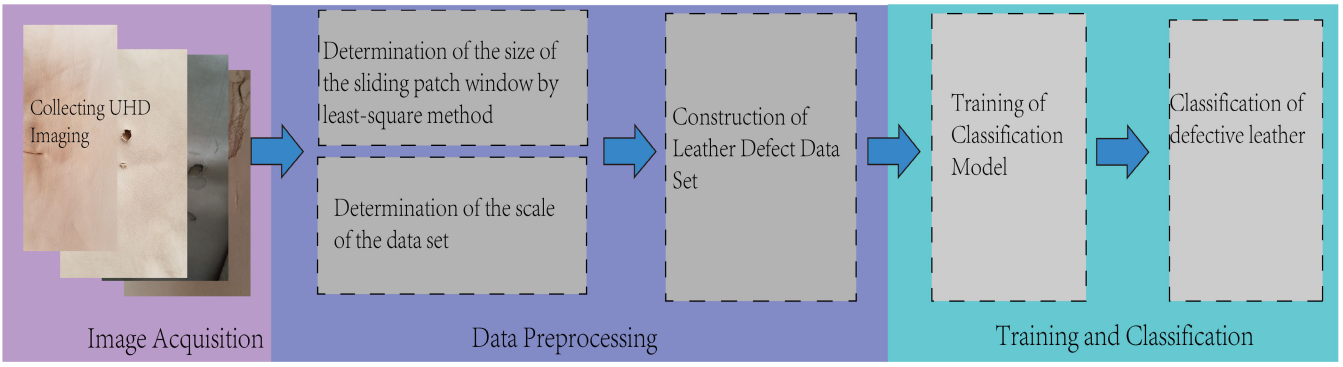

FIGURE 1. The framework for classifying the leather defects automatically and objectively. The leather samples were obtained using ultra high definition (UHD) imaging system. The data sets are obtained by cropping the original images by a sliding window. The data sets consist of the training sets and the test sets. The training set and the test set are constructed by performing the random extraction of the non-repetitive images from the data set. The ratio of the size of the test set to the size of the training set is set at 1:5. There were 2500 and 500 images for training and testing after the size of the dataset and the size of the sliding path window are determined.

polyps could be classified effectively by the Korbar's residual network.

Similar to the features of the colorectal polyps on the whole slide images, the leather images also have a regular texture as a whole. These kinds of defects (the scratches, the rotten surfaces, the holes and the needle eyes), the sizes, the shapes, the textures and the structures of defects are also similar to the polyps of the pathological images. In the view of the similarity of these two classification scenarios and the effective application of the deep residual network for classifying the colorectal polyps, a novel framework for classifying the leather surface defects based on a parameter optimized residual network is proposed in this paper. The main contributions are as follows:

(1) To solve the problem that the current researches [3]–[6] conducted on the leather defect classification did not classify the types of the leather defects automatically and objectively, a new framework of the classification of the leather surface defects based on a parameter optimized residual network is proposed.

(2) The least squares method is used to fit the computer numerical simulation data to determine the appropriate size of a sliding patch window.

(3) The appropriate size of the data volume is verified by the computer numerical simulations.

(4) Verify the performance of the proposed parameter optimized residual network by using the same leather defect dataset. The computer numerical simulation results show that the proposed network outperforms the baseline comparison methods, e.g. Lenet5 [21], Caffenet [22], Faster R-CNN [23], ConerNet [24], and can effectively classify the leather defects automatically and objectively. The accuracy of our proposed network can reach 94.6%.

II. PROPOSED FRAMEWORK

The framework proposed in this paper first obtains the ultra high definition (UHD) leather images through the UHD leather imaging system. Second, the least squares method is used to fit the computer numerical simulation data to determine the appropriate size of the sliding patch window. Also, the scale of the data set is determined by achieving the tradeoffs between the evaluated workload and the classification accuracy. Then, the residual network is constructed and other related parameters are optimized. Finally, the network is trained and the leather defects are classified into different types as illustrated in Figure 1. In the proposed framework, the features of the leather defects are extracted by the deep learning network automatically and objectively.

A. UHD LEATHER IMAGING SYSTEM

At present, the real-time detection of the leather surface defects is mainly based on the CCD (Charge Coupled Device) scanning and the imaging system. CCD is sensitive to light and used as the light-detecting component in digital camera. The imaging field of the common CCD camera is small and limited. It needs to move the CCD camera by a complex control system to scan the effective area of the whole piece of the leather. After taking the image, the image fusion technique is needed to apply to the scanned images for obtaining the whole piece of the leather image.

In order to address the problems of the small imaging field and the complex control devices of the common CCD scanning and the imaging system of the leather, a high definition and high efficiency leather imaging system is adopted. The system contains a fixing platform for placing the leather, an UHD camera, the light sources and the image processing workstation. In order to distinguish the leather from the background, the color of the fixing platform is set to the blue color according to the statistics. During the imaging procedure of the leather image, the leather is laid and fixed on a platform flatly. The UHD camera is employed to take pictures of the leather above the platform with ultrahigh definition. The size of the obtained leather image is $8688 * 5792$ pixels. After the leather is imaged by this UHD camera, the system transmits the image to the image processing workstation for performing the detection and the classification. Compared with the

common CCD imaging system of the leather, this system has the advantages of the simple structure, using only one-time imaging for the whole piece of the leather, not requiring the image fusion technique as well as fast, simple and effective process of images.

B. OPTIMIZING THE SIZE OF THE SLIDING PATCH WINDOW BY THE LEAST SQUARES METHOD

In order to have an appropriate size of the sliding patch window, the size of the window is increased progressively and tested by our framework. In particular, different datasets with different patch sizes are constructed by setting different sizes of the patch windows. The accuracy of the defect classification is figured out from these datasets. Then, the least squares method is applied to fit the relationship between the size of the sliding patch window and the accuracy results. Consequently, the appropriative size of the sliding patch window is obtained by finding the extreme point of the fitted curve. The description of the fitted curve is discussed below:

Let the fitted curve be f(x). It is a polynomial whose order is less than the sampling points as shown in:

$$f(x) = a_1 x^m + a_2 x^{m-1} + \cdots + a_m x + a_{m+1}. \quad (1)$$

where $a_1, \cdots, a_n$ are the coefficients determined by the least squares method. x is the size of the sliding patch window and f(x) denotes the accuracy of the defect classification. Assume that there are n discrete data points (x_i, y_i) for $i = 1, \cdots, n$. Here, x_i for $i = 1, \cdots, n$ are the sizes of the sliding patch windows and y_i are the accuracies of the defect classifications obtained from the computer numerical simulations. According to the least squares method, a fitted curve f(x) can be obtained by finding the minimum value of d, where d is the variance which is defined as:

$$d = \sum_{i=1}^{n} (f(x_i) - y_i)^2. \quad (2)$$

C. DEEP RESIDUAL NETWORK ARCHITECTURE

Compared with the common convolutional neural network such as the LeNet [21], the AlexNet [25], and the GoogleNet [26], the ResNet introduces a residual structure. This can solve the degradation problem caused by the increase of the network layers and improve the classification accuracy [27]. Thus, the residual structure is employed in our proposed framework. The details of the flow chart of the leather defect classification algorithm based on the deep residual network are shown in Figure 2. The module in the algorithm includes the image acquisition module, the data preprocessing module, the model training module and the defect classification module.

In our framework, a classification architecture based on the deep residual network is designed. It is shown in Figure 3. The proposed network extracts the features by the convolution and the pooling of samples. Finally, it classifies the sample images by a classifier.

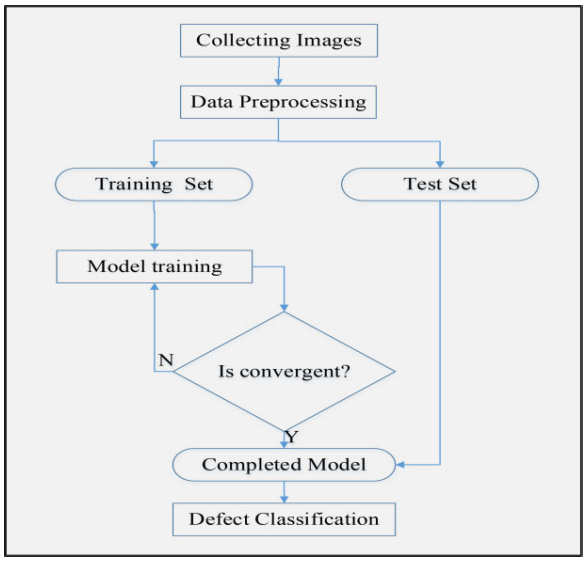

FIGURE 2. The flow chart of the leather defect classification based on the deep residual network. The module of data preprocessing can be split into three parts: (1) determination of the size of the sliding patch window by least-squares method, (2) determination of the scale of the data set and (3) construction of dataset for training and testing. Before classifying the leather defects, the deep residual network is trained by the leather training dataset until it is convergent. Then the leather images in the test set are fed to the trained model to be classified into different types.

The employed network is constructed by the input layer, the convolution layers, the pooling layers, the residual module and a softmax classifier as shown in Figure 3. The input layer is the initial layer of the network which loads the leather image patches as the input of the next network layers. The others are described in details below.

1) CONVOLUTION LAYERS

The features of the input images are extracted by the convolution layers. The input image is convoluted with a set of the convolution kernels. The outputs of the activation function are a set of the feature maps. Different feature maps represent different features of the input image. A feature map is obtained using a convolution kernel. That is, the weight averaging. The operation is shown in:

$$FM_j^{(L)} = f\left(\sum_{i \in M^{L-1}} FM_i^{(L-1)} * k_j^{(L)} + b_j^{(L)}\right). \quad (3)$$

where FM is the feature map and f is the activation function. To prevent the disappearance of the gradient, the ReLU is used as the activation function. L is the total number of layers, k_j^L is the convolutional kernel at the n^{th} layer feature map, M^{L-1} is the leather feature map at the $(L-1)^{th}$ input layer, $*$ denotes the convolution operation, and b is the bias term. In these convolutional layers, the local connection is applied to reduce the total number of parameters and the complexity of the algorithm. At the same time, the multi-convolution kernels are applied to obtain different features of the images.

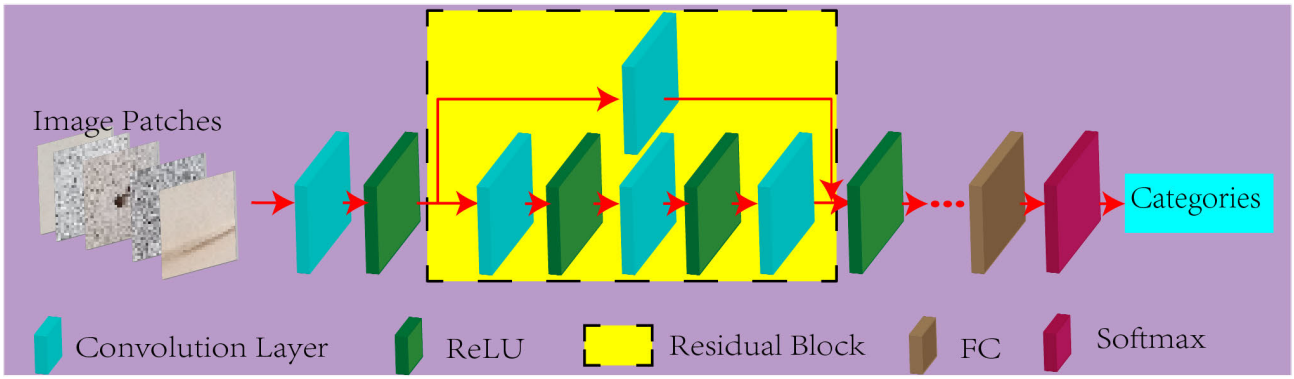

FIGURE 3. The residual network structure denoted by a dotted yellow rectangle.

2) POOLING LAYERS

The pooling layers aggregate the adjacent areas of the leather feature map by performing the downsampling. This mainly includes the average and the maximum pooling. An average pooling layer outputs the mean values of the rectangular regions. Similarly, the maximum pooling layer outputs the maximum values of the rectangular regions. The purpose of the pooling layer is to further reduce the network parameters and maintain certain feature invariances for the rotation and the translation.

3) RESIDUAL MODULE

Unlike the traditional neural networks, the residual structure is employed in the residual network. The residual structure is shown in the dotted yellow rectangle in Figure 3. It is realized by a neural block with a fast connection from the local input to the local output. The residual module does not increase the network parameters, but solves the vanishing gradient problem. Furthermore, the depth of the neural network becomes greater. Also, the computational complexity can be kept at a controllable level by introducing the residual module [27].

4) SOFTMAX CLASSIFIER AND THE LOSS FUNCTION

Finally, the extracted features can be classified by a softmax function. That is, the inputted softmax classifier is a common method for solving the multi-classification problems. Through this classifier, a one-dimensional vector can be obtained. The elements of this vector are in the interval (0,1). Also, the sum of these elements is equal to 1. They can be interpreted as the probabilities of different categories. The softmax classifier is defined in:

$$P(t) = \frac{e^{z_t}}{\sum_t^K e^{z_t}} \tag{4}$$

where P(t) is the probability corresponding to the t^{th} class, z_t is the input value corresponding to the t^{th} class and K is the total number of categories.

The loss function H applied in our deep network is the cross entropy as defined in:

$$H(\{y(t)\}, \{P(t)\}) = -\sum_t^K y(t) \log p(t) \tag{5}$$

where y(t) is probability of the ground truth class which is corresponding to P(t).

D. DATA VOLUME ENHANCEMENT

There are many parameters required to be adjusted during training the deep neural network. The purpose of the training is to optimize the parameters through the massive data. Training with a small volume of data leads to the overfitting of the network. In order to mitigate this limitation, the whole piece leather images are broken into the smaller and the overlapping patches. The neural network is applied to these patches. The data set in this paper is constructed by these patches. Compared with the volume of the original leather images, the volume of the cropped patches is much larger. The network parameters can be optimized by the cropped patches. The process of breaking down the whole piece of the leather image is shown in Figure 4.

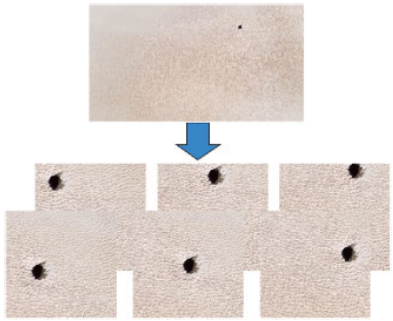

FIGURE 4. The schematic diagram of breaking down the whole piece leather image into the smaller and the overlapping patches. The size of the original leather image on the top is 8688 ∗ 5792 pixels. The size of the cropped ones is 536∗536 pixels.

III. COMPUTER NUMERICAL SIMULATION RESULTS AND DISCUSSIONS

The computer numerical simulations are performed on an ubuntu16.04LTS 64-bit PC equipped with an Intel Core i7 8700k 12 core 3.70GHz CPU and 16GB RAM. In order to reduce the training time, NVIDIA 1080ti GPU with 11 GB memory is used. Also, the caffe open source deep learning framework is used to implement the proposed deep learning model.

A. COMPUTER NUMERICAL SIMULATION DATA SET

In this paper, the original leather images are collected from the leather enterprises. The images obtained by the UHD leather imaging system are explained in section 2A. The resolution of the original leather image is 8688×5792 pixels. There are over 300 pieces of the leather images collected in the computer numerical simulations. Five types of common leather defects are considered in this paper. They are the scratch, the rotten surface, the holes, the needle eye and the defect free. All the defects in the collected data set are labeled by the factory technicians. In order to validate the proposed framework, three data sets are constructed as follows. Data set 1 is the data set for optimizing the size of the data volume. The size of Data Set 1 changes from 300 to 3000, step by step. Data set 2 is the data set for optimizing the size of the sliding patch window. In Data Set 2, there are 5 groups of sliding patch windows. There are 3000 images for each group of the sliding path windows. Therefore, there are 15000 images in Data set 2, totally. Data set 3 is the data set for performing the model training and achieving the performance improvement based on both the data set 1 and the data set 2. The size of Data Set 3 is 3000, in which there are 2500 images for training and 500 images for testing.

TABLE 1. Six kinds of data volume in Data set 1 (images).

Data	V 1	V 2	V 3	V 4	V 5	V 6
Train Set	250	500	1000	1500	2000	2500
Test Set	50	100	200	300	400	500

Note: V1 $\cdots$ V6 denote volume1 $\cdots$ volume6

Data set 1: The purpose of constructing this data set is to optimize the total number of images for both training and testing the deep learning network. The deep networks encounter the overfitting on the small data sets for performing the training. Although the deep network parameters can be adjusted thoroughly and correctly by a large data set, the computer numerical simulation workload increases as the data volume becomes larger. In order to achieve the balance between the computer numerical simulation workload and the classification accuracy, an appropriate size of the data set is required to be evaluated. In order to obtain the appropriate size of the data set, the relationship between the classification accuracy and the size of the data set is figured out by evaluating the classification accuracy with increasing the size of the data set. During this procedure, the size of the image patch is fixed at 224×224. As shown in Table 1, the size of the data set for performing the training and the testing increases from a small step to a large step. Based on this rule, six scales of images are constructed. There are five types of the leather defects in each scale of the data set. That is, the scratch, the rotten surface, the holes, the needle eye and the defect free. The total number of the image patches is the same for each type of the leather defects. For example, each of the scratch, the rotten surface, the holes, the needle eye and the defect free consists of 200 image patches from the training set of the 1000 image patches.

Data set 2: The purpose of constructing this data set is to optimize the size of the sliding patch window. The defect region of the leather is localized and small compared with the whole piece of the leather. In the computer numerical simulation, the defect features of the leather are extracted from the defect region. Therefore, whether the defect is included in the defect region completely or not is important to the feature extraction. In order to obtain the effective defect region, the whole piece of the leather is broken into the smaller and the overlapping patches. The size of the patch corresponds to the size of the defect region. However, if the patch size is too small, then the defect cannot be included in the patch completely. This leads to the extracted defect features lack of the comprehensiveness. On the contrary, if the patch is too large, then the proportion of the defect region is too small compared with the whole patch region. This leads to the extracted defect features lack of the validity. In order to solve this problem, a comparative computer numerical simulation of different sizes of the patch windows is carried out. The optimal size of the patch window is determined by fitting the computer numerical simulation data with the least squares method. In the computer numerical simulations, five groups of sliding patch windows are used to construct the computer numerical simulation data. They are 224×224, 336×336, 448×448, 560×560 and 672×672. The data set of each group of the sliding patch windows contains 3,000 images where 2,500 images are used for training and 500 images are used for testing.

Data set 3: The purpose of constructing this data set is to train and improve the classifying network based on the computer numerical simulation results of both the data set 1 and the data set 2. After carrying out the computer numerical simulations on both the data set 1 and the data set 2, the size of the whole data set and the size of the sliding patch window can be determined. The data set 3 is constructed based on the determined size of the whole data set and the determined size of the sliding path window.

B. TRAINING STRATEGY

In order to ensure the validity of the computer numerical simulation results, there is no intersection between the training set and the test set. The training set and the test set are constructed by performing the random extraction of the non-repetitive images from the data set. The ratio of the size of the test set to the size of the training set is set at 1:5. In addition, the advantages of the proposed network are verified by carrying out the comparative computer numerical simulations based on the same data set for different network models.

In the computer numerical simulations based on the data set 3, the parameters of the residual network are optimized. In the optimization process, the ''multi-step'' learning strategy is adopted to the update of the learning rate. After performing the validation of the computer numerical simulations, the adjustment level of the learning rate is set at 0.5 to achieve the better training results. For example, let the initial

learning rate be 0.1. The second adjustment of the learning rate is 0.05, and the third is 0.025.

C. EVALUATION CRITERIA

For the defect classification of the leather patches, the accuracy of the computer numerical simulation is defined in:

$$CA = \frac{N_r}{N} \times 100\%. \tag{6}$$

where N_r is the total number of the images correctly recognized in the test set and N is the total number of the test images. Through computing the classification accuracy (CA), the effects of various networks on the leather defect classification can be evaluated. Also, the validity of the network model can be determined.

D. COMPUTER NUMERICAL SIMULATION RESULTS AS WELL AS THE DISCUSSIONS ON THE PARAMETER OPTIMIZATION AND CLASSIFICATION

1) OPTIMIZING THE SIZE OF THE DATA SET

The optimization on the size of the data set is performed on the data set 1. A small data set will result to the overfitting of the deep network. Although the parameters of the deep network can be trained thoroughly by a large data set and the obtained accuracy can be very high, the larger the data set will result to the higher the computer numerical simulation workload. In order to obtain the appropriate size of the data set, the classification accuracy is checked based on the size of the data set shown in Table 2.

TABLE 2. Comparisons on different data sets.

	S1	S2	S3	S4	S5	S6
Training Set(images)	250	500	1000	1500	2000	2500
Test Set(images)	50	100	200	300	400	500
CA (%)	15	22	42.2	60.8	63	63.2
Increment of CA (%)	–	7	20.2	18.6	2.2	0.2

Note: All the size of the image in data sets are 224×224
$S1 \cdots S6$ stands for the scale1 $\cdots$ scale6. The scale denotes the size of the dataset.

As shown in Table 2, when the scales of the training data are set to 250, 500 and 1000, the classification accuracy is only 15%, 22% and 42.2%, respectively. However, with the increase of the size of the data set, the classification accuracy increases from 15% to 42.2%. This corresponds to 2.8 times of the improvement. When the size of the training data set changes from 1500 to 2500, the accuracy increases slowly from 60.8% to 63.2%. This shows that the increment of the classification accuracy increases from 7% to 20.2%. Then, it decreases to 0.2% for the size of the training data set increased from 250 to 2500 while other parameters remain unchanged. Specially, when the size of the training data set is increased to 2500, the increment of the classification accuracy is only 0.2%. This refers to the saturation case. Therefore, the size of the data set is set to 3000 including 2500 images in the training set and 500 images in the test set. Besides, the tradeoff between these two factors for evaluating the workload and the classification accuracy is shown in Table 2.

2) OPTIMIZING THE SIZE OF THE SLIDING PATCH WINDOW

The sliding patch window is the window with a specific pixel size to crop the collected leather sample image of ultra-high definition according to certain moving rules. In this study, the size of the high-definition leather image of is $8688 * 5792$ pixels, which will cause the following problems. (1) A larger input image will increase more computation burden, cause more time consumption and occupy more computation resources. (2) The input image needs to be normalized to the image with certain size for the deep learning network. Normalizing the size of the holistic piece of large leather image may destroy the shape and the content of the defect regions. That may lead to incomplete or wrong feature extraction. (3) Most of the areas in the leather image to be classified are the areas without surface defects. It is difficult to collect a large number of leather images with surface defects. Therefore, we crop the holistic piece of leather image into many patches assisted by the sliding path window.

The computer numerical simulations on optimizing the size of the sliding patch window is carried out on the data set 2. Five kinds of the sliding patch windows are designed by increasing the window size step by step. The classification accuracy is evaluated based on these data sets. The results are shown in Table 3.

TABLE 3. Comparisons based on different sizes of the sliding patch window.

Sliding-patch window size	224× 224	336× 336	448× 448	560× 560	672× 672
CA (%)	63.2	63.6	76.6	71.4	70.2
Increment of CA (%)	–	0.6	20.4	−6.8	−1.7

As shown in Table 3 that the size of the sliding patch window increases from 224×224 to 672×672, the step length is 112. In order to figure out the relationship between the size of the patch window and the classification accuracy, the classification accuracy is evaluated along with the change of the size of the sliding patch window while the other relative parameters remain unchanged. When the size of the sliding patch window is 224×224, the classification accuracy is 63.2%. With the increment of the sliding patch window, the accuracy increases. When the size of the patch window is 448×448, the classification accuracy increases to 76.6%. However, when the size of the patch window increases from 448×448 to 560×560, the accuracy deceases by 6.8%. When the sliding patch window increases to 672×672, the accuracy continues to decrease by 1.7%. That is because a too small patch window cannot fully contain the defect area. As a result, the corresponding defect feature extraction is inaccurate in the network training. Therefore, the accuracy based on the patch window with the size 224×224 is less than that with the size 448×448. On the other hand, a too large patch window

may weaken the effectiveness of the extracted defect features. Therefore, the accuracy becomes lower when the size of the patch window becomes larger such as the window size being equal to 448 × 448.

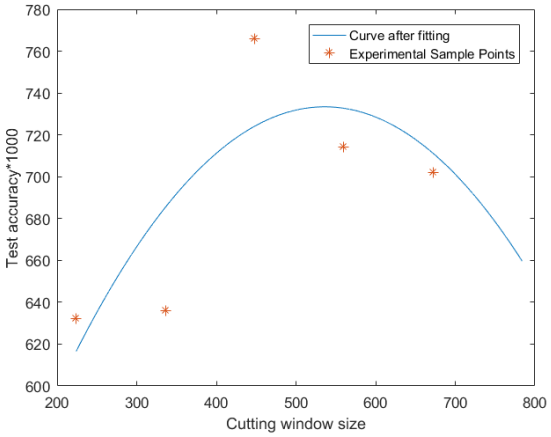

FIGURE 5. The itting urve between the size of the patch window and the classification accuracy.

It is worth noting that the relationship between the size of the patch window and the classification accuracy is convex. There is a change of the trend of the classification accuracy as the window size increases. This relationship has an extreme point. In order to locate this extreme point, the relationship between the window size and the accuracy is fitted by a polynomial based on the least squares method described in section 2B. The extreme point can be found based on the obtained polynomial as shown in Figure 5. In Figure 5, the x axis represents the size of the sliding patch window and the y axis represents the classification accuracy. In order to find the size of the patch window, the magnitude of the classification accuracy is magnified by 1000 times. According to least squares method, the relationship between the size of the patch window and the classification accuracy is shown in:

$$f(x) = -0.0012x^2 + 1.2865x + 388.8. \quad (7)$$

where x is the size of the sliding patch window, $f(x)$ is the fitted classification accuracy by the least squares method. To demonstrate the result more clearly, the fitted classification accuracy $f(x)$ is plotted with a blue line in Figure 5. The classification accuracy obtained by the computer numerical simulation is represented by the red stars. The extreme point of the fitted classification accuracy can be located by finding the maximum of the function $f(x)$. After the locating the extreme point, this extreme value is used as the appropriate size of the sliding patch window. In this example, it is found to be 536.

3) COMPARISONS WITH DIFFERENT DEEP NETWORKS

The comparisons with different deep networks are carried out based on the data set 3. After performing the computer numerical simulations and the analysis on the appropriate size of both the data set and the sliding patch window, the size of the data set for classifying the leather defects is set at 3000 patch images including 2500 images for training and 500 images for testing. The size of the sliding patch window is set at 536 × 536 based on the least squares method above. Inspired by the Korbar's work [12] for classifying the colorectal polyps, the ResNet-50 is applied as a deep residual network for classifying the leather defects. In order to verify the performance of the applied network, the Lenet, the Caffenet, the Faster R-CNN, the CornerNet, the SSD, the ResNet50-1 and the ResNet50-2 are used for performing the comparisons. The computer numerical simulation results are shown in Table 4. Here, the difference between the ResNet50-1 and the ResNet50-2 is that some of the network parameters in the ResNet50-2 were adjusted. The adjusted network parameters are the learning rate, the training iteration and the adjustment level of the learning rate.

TABLE 4. Comparisons with different deep networks.

Network Model	CA(%)	Comments
Lenet5	70	It only distinguishes the defect of leather, but fails to classify the defects in detail.
Caffenet	73	A network model that comes from Caffe.
Faster R-CNN	93.6	Defect marking in pixel-level is needed in network
CornerNet	69.6	preprocessing
SSD [28]	97.8	
ResNet50-1	81.6	ResNet-50 with unadjusted parameters
ResNet50-2	94.6	ResNet-50 with adjusted parameters.

In Table 4, the ResNet50-2 outperforms other deep networks based on the same data set and the same size of the sliding patch window. The classification accuracy of the ReNet50-2reaches 94.6%. This is much larger than that of the Lenet 5 and the Caffenet. For the Faster R-CNN and the SSD, the classification accuracy can reach 93.6% and 97.8%, respectively, after training by the same data set. However, both the Faster R-CNN and SSD are supervised by pixel-level labels, which is time-costly and needs cumbersome labors. Although the classification accuracy of the network applied in this work is less than SSD, the applied network is supervised only by image-level labels, which is time-efficient and easy to prepare.

In order to show the advantage of using the applied network, the relationships between the classification accuracy, the training loss and total number of iterations are plotted in Figure 6. In Figure 6, although the maximum of the total number of iterations is 10000, the network becomes static rapidly and remains almost static with the total number of iterations less than 2000 for the ResNet50-2. This shows that the ResNet50-2 has an advantage of the fast convergence.

The test results of each type of leather defects are shown in Table 5 below. The classification accuracy of leather sam-

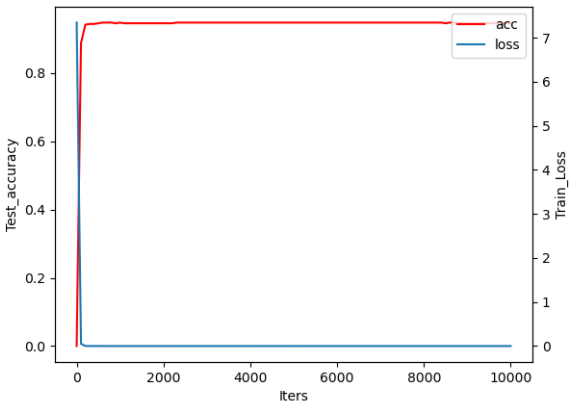

FIGURE 6. The relationships between the classification accuracy, training loss and the total number of iterations.

TABLE 5. Test accuracy(*CA*) of various types of defects by the proposed method.

	CA(%)
the scratches	96
the rotten surfaces	95
the holes	93
the needle eyes	92
Flawless	97
total	94.6

ples of each type is more than 92%, and the classification accuracy of scratch, rotten surface and defect free type is more than 95%.

IV. CONCLUSION

In the leather industry, the existing works are focused on the leather surface defects. It was mainly limited to the detection of the leather defects locally without classifying the defects automatically. Meanwhile, the leather defects are classified manually, this requires a very high workload and the process is subjective. In order to classify the surface defects on the whole piece of the leather automatically and objectively, a novel framework for classifying the leather surface defects based on a parameter optimized residual network is proposed. The proposed framework first obtains the UHD leather images through the UHD leather imaging system. Second, the volume of the data set is expanded to an appropriate scale by applying a sliding patch window on the original UHD images to avoid the overfitting of the insufficient samples. Two of the network parameters, the size of the data set and the size of the sliding patch window, are optimized. The size of the data set is obtained by achieving the tradeoffs between the evaluated workload and the classification accuracy. The size of the sliding patch window is achieved by using the least squares method to fit the computer numerical simulation data. Inspired by the Korbar's work [12] for classifying the colorectal polyps, the ResNet50 is applied as a deep residual network for classifying the leather defects. The other network parameters of the deep residual network are optimized through the computer numerical simulations. Finally, the residual deep network is trained and the leather defects are classified into different types. In order to demonstrate the effectiveness of using the applied network, different deep learning networks are used for performing the comparisons. The computer numerical simulation results show that the applied ResNet50-2 outperforms the Lenet5, the Caffenet, the Faster R-CNN, the CornerNet, and the ResNet50-1. It is time-efficient and easy to prepare while compared with the SSD. The classification accuracy of the applied reaches 94.6%. This meets the requirements of the automatic production in the leather industry. With the aid of our work, the surface defects on the whole piece of leather can be classified into different types automatically and objectively. The workers in leather industry can be released from the heavy workload of classifying the leather defects manually. And the procedure of classifying the leather defect need not be influenced by the workers' personal subjective factors.

REFERENCES

[1] S. Yi, Z. Ming, and G. Yanxin, "Current situation and outlook of leather surface defect detection," *West Leather*, vol. 37, no. 20, pp. 28–31, 2015.
[2] S. Chudzik, "Quality inspection of natural leather using non-destructive testing technique," *Quant. Infr. Thermogr. J.*, no. 134141, pp. 1–19, 2019.
[3] D.-M. Tsai and T.-Y. Huang, "Automated surface inspection for statistical textures," *Image Vis. Comput.*, vol. 21, no. 4, pp. 307–323, Apr. 2003.
[4] D.-M. Tsai and B. Hsiao, "Automatic surface inspection using wavelet reconstruction," *Pattern Recognit.*, vol. 34, no. 6, pp. 1285–1305, Jun. 2001.
[5] C. Yang, *Study on Image Detection Technique and its Application on Detecting Defects of Leather*. Hangzhou, China: Zhejiang Univ., 2004.
[6] C. Hong, *The Research of Leather Defect Detection Using Texture Analysis Techniques*. Chengdu, China: Sichuan Normal Univ., 2011.
[7] L. Jian, H. Wei, and Z. Dong, "Research of method for inspection and classification of leather surface defects," *Comput. Meas. Control*, vol. 18, no. 12, pp. 2733–2734, 2010.
[8] E. Q. Santos Filho, P. H. F. de Sousa, P. P. Rebouças Filho, G. A. Barreto, and V. H. C. de Albuquerque, "Evaluation of goat leather quality based on computational vision techniques," *Circuits, Syst., Signal Process.*, vol. 39, no. 2, pp. 651–673, Feb. 2020.
[9] D. Liu and W. Zijie, "Deep convolution neural networks for vehicle classification," *Appl. Res. Comput.*, vol. 33, no. 3, pp. 930–932, 2016.
[10] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.
[11] H. Xueying, H. Zhongyi, and W. Benzheng, "Breast cancer histopathological image auto-classification using deep learning," *Comput. Eng. Appl.*, vol. 54, no. 12, pp. 121–125, 2018.
[12] S. Hassanpour, B. Korbar, A. Olofson, A. Miraflor, C. Nicka, M. Suriawinata, L. Torresani, and A. Suriawinata, "Deep learning for classification of colorectal polyps on whole-slide images," *J. Pathol. Informat.*, vol. 8, no. 1, p. 30, 2017.
[13] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
[14] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognit. Syst. Res.*, vol. 57, pp. 147–159, Oct. 2019.
[15] J. Zhang, S. Pan, H. Hong, and L. Kong, "Blending ensemble of fine-tuned convolutional neural networks applied to mammary image classification," *J. Med. Imag. Health Informat.*, vol. 9, no. 6, pp. 1160–1166, Aug. 2019.
[16] L. Zhang, J. Jia, G. Gui, X. Hao, W. Gao, and M. Wang, "Deep learning based improved classification system for designing tomato harvesting robot," *IEEE Access*, vol. 6, pp. 67940–67950, 2018.
[17] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Appl. Artif. Intell.*, vol. 31, no. 4, pp. 299–315.

[18] Q. Chen and H. Zhang, "Hybrid model for compressive image recovery: Integrating ResNet-based denoiser into GAMP," *Signal Process.*, vol. 173, Aug. 2020, Art. no. 107583.
[19] J. Duan *et al.*, "A novel ResNet-based model structure and its applications in machine health monitoring," *J. Vib. Control*, 2020, Art. no. 107754632093650.
[20] X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, and W. Li, "Moving object detection method via ResNet-18 with encoder–decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108152–108160, 2019.
[21] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[22] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACMMM*, 2014, pp. 675–678.
[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[24] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 734–750.
[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[28] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 21–37.

JIEHANG DENG received the B.S. and M.S. degrees from the Xi'an University of Technology, China, in 2002 and 2005, respectively, and the Ph.D. degree from the University of Fukui, Japan, in 2009. He is currently an Associate Professor with the School of Computers, Guangdong University of Technology, China. His research interests include image processing and pattern recognition.

JIAXIN LIU is currently pursuing the M.S. degree with the School of Computers, Guangdong University of Technology, China. Her research interests include image processing and pattern recognition.

CHANGZHENG WU received the B.S. degree from Anhui Agricultural University, China, in 2016. He is currently pursuing the M.S. degree with the School of Computers, Guangdong University of Technology, China. His research interests include image processing and pattern recognition.

TAO ZHONG is currently pursuing the B.S. degree with the School of Computers, Guangdong University of Technology, China. His research interests include image processing and pattern recognition.

GUOSHENG GU received the M.S. degree in applied mathematics from South China Normal University, Guangzhou, China, in 2004, and the Ph.D. degree in computer application technology from the South China University of Technology, Guangzhou, in 2007. He is currently a Teacher with the School of Computers, Guangdong University of Technology, Guangzhou. His research interests include multimedia information security and image processing.

BINGO WING-KUEN LING (Senior Member, IEEE) received the B.Eng. (Hons.) and M.Phil. degrees from the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, in 1997 and 2000, respectively, and the Ph.D. degree from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, in 2003.

In 2004, he joined King's College London as a Lecturer. In 2010, he joined the University of Lincoln as a Principal Lecturer and promoted to a Reader in 2011. In 2012, he joined the Guangdong University of Technology as a Full Professor. He has published an undergraduate textbook, a research monograph, five book chapters, a book review published in an IEEE journal, more than 160 internationally leading journal articles, and more than 130 highly rated international conference papers. His research interests include time frequency analysis, optimization theory, symbolic dynamics of digital signal processing systems, biomedical signal processing, multimedia signal processing, and fuzzy and impulsive control theory. He is also a Fellow of IET, a China National Young Thousand-People-Plan Distinguished Professor, a Guangdong Province Pearl Scholar Distinguished Professor, and a University Hundred-People-Plan Distinguished Professor. He also serves in the technical committees of the nonlinear circuits and systems group, the digital signal processing group, and the power electronics and systems group of the IEEE Circuits and Systems Community. He was awarded the best reviewer prizes from the IEEE Instrumentation and Measurement Society in 2008 and 2012. He has also served as the Guest Editor-in-Chief of several special issues of highly rated international journals, such as the *Circuits, Systems, and Signal Processing*, and the *American Journal of Engineering and Applied Sciences*. He is also an Associate Editor of the *Journal of The Franklin Institute*, the *Circuits, Systems, and Signal Processing*, the *IET Signal Processing*, the *Measurement*, the *Measurement: Sensors*, and the *Journal of Industrial and Management Optimization*.