IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 22, 2020, accepted June 3, 2020, date of publication June 10, 2020, date of current version June 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001354

Fusing Convolutional Neural Network Features With Hand-Crafted Features for Objective Fabric Smoothness Appearance Assessment

JINGAN WANG[1], KANGJUN SHI[1], LEI WANG[1], ZHENGXIN LI[2], FENGXIN SUN[1], RURU PAN[1], AND WEIDONG GAO[1]

[1]Key Laboratory of Eco-Textiles, Ministry of Education, College of Textile Science and Engineering, Jiangnan University, Wuxi 214122, China
[2]School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China

Corresponding author: Weidong Gao (gaowd3@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0309200, in part by the National Natural Science Foundation of China under Grant 61802152, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180602 and Grant BK20180589, and in part by the Fundamental Research Funds for the Central Universities under Grant JUSRP52007A.

ABSTRACT In the textile and apparel industry, it remains a challenging task to evaluate the fabric smoothness appearance objectively. In existing studies, with computer vision technology, researchers use the hand-crafted image features and deep convolutional neural network (CNN) based image features to describe the fabric smoothness appearance. This paper presents an image classification framework to evaluate the fabric smoothness appearance degree. The framework contains a feature fusion module to fuse the hand-crafted and CNN features to take both advantages of them. The framework uses the multi-scale spatial masking model and a pre-trained CNN to extract hand-crafted and CNN features of fabric images respectively. In addition, a mislabeled sample filtering module is set in the framework, which helps to avoid the negative impact of mislabeled samples in training. In the experiments, the proposed framework achieves 85.2%, 96.1%, and 100% average evaluation accuracies under errors of 0 degree, 0.5 degree, and 1 degree respectively. The experiments on the feature fusion and mislabeled sample filtering verified their effectiveness in improving the evaluation accuracies and the label noise robustness. The proposed method outperforms the state-of-the-art methods for fabric smoothness assessment. Promisingly, this paper can provide novel research ideas for image-based fabric smoothness assessment and other similar tasks.

INDEX TERMS Fabric smoothness, textile testing, convolutional neural network, feature fusion, mislabeled sample filtering.

I. INTRODUCTION

Fabric smoothness after laundering is treated as a vital characteristic of the fabric to evaluate the tendency for the fabric to wrinkle, which quantizes the wrinkles on the fabric after being subjected to laundering procedures, has a bearing on 'ease-of-care' related properties (durable press, easy-care, minimum-iron, after wash appearance, etc.) in the textile and garment industry [1]. In the current industry application, the smoothness of fabric samples after laundering is mainly evaluated subjectively by trained testers, according to the similarity of the fabric sample and the appearance of the smoothness replicas in six grades (provided by the standards [2], [3] as shown in Figure 1). However, the subjective testing is easily disturbed by the test environmental factors and the testers' physiological and psychological states, and has the disadvantages of low accuracy, poor stability, high cost, and poor reproducibility [4]. With the increasing labor costs and the continuous improvement of production efficiency and quality requirements of the textile industry, subjective method terns to be difficult to meet the needs of practical applications. Therefore, it has become a great demand for the textile industry to propose an objective, stable, accurate and efficient method for evaluating the fabric smoothness appearance. In recent years, researchers in the field uses machine vision technology to establish subjective evaluation methods. In these studies, fabric smoothness evaluation is defined as image classification task. Their studies

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

generally include three main steps: fabric appearance data acquisition, smoothness features extraction, and smoothness degree classification.

In terms of fabric appearance data acquisition, the existing methods can be divided into two categories, namely 2D methods and 3D methods. The 2D methods [5]–[12] use an industrial camera or a scanner to collect the fabric images within shadow caused by wrinkles on the fabric surface. 3D methods utilize techniques as laser triangulation [13]–[17], photometric stereo vision [18], [19] and binocular stereo-vision [20]–[23] to reconstruct the fabric surface. For the 3D methods have better adaptability to the image acquisition environment and fabric color patterns, they attracted more attention in recent years. However, the 3D methods have the disadvantages of low efficiency, high cost, high calibration complexity, and unreliable accuracy. Our previous research [24] built and optimized a fabric 2D image acquisition platform. The possibility and effectiveness of the objective evaluation for the smoothness appearance of fabric without color pattern on the basis of 2D fabric images was demonstrated. In addition, to extend the variety adaptability of 2D methods, another study [25] has proposed a decoloration method for images of multi-color fabrics, which can eliminate the color pattern in the fabric images while preserve the wrinkles information in images. Therefore, the task of fabric smoothness appearance evaluation after image acquisition can be divided into two main steps, i.e. multi-color fabric image decoloration and monochrome fabric smoothness appearance evaluation. This paper is mainly focus on the second step of the main task.

In terms of feature extraction, a lot of hand-crafted features for 2D images are proposed in early years, which are constructed by edge and shade area [5], grey level distribution [6], gray level co-occurrence matrix (GLCM) [7], [8], discrete Fourier transform (FFT) [11], or wavelet decomposition [12]. For 3D methods, features are extracted based on the fractal dimension [20], depth pixel distribution [26], shape and distribution of wrinkles [15], [21], [27] of the 3D images. These hand-crafted features are lack of the description of human vision. However, the fabric smoothness appearance valuation task is closely related to human visual perception. A recent study for 3D image feature extraction [4] proposes to use the dense-sift features encoded by linear spatial pyramid matching using sparse coding(ScSPM), which simulates the sparse structure of human visual system (HVS) [28]. For the 2D methods, our previous study [29] established a multi-scale spatial masking model to extract hand-crafted features for fabric smoothness appearance, which models the spatial masking effect of the HVS.

Differing from the hand-crafted features constructed by human, convolutional neural networks (CNNs) can extract image features by self-driven learning, which are built with reference to the hierarchical information processing structure of the HVS [30]. CNNs have been widely used in different computer vision tasks and achieved a superior performance. Our previous study [31] explored the application of the deep CNN models in the evaluation of the fabric smoothness appearance. It was proved that the features extracted by the CNN model have great ability to describe the fabric smoothness appearance. With the various image features, different classification models have been applied, e.g. neural network [7], [14], support vector machine [4], [12], logistic regression [27], and CNN. However, it has not yet reached the requirements for industrial application, and the prediction performance of existing methods need to be further improved.

The subjective evaluation of the fabric smoothness appearance is a response of the HVS to optical stimuli, which includes quantitative feedback on the strength of the visual stimulus, and the prediction of the possibility of wrinkle recovery according to the human experience. This process involves the bottom-up visual perception patterns in the HVS, which are spontaneously constructed with simple visual stimuli such as brightness, contrast, edge, and shape. It also involves the top-down prior knowledge of the tester, such as the prediction of the wrinkle recovery trend, the classification of smoothness appearance level and other highly abstract visual perception patterns.

According to the feature extraction mechanism, most hand-crafted features consider the bottom-up perception only, while the CNN features contain the top-down knowledge in their task-driven training. Combining the two can describe the fabric smoothness appearance evaluation process more comprehensively. As the research basis, our previous studies have proposed a set of effective hand-crafted feature [29] and demonstrated the superiority of the CNN features [31] for fabric smoothness assessment. The experimental prediction results from these two types of features showed a certain complementarity in the sample distribution, which supports our viewpoints. Therefore, this study aims at combining the benefit of both and further improve the performance of the objective fabric smoothness appearance evaluation.

The idea of fusing CNN features and hand-crafted features to improve model performance is widely used in the field of existing computer vision, such as breast cancer detection [32], industrial superheat identification [33], ship classification [34], vehicle detection [35], person re-identification [36], osteoporosis diagnoses [37], lung nodule classification [38]. In the textile industry, fabric defect detection is the only task for which researchers have used the feature fusion idea [39]. According to their implementation, the feature fusion frameworks can be categorized into three types, i.e. feature level fusion [32]–[34] with simple concatenation of features; fusion by ensemble [35] with combining multiple classifiers trained by hand-crafted and CNN features in respective; joint level fusion [39] with combining different features in the CNN framework. In general, the feature level fusion fails to consider the difference of the dimension and range between the features, which may result in an unbalanced combination. The ensemble method always facing with the difficulty of training classifiers with the large dimensional CNN features. As result, considering the joint fusion,

FIGURE 1. AATCC smoothness appearance replicas.

we established a feature fusion module with learnt parameters to fuse the information from hand-crafted and CNN features in a self-adjusted way.

On the other hand, in the manual fabric smoothness appearance assessment, caused by the individual physical, psychological and environmental factors of human testers, the samples' label can be incorrect. Samples with wrong labels in the training data set may mislead the training progress and decrease the model performance. Differing from the nominal classification task, the label of fabric smoothness appearance assessment task is ordinal. In this study, we discuss the issue with the consideration of the ordinal label attribute of the task, and propose a training strategy to eliminate the mislabeled samples in training.

In this paper, to further improve the performance of 2D-image-based objective fabric smoothness assessment methods, an image classification framework with a multi-level feature fusion module and a mislabeled sample filtering module is proposed. The feature fusion module fuses a set of hand-crafted features extracted by the multi-scale spatial masking model and a set of CNN features extracted by a pre-trained CNN model. The hand-crafted features have richer local information representing the bottom-up perception of HVS. It complements the abstract top-down information from the CNN features to describe the fabric smoothness in a comprehensive way. In addition, the mislabeled sample filtering module trained with the hand-crafted features help avoid the negative impact of the mislabeled samples in training.

The main contribution of the study can be summarized as follows:

1. The multi-level feature fusion module in the proposed image classification framework can take the advantages of both hand-crafted and CNN features when evaluating the fabric smoothness appearance. It is the first study in this task that explore the idea that combining features at different levels.

2. We propose a mislabeled sample filtering module in the framework, which helps mitigate the impact of mislabeled samples in training for image classification tasks with subjectively labeled and sequence tags.

3. The experimental results of the proposed framework in the fabric smoothness assessment task are superior to existing methods, and the proposed framework and modules can be easily extended to other tasks.

II. RELATED WORKS

A. IMAGE ACQUISITION

According to the different data acquisition technologies used, the computer-vision-based methods proposed in the existing research can be divided into two categories, namely 2D methods and 3D methods. The 2D methods [5]–[12] usually use an industrial camera or a scanner to collect the fabric surface data, whose data format is a gray-level 2D digital image. The 2D image expresses the fabric smoothness information as the shadow caused by wrinkles on the sample surface, which is an indirect expression progress with a high environmental dependence. In addition, if the fabric sample has color texture, the color texture will be mixed with the smoothness information in the image space and difficult to be distinguished in the subsequent feature extraction process. This limits the adaptability of the 2D method in the objective assessment of fabric smoothness appearance.

Hence, more researchers have recently tended to use 3D methods. Technologies such as laser triangulation [13]–[17], photometric stereo method [18], [19] and binocular stereo-vision [20]–[23] are used in existing studies, to obtain the smoothness information of the fabric as the reconstructed surface depth map. Compared with the 2D methods, 3D methods are kinds of direct expression of the sample's surface, which can effectively avoid the disadvantages of the 2D method. However, the 3D methods have the disadvantages of low efficiency, high cost, high calibration complexity, and unreliable accuracy.

In our previous studies [24], [29], [31], we put forward the following point: although the observation of the 2D method is indirect, they can still provide sufficient information for the smoothness evaluation of the fabric. we have established a 2D fabric image acquisition system and optimized the illumination environment for fabric smoothness information extraction [24], and the view was verified by the experiments in the studies. Additionally, to improve the multi-color fabric adaptability of the 2D method, we have proposed a decoloration method for multi-color fabric images in the previous study [25]. In this study, we further explore the improvement of the 2D method, and the fabric image data set is the same as our previous studies captured by the proposed system.

B. HAND-CRAFTED FEATURES

Researchers in the field have proposed different features to express the fabric smoothness. Depending on the data format, the features are listed below.

(1) Features from 2D images:
Edge area; shade area [5]; variation of the grey level intensity [6]; angular second moment, contrast, correlation, entropy of the gray level co-occurrence matrix (GLCM) [7], [8]; gray-scale range around the edges [9], [10]; FFT subtotal in a specific frequency range [11]; orientation, hardness, density, and contrast of wavelet coefficients [12].

(2) Features from 3D depth maps:
Length, surface area, volume under the surface, mean principle curvatures, mean max twist of every sub-block of the depth map [13]; arithmetic average roughness, root mean square roughness, 10-point height, bearing surface ratio, wrinkle sharpness, wrinkle density [14], [22]; fractal dimension [20]; maximum amplitude, sharpness, density, maximum amplitude of the first derivative of the cross profile of the edges [15], [21]; mean, mean deviation, and standard deviation of the height values in every row [26]; wrinkling density, wrinkling hardness, tip-angle, wrinkling roughness [27]; dense-SIFT feature with sparse coding [4].

In our previous study, we extract the image features by a visual masking model [29] for human visual system. The experiment in the study proved that this feature has a strong ability to characterize the smoothness appearance of the 2D fabric images.

C. CNN FEATURES

In recent year, CNN models have been widely used in different tasks in computer vision and reached the state-of-the-art performance [30]. During the training progress, the CNN models can extract the features from the input images with high-level and abstract information. However, the training of CNN heavily depends on the amount of training data to avoid overfitting. In our recent study [31], we have built a compact CNN model for the fabric smoothness assessment task with small sample size, and demonstrated that the CNN features can express the human perceptual prediction of the tendency of wrinkles to recover to some extent.

Another way to mitigate the overfitting of CNN training on small size data sets is using the transfer learning [40], [41] with a pretrained CNN model. However, in our experiment, the pre-trained deep CNNs did not show superior performance. The reason may be that the CNN features extracted by the pre-trained deep CNNs lose too much low-level information. In this paper, we will try to improve the performance of the pre-trained CNNs with the fusion of hand-crafted features.

D. FUSION OF HAND-CRAFTED FEATURES AND CNN FEATURES

Due to different feature construction methods, hand-crafted features and CNN features have some differences in describing the problem. To take advantages of both, researchers has paid attention to the fusion of them. For image classification, the studies can be categorized into feature level fusion, fusion by ensemble and joint level fusion.

For feature level fusion, the hand-crafted features and CNN features are directly combined by concatenation, then reduced and utilized to train the classifier [32]–[34]. In [38], for lung nodule classification, a set of hand-crafted features and the CNN feature are fused by the cascade method. The cascaded feature is then processed by the principal component analysis (PCA) method and used to train a cost-sensitive random forest model. In [37], for osteoporosis diagnoses, researchers propose encoded features combined with the CNN features. After the feature selection by a minimum-redundancy maximum-relevance (mRMR) algorithm, the fusion features are used to train an SVM classifier. However, direct concatenation does not consider the different dimensions and scales of the features.

For fusion by ensemble, multiple classifiers are trained with the hand-crafted features and CNN features and then combined into an ensemble [35]. Some studies in this category also take the feature level fusion into consideration by using an additional classifier trained with the fusion features. For examples, in a study for intelligent human action recognition [42], the histogram of oriented gradients (HoG) is utilized as the hand-crafted feature and fused with the CNN feature by a feature level fusion method with feature reduction selection. And the multi-class support vector machine (SVM) is used as the classifier. Classifiers map the features into the problem solution space, which help avoid the disadvantages of the feature level fusion. However, the training process of multiple classifiers is time-consuming. And it is difficult to train a traditional classification model with the large-dimensional CNN features.

Joint level fusion frameworks combine the hand-crafted feature and the CNN feature in the constructed module in the whole CNN framework, and output the prediction result comprehensively [36]. The classification is then given by a softmax layer. In [39], for fabric defect detection, the researchers proposes a defect probability map as the hand-crafted features, which is fused in the convolutional level of the CNN. This is the only example of the hand-crafted and CNN feature fusion work in textile research at present.

In summary, the joint level fusion can achieve the feature fusion in one comprehensive CNN model, avoiding multi-step model training. In addition, joint level fusion can fuse the features in different scales without normalization, making better global optimization. Therefore, in this paper, the joint level fusion is adopted. And the other two categories of feature fusion frameworks are discussed in the experiment.

E. SMOOTHNESS DEGREE CLASSIFICATION

To the aspect of smoothness appearance degree classification, minimum distance algorithm [11], fuzzy priority similarity comparison method [8], neural network [7], [14], logistic regression [27], and support vector machine (SVM) [4], [12] have been widely used. These models are generally trained with hand-crafted features of the input images, and perform differently in different feature domains. Differing from the traditional classification models, the CNN models can be directly trained with the raw image data and output the result [30]. In this paper, a CNN classifier is established which is trained with CNN features from transfer learning and hand-crafted feature from spatial masking model.

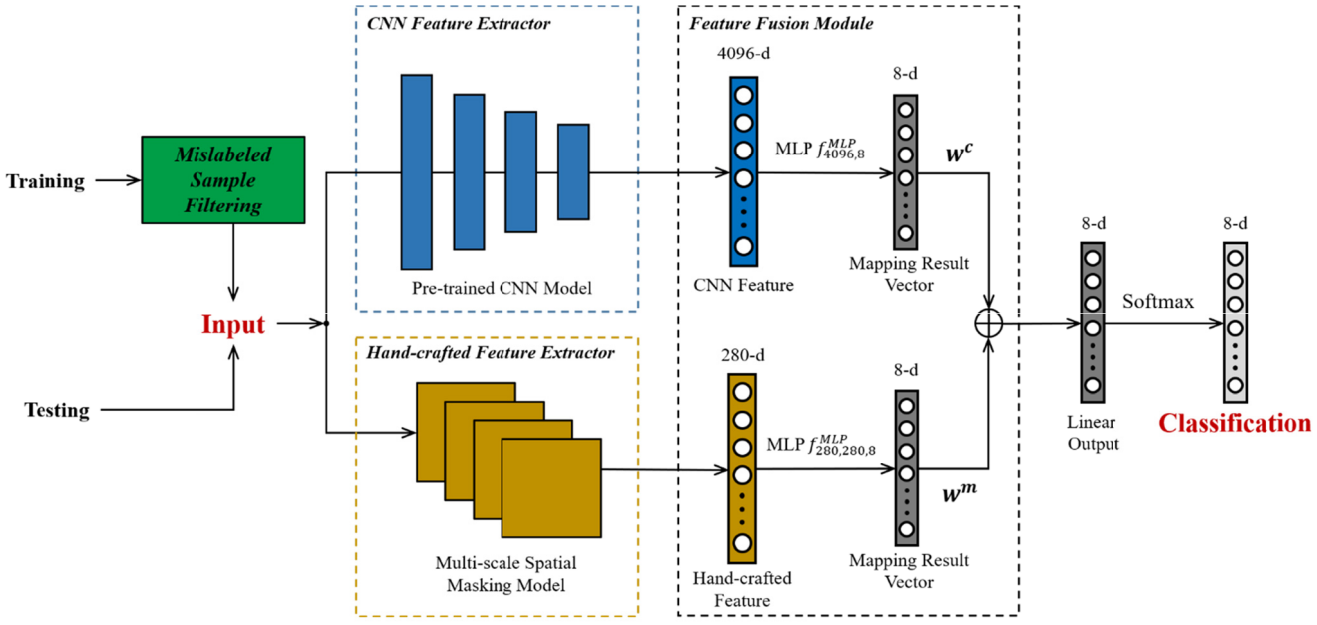

FIGURE 2. A flowchart of the proposed image classification framework with multi-level feature fusion and mislabeled sample filtering; where $f_l^{MLP}(\cdot)$ represents the multilayer perceptron (MLP) whose integer subscript sequence indicates the feature dimension of each layer; $w^m \in \mathbb{R}^8$ and $w^c \in \mathbb{R}^8$ are the weight vectors for every connection.

F. OBJECTIVE METHODS FOR FABRIC WRINKLE EVALUATION

In addition to the fabric smoothness appearance evaluation method reviewed above, there are many other objective methods for the evaluation of the fabric wrinkle property. By the measuring object, the methods can be divided into direct methods and indirect method. The direct methods measure the wrinkle on the sample surface while the indirect methods measure the other indicators of the fabric and mapping it to wrinkle properties.

For direct methods, in [6], a wrinkling generation device is established. Differing from the standard sample preparation method, it generates wrinkling on fabric by a motor driven cylinder creux automatically. In addition, instead of testing the fabric smoothness appearance on fabric sample, researchers propose a series of wearing simulation devices to generate samples with similar wrinkles in actual wearing [43], [44]. Then the smoothness appearance of the samples is also evaluated by image analysis algorithms. In another study [45], the proposed method can evaluate the smoothness appearance for real garment. In addition, image-based wrinkle recovery angle are also used to evaluate the wrinkle properties of fabrics [46], [47]. Differing from the random wrinkle generation method (standard laundering), it uses a quantitative wrinkle sample preparation method, and measures the wrinkle recovery angle to evaluate the wrinkle property of fabrics.

In addition, there is a series of methods that does not measure the wrinkles on the sample surface, namely indirect measurement. These methods evaluate the fabric wrinkle property by measuring the fabric weave parameters, mechanical properties [48], in-situ characterization [49], and shape retention properties [50] of the fabric sample and mapping to the smoothness appearance degree or wrinkle recovery angle.

The fabric smoothness appearance evaluation studied in this article is parallel to the above methods. They are all used to evaluate the wrinkle property of fabrics. But at present they cannot replace each other. Moreover, at present, only two methods, i.e. the evaluation of the wrinkle recovery and fabric smoothness appearance after laundering, are included in the relevant standards, which are commonly used in the industry.

III. PROPOSED IMAGE CLASSIFICATION FRAMEWORK

In this paper, an image classification framework is established to solve the fabric smoothness assessment task. As given in Figure 2, the framework has a multi-level feature fusion module which fuses the hand-crafted and CNN features, and contains a mislabeled sample filtering module to reduce the participation of mislabeled samples in training.

A. HAND-CRAFTED FEATURES

The smoothness appearance of fabrics is a subjective concept, which is mainly concerned by the perception of the human vision under a specific environment. A feature set named multi-scale spatial masking feature (MS-SMF) [29] is proposed with the consideration of the relationship between the HVS characteristics and the fabric smoothness appearance in images. As it was verified that the MS-SMF has a superior ability to describe the fabric smoothness appearance than other hand-crafted features used in previous studies, the

MS-SMF is used as the hand-crafted features used in the proposed image classification framework.

Let I denotes an input fabric image, the MS-SMF extraction can be expressed as

$$v^m = f^M(I), \tag{1}$$

where $v^m \in \mathbb{R}^d$ is the MS-SMF of dimension d; $f^M(\cdot)$ represents the MS-SMF extraction. The operation of $f^M(\cdot)$ includes three main steps, i.e. difference-of-Gaussian (DoG) scale space decomposition, spatial masking effect maps calculation, and feature extraction in different scales. They are detailed as follows.

1) DOG SCALE SPACE DECOMPOSITION

In MS-SMF extraction, the input image I is decomposed into the DoG scale space to model the multi-scale perception ability of the human testers. the Gaussian scale space can be generated from a series of low-pass filtering and down-sampling, and expressed as a sequence of images $D = \{I_0, I_1, \ldots, I_{N_g}\}$ formulated as follow [51], [52]:

$$I_n = G_{q^n\sigma} * I_{n-1}, \tag{2}$$

where $n = 1, 2, \ldots, N_g$ is the sequence number; $I_0 = I$; $G_{q^n\sigma}$ is a Gaussian kernel with the scale $q^n\sigma$ (variation of Gaussian); σ is a constant basic scale factor; $*$ is the convolution operator; q is a constant multiplicative factor which determines the interval of the decomposition; N_g is the scale parameter which determines the number of images in the decomposition result. The DoG scale space can be expressed as a sequence of images $D^- = \{I_0^-, I_1^-, I_2^-, \ldots, I_{N_g-1}^-\}$ calculated from the difference between every two consecutive images in the Gaussian scale space as follow [53]:

$$I_n^- = I_n - I_{n+1}, \tag{3}$$

where $n = 0, 1, 2, \ldots, N_g - 1$.

2) SPATIAL MASKING MAP CALCULATION

In the HVS, spatial masking effect refers to the phenomenon that a visible stimulus (target) turns to undetectable due to the presence of another (masking) [54]. In fabric smoothness appearance perception, a more wrinkled fabric surface can causes a stronger spatial masking effect [29]. For every scale $n \in \{0, 1, 2, \ldots, N_g\}$, the spatial masking map M_s^n calculation is detailed in follows. For the operation is the same to every scale, the sequential variable n is omitted.

Step 1: Pattern Complexity Calculation

The pattern complexity of a pixel $x^p \in X^p$ is described as the gradient direction irregularity in the local region $R(x^p)$ centered on x^p. For a scale n, with the scale images I_n, I_{n+1} and I_n^-, the orientation θ (the subscript n is omitted) for each pixel $x^p \in X^p$ can be formulated as follow:

$$\theta\left(x^p\right) = \frac{1}{2}\left(arctan\frac{G_v^0\left(x^p\right)}{G_h^0\left(x^p\right)} + arctan\frac{G_v^1\left(x^p\right)}{G_h^1\left(x^p\right)}\right), \tag{4}$$

where $G_v^0(x^p)$ and $G_h^0(x^p)$ are the vertical horizontal gradient of the image I_n at pixel x^p; x^p is the coordinate space of I_n; and $G_v^1(x^p)$ and $G_h^1(x^p)$ are the vertical horizontal gradient of the image calculated from I_{n+1}.

The pattern complexity C_p for a pixel $x^p \in X^p$ is calculated as the sparsity of the orientation histogram for the local region $R(x^p)$ centered on x^p:

$$C_p\left(x^p\right) = \left\|H\left(t \mid x^p, T\right)\right\|_0, \tag{5}$$

where $H(t \mid x^p, T)$ is the histogram of θ in the local region $R(x^p)$; T is the histogram interval, which is selected as 12° according to the subjective masking experiments [55]; t is the coordinate value of the histogram; $\|\cdot\|_0$ denotes the L_0 norm computing on t for the pixel x^p, which counts the number of non-zero elements in H.

Step 2: Luminance Contract Calculation

The luminance contrast C_l can be expressed by the local gradient strength of a Gaussian image, which is calculated as the magnitude of the DoG image I_n^-, that is:

$$C_l\left(x^p\right) = \left|I_n^-(x^p)\right|, \tag{6}$$

where $x^p \in X^p$ represents the image coordinates.

Step 3: Pattern Masking Effect Calculation

As an important component in spatial masking effect, the pattern masking effect is firstly computed. Based on the subjective experiment about the visual gain control and pattern masking, the pattern masking effect M_p is modeled as [55], [56]

$$M_p = f_l(C_l) \cdot f_p\left(C_p\right) = \log_2(1 + C_l) \cdot \frac{a_1 C_p^{a_2}}{C_p^2 + a_3^2}, \tag{7}$$

where a_1 is a constant of proportion indicating the output gain of the visual stimuli in human visual system; a_2 is an exponential parameter representing the excitatory of human visual neurons; and a_3 is a small constant to avoid the zero denominator, which helps describe the strength of pattern masking effect when the pattern complexity is zero. In the subjective test [55], [56], pattern masking strengths (expressed as the contrast gain of human testers) caused by different visual input pattern are recorded. The parameters in the model are then obtained by the fitting method as $a_1 = 0.8$, $a_2 = 2.7$, and $a_3 = 0.1$.

Step 4: Contract Masking Effect Calculation

As another important component in spatial masking effect, the constrast masking effect M_c in the MS-SMF is modeled. In perceptual studies, the spatial masking caused by the luminance contrast should be expressed by a nonlinear transducer for luminance contrast [57], [58]:

$$M_c = 0.115 \cdot \frac{\alpha \cdot C_l^{2.4}}{C_l^2 + \beta^2}, \tag{8}$$

where α is a constant of proportion; and β is the positively accelerating and compressive regions of the nonlinearity. In subjective experiments, the luminance masking effect strength is evaluated by the just noticeable difference (JND)

in images perceived by human testers. By fitting the model with the JND in different luminance contrasts, the parameters are set as $\alpha = 16$ and $\beta = 26$.

Step 5: spatial masking effect formulation

The spatial masking model in MS-SMF takes both the contrast masking and pattern masking into account. For the stronger one in the two types of masking effects plays a dominant role in some in general, the combination of contrast masking and pattern masking is calculated by a maximum function to express spatial masking effect M_s.

$$M_s\left(\boldsymbol{x}^{\boldsymbol{p}}\right) = max\left\{M_p\left(x^{\boldsymbol{p}}\right), M_c\left(x^{\boldsymbol{p}}\right)\right\}. \tag{9}$$

As the sequential variable $n \in \{0, 1, 2, \ldots, N_g - \}1$ is omitted in the above model expressions, the M_s should have a Sequence superscript n. Thus M_s^n is the spatial masking map for the fabric image in n^{th} scale.

3) MULTI-SCALE SPATIAL MASKING FEATURE

As results, the sequence $\{M_s^0, M_s^1, \ldots, M_s^{N-1}\}$ can describe the spatial masking effect in different scales. Therefore, the feature vector v^m is express as follow:

$$v^m = \left\{pr\left(M_s^n, i_p\right)\right\}, \tag{10}$$

where $n \in \{0, 1, \ldots, N_g - 1\}$; $pr\,(M, p)$ is the percentile calculation function which returns the p^{th} percentile of the data set M; and $i_p \in \{60, 70, 80, 90\}$ is the percentiles used in feature extraction. In the experiment in this study, for the scale length N_g is set as 35, the feature dimension is 140 for one fabric image. As we use two images captured under orthogonal position angles of the light source for one fabric sample, the dimension of a fabric sample is 280.

B. CNN FEATURES

In the recent research on image classification, the deep CNN models have shown excellent performance. It is believed that the deeper CNN model may extract more abstract image features, which have better capabilities to describe the abstract contents in images than the traditional hand-crafted features. The fabric smoothness appearance is a subject concept abstracted by the human vision, which includes the prediction of the possibility of fabric wrinkle recovery and some aesthetic perception. The CNN models may be able to extract such abstract features. Therefore, CNN features are used in the proposed image classification framework as a complement to the hand-crafted features.

Let I denotes an input fabric image, consider a CNN classification model $\Psi\,(I; \Theta) = g(f^c\,(I; \Theta^c)\,; \Theta^g)$, the CNN features extraction can be expressed as

$$v^c = f^c\left(I; \Theta^c\right), \tag{11}$$

where $f^c(\cdot)$ is the calculation before the last fully connected layer of the CNN $\Psi\,(I; \Theta)$ with parameters Θ^c; $v^c \in \mathbb{R}^q$ represents the CNN features of dimension q; Θ is the parameters of the CNN model; and $g(\cdot)$ is the last activation and output layer of the CNN with the parameters Θ^g.

FIGURE 3. CNN model structure used in the experiment, where Conv represents convolutional layers; ReLU represents the rectified linear unit activation function; Max Pool represents max pooling layers; FC represents fully-connected layers.

In the proposed image classification framework, different CNN classification models that is the state-of-the-art can be used. To satisfy the small data size of the fabric smoothness appearance evaluation task, the models are pre-trained with the Imagenet [59] data. Take the AlexNet [60] as a well performed example, the model $\Psi\,(I; \Theta)$ is trained with the Imagenet and the dimension q of the features v^c is 4096. As given in Figure 3, the AlexNet model is constructed by five convolutional layers, three max pooling layers, and three fully-connected layers. The input image is resized into 224 × 224 pixel. For the convolutional layers (Conv), the parameters provided in Figure 3 includes the kernel size, step size, channel number, and activation function. For fully-connected layers (FC), the parameters include neuron number and activation function. In addition to the application example of AlexNet, the performance of features extracted by different CNN models were tested in our experiment.

C. SAMPLE FILTERING

As the label of the samples in fabric smoothness appearance assessment are evaluated by human testers, there must be noises in the labels. Intuitively, in the fabric smoothness assessment task, the wrong label is more likely to be close to the correct label, that is, SA-1 sample is more likely to be wrongly rated as SA-2 instead of SA-5. Based on this characteristic of the problem, in this paper, we propose to use a nominal classification model to be the label-error estimator, and filter the samples with large error.

For a training set $\mathcal{S} = (\boldsymbol{x}_i, \tilde{y}_i = y_i + \Delta y_i)_{i=1}^N$ with label noise Δy. The trained classification model r can be viewed as a ground truth estimator. And the estimated output can be used to estimate the noise for the samples in the testing set $\mathcal{T} = (\boldsymbol{x}_i, \tilde{y}_i = y_i + \Delta y_i)_{i=1}^M$ as

$$\Delta\hat{y} = \left|r\left(\boldsymbol{x}_i\right) - \tilde{y}_i\right|, \tag{12}$$

where $i = 1, 2, 3, \ldots, M$; r is the classification model trained with S. In the experiment in this paper, $\boldsymbol{x}$ is the hand-crafted feature v^m and the samples with $\Delta\hat{y}$ greater than 0.5 degree

are filtered out in the framework. A multi-class SVM model is used as the classification model.

D. FEATURE FUSION AND CLASSIFICATION

Figure 2 shows the structure of the proposed image classification framework. The framework is constructed by three main modules, i.e. hand-crafted feature extractor f^m, CNN feature extractor f^c, and feature fusion module f^s. As an input image I, the hand-crafted features $v^m \in \mathbb{R}^d$ and the CNN features $v^c \in \mathbb{R}^q$ are extracted according to Equation (1) and (11) respectively. The hand-crafted features describe more information about the intensity of visual stimuli caused by the wrinkles. And the CNN features contain more abstract description about the aesthetic perception and the wrinkle recovery prediction of human vision. In the expectation of the development of the feature fusion module, these features may constitute a complementary structure.

As these two categories of features are not distributed on the same scale, it is unreasonable to directly concatenate them and use them for classification, which may cause the unbalanced contributions of them. Therefore, in the feature fusion module, the features are mapped to the solution space of K dimension by a multilayer perceptron (MLP), respectively, where K is the number of categories of the task target. The pre-mapping of the features into the same space can help to fuse them in a same scale.

For the CNN features v^c, due to its strong information abstraction ability, we use a single layer MLP to map it to the problem solution space. That is

$$z^c = f_{q,K}^{MLP}\left(v^c, \Theta^{m_1}\right), \tag{13}$$

where $z^c \in \mathbb{R}^K$ is the mapping result vector in the problem solution space; $f_{q,K}^{MLP} : \mathbb{R}^q \to \mathbb{R}^K$ is the function of the single layer MLP with q-dimensional input and K-dimensional output; Θ^{m_1} is the parameters of the MLP $f_{q,K}^{MLP}$; and the activation function in MLP is the ReLU function.

While for the hand-crafted features v^m, considering the low dimension and abstraction of it, we use a two-layer MLP to map it, that is

$$z^m = f_{p,p,K}^{MLP}\left(v^m, \Theta^m\right), \tag{14}$$

where $z^m \in \mathbb{R}^K$ is the mapping result vector in the problem solution space; $f_{p,p,K}^{MLP} : \mathbb{R}^p \to \mathbb{R}^K$ is the function of the two-layer MLP with p-dimensional input and K-dimensional output, and p-dimensional middle layer; Θ^{m_2} is the parameters of the MLP $f_{p,p,K}^{MLP}$; and the activation function in MLP is the ReLU function.

Finally, as the hand-crafted and CNN features are both mapped into the problem solution space, the outputs z^m and z^c could contain a certain correlation between the prediction results. As a result, we fused them by an element-weighted linear layer. As shown in figure 2, in the element-weighted linear layer, each element in the output of this layer is linearly contributed by the weighted corresponding element with the same index in the two inputs, which can be formulated as follow:

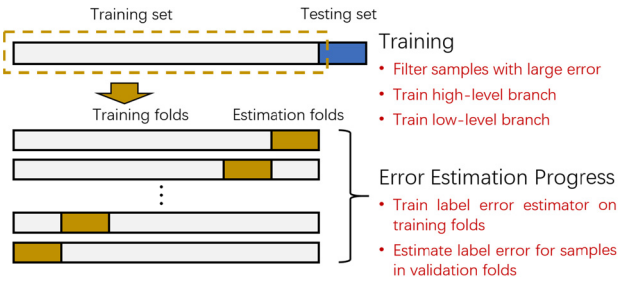

FIGURE 4. Training process illustration of the proposed image classification framework with mislabeled sample filtering.

$$p^{out} = f^s\left(z^{out}\right), \tag{15}$$

where p^{out} is classification probability activated by a SoftMax function $f^s(\cdot)$, and

$$z_i^{out} = w_i^m z_i^m + w_i^c z_i^c, \tag{16}$$

where $i = 1, 2, 3, \ldots K$ is the vector index in the vectors; $w^m \in \mathbb{R}^K$ and $w^c \in \mathbb{R}^K$ are the weights for every connection; $z^{out} \in \mathbb{R}^K$ is the linear output of the element weighted layer. In our experiment, the dimension p of the hand-crafted features v^m is 280; the dimension q of the CNN features v^m is defined by the CNN model in use (4096 for AlexNet); and the dimension K of problem solution space is 8 for our fabric smoothness dataset.

E. MODEL TRAINING

An illustration of the training progress of the proposed framework is shown in Figure 4. The framework firstly conducts an error estimation progress. In this progress, the training set is split into m folds, then every fold is treated as the estimation fold in turn, while the others are used as training fold. The estimation model is trained on the training folds and predicts the label-errors of the estimation folds in turn.

Based on the estimation of the label-errors of all the samples in the training set, the framework filters the samples with large error. In the experiment, the samples with estimated label-error greater than 0.5 degree are filtered. Then, on the training set after filtering, the framework trains the hand-crafted and CNN modules in an alternating way.

Although the proposed label-error estimator may have some mistakes, in our experiment, the filter can mitigate the influence of wrong label samples in training to some extents. In addition, since the pre-trained CNN and the hand-crafted feature extractors have certain unsupervised properties, the framework can be trained on small datasets properly.

IV. EXPERIMENT

A. IMAGE ACQUISITION AND PREPROCESSING

As introduced in our previous study [29], an image acquisition system was established. It has ability to capture images

TABLE 1. Fabric samples information and the sample number in different smoothness appearances.

Fabric			Sample Size							
Sample number	Weave	Fiber content	SA-1	SA-1.5	SA-2	SA-2.5	SA-3	SA-3.5	SA-4	SA-5
1	Plain	100% C	25	13	15	15	7	7	10	6
2	Plain	65% T 35% C	2	1	3	4	7	25	25	29
3	Twill	100% C	18	12	13	15	9	11	12	4
4	Twill	65% T 35% C	3	2	1	4	6	24	23	34
Total			48	28	32	38	29	67	70	73

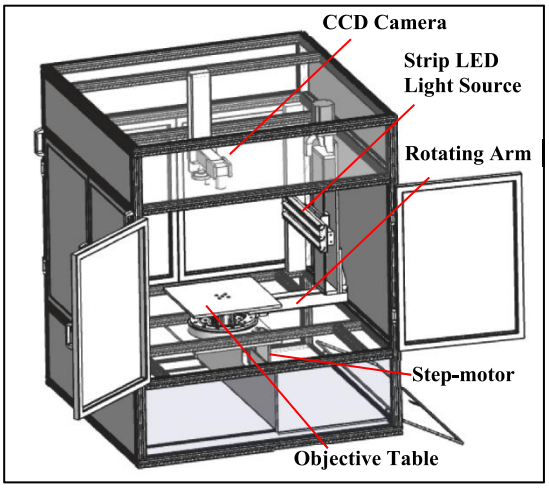

FIGURE 5. The fabric image acquisition system.

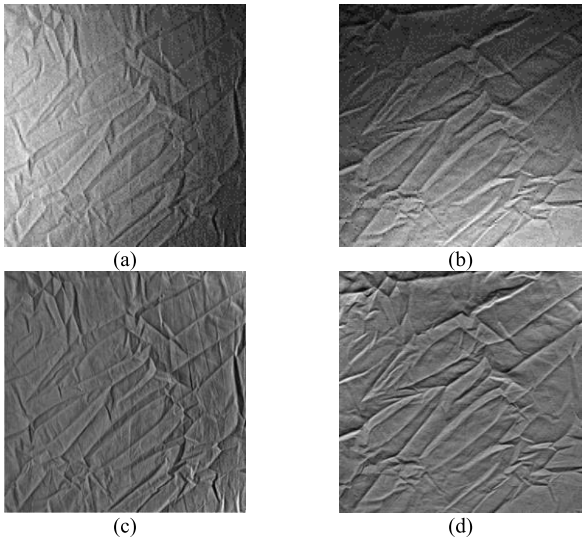

FIGURE 6. Fabric images with different light source position angles: (a) 0°, (b) 90° and their preprocessing results: (c), (d).

of the fabric samples under different image acquisition environments (as illustrated in Figure 5). And the image acquisition environment was optimized. It was demonstrated that two images with two orthogonal light source position angles contains sufficient fabric smoothness appearance information for a fabric sample. Image data of a fabric sample is illustrated in the first row of Figure 6.

For there is general brightness unevenness in fabric images captured by the system, we have proposed an image preprocessing algorithm to adjust it. Denote the original fabric image is I_o, the preprocessing can be expressed as follow:

$$I = I_o / I_f, \tag{17}$$

where I_f is the two-dimensional binomial fitting of I_o. As given in the second row of Figure 6, the unevenness has been eliminated. The effectiveness of the preprocessing has been verified in our previous study [31].

B. EXPERIMENT SETUP

The experiments in this study were conducted on a personal computer with Intel(R) Core(TM) I7-4790 CPU(3.6 GHz) and 16GB RAM and GPU of Nvidia(R) GTX 1080Ti. All the algorithms were implemented by MATLAB and Python under the Windows 10 and Linux operating system. The SVM was implemented by the LIBSVM [61]. The deep learning algorithms were implemented by the PyTorch [62].

C. MATERIALS

The fabric image data set used in this study is the same as our previous studies [24], [29], [31] within 385 fabric samples, including four varieties of fabrics in different weave structures and fiber compositions. According to the AATCC standard [2], the fabric samples were cut into 380 mm × 380 mm, then laundered in different modes to generate diverse fabric smoothness appearances. The smoothness degree of the samples was evaluated by the subjective method. The detail information of the fabric samples is given in table 1. In our experiment, the characteristics of white fabrics reveal quite negligible effect in the fabric images.

D. DATA AUGMENTATION

For the performance of CNN models are heavily relies on the data set size, we used two data augmentation methods. Firstly, the illumination brightness was extended into three levels in image acquisition, which expanded the data set size three times. Secondly, the fabric images (with brightness extended) in data set were noised, rotated, and flipped randomly

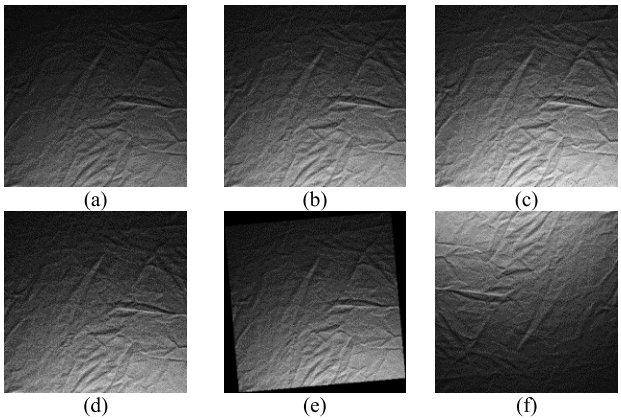

FIGURE 7. Fabric image data augmentation examples: illumination augmentation (a), (b), (c); and noising (d), rotating (e), flipping (f) on (a).

before training, which expanded the data set size four times. Thus, the final image data set size was expanded by twelve times to 4620. A data augmentation example is given in Figure 7.

E. TRAINING DETAIL

In the experiments in this research, we applied mini-batch Adam optimizer [63] to train the CNN models. The batch size was128 and the epoch amount was 100. We set the learning rate decayed by 10 by every 25 epochs. The best learning rate and weight decay are optimized by grid search in $[3\times10^{-4}, 3\times10^{-3}, 3\times10^{-2}]$. The performance of the CNN models was evaluated by 10-fold cross validation. In this process, the data set was divided into 10 folds. And then they were used as testing folds and the reset as training folds in loop.

The training folds were augmented before model training. Thus, the training sample size was 4104 and the testing sample size was 39 in average for each loop. Due to the small sample size, it is difficult to control the sampling bias, which may cause the test results difficult to evaluate the generalization of the model. Therefore, we did not set up a separate test set.

F. EVALUATION INDICES

The target of the fabric smoothness appearance assessment task is the different smoothness degrees in sequence. When evaluate the performance of the assessment methods, the closer the prediction is to the label means the better the model performs. As a result, in our study, classification accuracies under errors of 0 degree, 0.5 degree, and 1 degree are calculated, namely acc0, acc0.5, and acc1 respectively. The classification accuracy under a specific error ε can be formulized as the percentage of samples whose prediction error is not larger than ε. Therefore, for a testing set $\mathcal{T} = (\boldsymbol{x}_i, y_i)_{i=1}^{M}$, the acc0, acc0.5, and acc1 for a predictor r can be calculated as follow:

$$acc0 = \frac{1}{M}\sum_{i=1}^{M} h\left(\left|r\left(\boldsymbol{x_i}\right) - y_i\right|, 0\right), \tag{18}$$

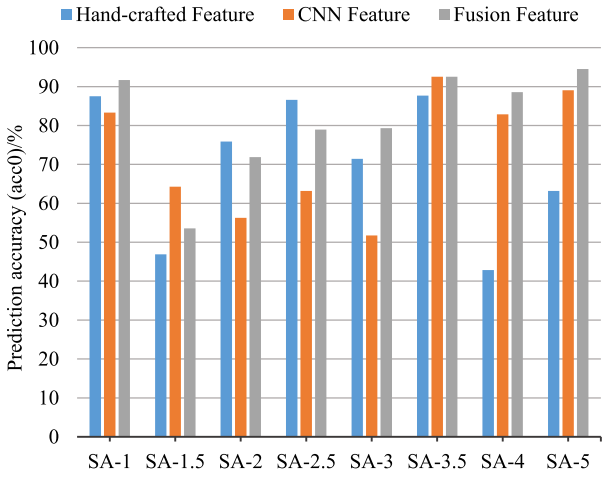

FIGURE 8. The testing accuracy of the classifiers trained with hand-crafted features, CNN features, and the proposed fusion features.

$$acc0.5 = \frac{1}{M}\sum_{i=1}^{M} h\left(\left|r\left(\boldsymbol{x_i}\right) - y_i\right|, 0.5\right), \tag{19}$$

$$acc1 = \frac{1}{M}\sum_{i=1}^{M} h\left(\left|r\left(\boldsymbol{x_i}\right) - y_i\right|, 1\right), \tag{20}$$

where $\hbar(\cdot)$ is a discriminant function as

$$\hbar\left(\varepsilon_0, \varepsilon\right) = \begin{cases} 1 & \varepsilon_0 \le \varepsilon \\ 0 & \varepsilon_0 > \varepsilon \end{cases} \tag{21}$$

Compared with the commonly used prediction accuracy, the accuracies under different errors can better evaluate the performance of the methods on such problems with different error tolerances.

V. RESULT AND DISCUSSION

A. VERIFICATION OF COMPLEMENTARITY BETWEEN FEATURES

In this study, the proposed image classification framework fuses the hand-crafted features and the CNN features. To explain the motivation and demonstrate the effectiveness of the fusion features, in this experiment, the hand-crafted features, the CNN features, and the fusion features were used to train the proposed image classification framework respectively. The testing accuracies (acc0) on different smoothness appearance degrees of them are given in Figure 8. The model trained with the hand-crafted features resulted higher accuracies on the samples in SA-1, SA-2, SA-2.5, and SA-3, while the CNN feature showed better adaptability on the samples in SA-1.5, SA-3.5, SA-4, and SA-5. This experimental phenomenon shows that these two types of features have a certain difference in the ability to describe the fabric smoothness appearance. It is an inspiration for the research work in this article.

On the other hand, by combining these two types of features, the fusion features showed the complementary use of the above two features in the experiment. It can be observed from Figure 8 that, in most smoothness appearance degrees,

TABLE 2. The comparison results of different methods.

Method	acc0 (%)	acc0.5 (%)	acc1 (%)	Parameters	Operation time (s)
Edge Features+kNN [9, 10]	76.9	95.6	99.74	-	0.1178
GLCM+BP [7]	68.9	92.2	99.74	-	0.0121
GLCM+SVM [8]	77.4	93.3	100	-	0.0136
Wavelet+SVM [12]	72.2	90.9	99.48	-	0.0243
Fourier [11]+SVM	79.2	94.6	100	-	0.0269
MS-SMF +SVM [29]	82.6	95.8	100	-	1.0916
Shape features + SVM [27]	79.5	95.1	100	-	0.2670
ScSPM+SVM [4]	69.9	90.1	99.74	-	2.1051
AlexNet [60]	76.3±1.30	91.1±0.76	99.74	6.1×10^7	0.0459
ResNet18 [64]	79.6±2.31	92.9±1.98	100	1.2×10^7	0.1180
ResNet34 [64]	79.9±2.40	92.1±2.30	100	2.2×10^7	0.2187
Inception v3 [65]	80.0±1.64	90.9±1.78	99.56	2.7×10^7	0.3569
VGGNet [66]	77.8±1.98	92.4±2.41	100	1.4×10^8	0.7738
cCNN [31]	84.0±1.00	95.4±1.48	100	4.3×10^5	0.0312
The proposed framework	**85.2±0.92**	**96.1±0.37**	100	6.1×10^7	1.1350

the testing accuracy of model trained with the fusion features was higher than the other two, or close to the better of the other two. The only exception is the results in SA-1.5, where the proposed feature fusion framework only reached the middle value (53.6%) of the others (46.9% and 64.3%). In fact, the general performance of all the models on SA-1.5 was undesirable. The reason may be that the sample size at this level is smaller than others (as can be seen in Table 1), resulting in insufficient model training.

B. COMPARISON RESULTS

In the 2D-image-based objective fabric smoothness appearance assessment field, different methods have been proposed with different image features and classification models. In this experiment, we compared the proposed framework with the existing different methods on the same data set. On the other hand, in the studies of 3D methods, researchers have adopted more advanced techniques in feature extraction. For example, Ouyang [27] defined a set of shape features (wrinkling density, wrinkling hardness, tip-angle, wrinkling roughness), and Xu *et al.* [4] used the dense-sift feature encoded by linear spatial pyramid matching using sparse coding(ScSPM). These features were transferred into 2D fabric images and tested in this experiment.

In addition, for the CNN models have been widely used in different image classification tasks, we discussed the performance of a series of CNN models on fabric smoothness appearance assessment, i.e. AlexNet [60], GoogLeNet: Inception V3 [65], ResNet [64], and VGGNet [66]. The main purpose of this experiment is to verify the effectiveness and superiority of the proposed method. Because the results of each training of the CNN models have a certain randomness, the CNN model was trained and tested three times, and the results were recorded as the average value and standard deviation of the accuracies, i.e. acc0, acc0.5, and acc1. The average accuracies reveal the general performance of the models on the data set, and the standard deviations of the accuracies describe their training stability.

The comparison results are given in Table 2. In comparison, the proposed framework achieves the best average values of acc0 and acc0.5, which are 85.2% and 96.1%, respectively. Compared with methods in fabric smoothness evaluation, the proposed framework further improves the evaluation performance. Compared with the general CNN methods for image classification, the proposed framework shows better adaptability in this particular task. In addition, the image samples for the standard replicas are predicted correctly by the proposed framework, which means the trained framework learned an effective distribution. Generally, the comparison experiment demonstrates the superior effectiveness to fabric smoothness evaluation of the proposed framework.

The standard variation (std) of the classification accuracies in Table 2 shows the training stability of the CNN models. As can be seen, the proposed method has the lowest acc0-std and acc0.5-std compared with the other CNN models. These results mean that the proposed method has the best stability in different CNN models. A possible reason is that the hand-crafted features and pre-trained CNN model are both unsupervised in the framework, which makes the training more stable.

In addition, for the acc1s that less than but very close to 100%, because only a very small number of samples are mis-predicted, the standard deviation of the them is always lower than 0.01. These too low standard deviations are not only difficult to be written into the table, but also lose the significance of expressing training stability. In addition, from the perspective of acc1 only, for most models have reached an acc1 of 100%, it already reveals the undesirable training stability performance of the models with acc1 lower than 100%. Thus, the acc1 is a sufficient indicator even without standard deviation of it.

TABLE 3. Accuracy loss of different methods under varying illumination intensity.

Method	acc0-loss (%)
Edge Features+kNN [9, 10]	7.27
GLCM+SVM [8]	2.08
Fourier [11]+SVM	1.30
Wavelet+SVM [12]	2.60
Masking Model+SVM [29]	1.04
CNN	1.89±0.92
The proposed method	1.69±0.63

In terms of computational efficiency, the proposed framework performs not so well. However, this does not affect the practical application of the framework. For the fabric smoothness assessment is not a strong time-dependent task, the relatively low computation efficiency of the model does not affect its application in the industry.

In summary, the performance of the proposed framework in this experiment can well meet the needs of industrial applications.

C. MODEL ADAPTABILITY TO IMAGE ACQUISITION ENVIRONMENT

In instrumental application, the image acquisition environment may change in different situations. The environment adaptability of the method determines the ability of the trained model to expand its application. In actual application, most attributes of the image acquisition environment can be setup stably by a well-set image acquisition system, such as camera parameters, light source position, and sample position. The only uncertainty is illumination intensity.

In this experiment, we test the illumination intensity adaptability of the proposed method comparing with different previous methods. The acc0 loss caused by changes in light intensity was used as the evaluation index. The lower the acc0 loss, the better the model's adaptability to lighting intensity.

Table 3 shows the experimental results. Compared with other methods, the average acc0 loss of the proposed method is not the lowest, being 1.70%. But compared with the best performing method in [29], the acc0 loss of the proposed method is only 0.66% higher than it, which is still at an acceptable level. From the perspective of standard deviation, the proposed framework shows a better performance on average accuracy and deviation than the CNN model. This verifies that the fusion of hand-crafted feature can improve the model's adaptability to the illumination intensity and training stability to a certain extent.

D. MODEL ADAPTABILITY TO FABRIC VARIETY

Since we separate the fabric image decoloration task from the smoothness evaluation, this study does not discuss the effect of the color texture to the smoothness evaluation of the fabric. Besides the color texture, the visual texture produced by the fabric structure, namely structural texture, may produce noise in the images for smoothness appearance assessment. In this experiment, models were tested on the different varieties of fabric samples to discuss the effect of structural texture on the evaluation of fabric smoothness appearance.

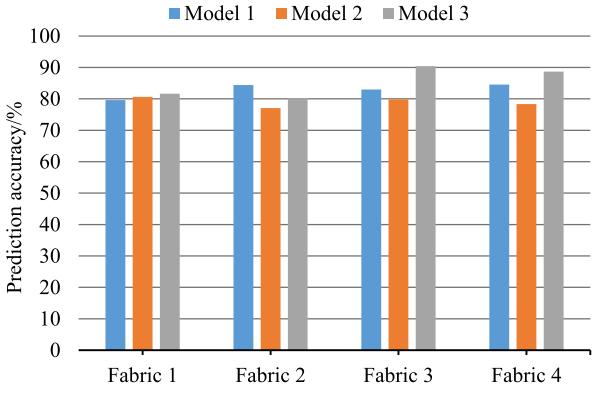

FIGURE 9. Distribution of the prediction accuracy (acc1) of models on different fabric varieties, where Model 1 is the SVM model trained with the hand-crafted features; Model 2 is the CNN model; Model 3 is the proposed image classification framework.

In Figure 9, the model 1 is the SVM model trained with the hand-crafted features; model 2 is the CNN model; and model 3 is the proposed image classification framework. And the four varieties of fabric correspond to the four fabrics introduced in table 1, which have different fabric structures. It can be observed that the testing accuracy of the models does not show a significant difference on different varieties, which proves the variety adaptability of such 2D-image-based methods. This is consistent with the fact that the structural texture is difficult to be visually observed in our fabric images (as can be seen from the fabric images in Figure 6 and 7). The reason may be that the image resolution we used is relatively to low (lower than 20 pixels per inch) to capture the structural texture on the fabric surface. However, the varieties in our data set is too few to fully prove this conclusion. The fabric image data set needs to be extended in our further study.

E. EFFECTIVENESS OF FEATURE FUSION AND SAMPLE FILTERING

It is expected that the introduction of hand-crafted features with unsupervised attributes will improve the training stability of the model, and the fusion of CNN features and hand-crafted features will result in a complementary effect and improve the overall accuracies in test. In addition, the mislabeled sample filtering is expected to improve the general performance of the model.

In this experiment, to evaluate the effectiveness of the sample filtering module and the multi-level feature fusion module, we performed a set of ablation experiment on the proposed image classification framework. In the experiment, pre-trained AlexNet [60], ResNet18 and ResNet34 [64] were tested as basic CNN models. For every basic CNN model, the performance of the model was tested with the hand-crafted feature fusion (hcf) and the mislabeled sample filter (msf) conducted in turn.

The test results are shown in Table 4. For each basic model, the best results obtained in the ablation experiment are

TABLE 4. Ablation experiment results for hand-crafted feature fusion and mislabeled sample filtering.

Method	acc0 (%)	acc0.5 (%)	acc1 (%)
AlexNet	76.3±1.30	91.1±0.76	99.73
AlexNet+hcf	84.4±**0.56**	95.9±**0.24**	100
AlexNet+hcf+msf	**85.2**±0.92	**96.1**±0.37	100
ResNet18	79.6±2.31	92.9±1.98	100
ResNet18+hcf	82.8±0.40	94.7±**0.30**	100
ResNet18+hcf+msf	**83.9±0.26**	**96.1**±0.52	100
ResNet34	79.9±2.40	92.1±2.30	100
ResNet34+hcf	83.3±**0.40**	**96.0±0.15**	100
ResNet34+hcf+msf	**84.0**±0.79	95.8±0.45	100

hcf: fusion with hand-crafted features;
msf: mislabeled sample filtering.

shown in bold in the table. From the perspective of accuracy and stability, the implementation of the hand-crafted feature fusion effectively improves the performance of the CNN models. Such test results verify the expectations presented in this paper that the fused hand-crafted features and CNN features result in a complementary effect in the proposed image classification framework.

In addition, from the perspective of accuracy, it can be observed that, in general, the test accuracies of the models increase with the implementation of the mislabeled sample filter (msf). The only exception is that the implementation of msf on ResNet34 reduces the acc0.5 of the model by 0.18%. However, the experimental results show that the implementation of msf widely improves the standard deviations of the model training results. This means that the implementation of msf has a certain negative effect on the stability of model training. The main reason could be that the filtering process reduces the training sample size, which makes the training more difficult. However, this stability weakening is relatively low and acceptable in application, and this problem is excepted to be mitigated by expanding the training sample size.

F. MODEL ADAPTABILITY TO LABEL NOISE

The proposed image classification framework contains a mislabeled sample filtering module, which is established to mitigate the negative effect of label noise in training process. To further verify the effectiveness and noise adaptability of this module, we added quantitative label noises to the training data set, and then discuss the testing performance of the framework with or without the mislabeled sample filtering module. The noise z was randomly added to a certain proportion (namely noise rate, γ) of sample labels in the training set. The noised sample label was randomly changed to one of its adjacent labels (simulating the actual subjective evaluation). In the experiment, γ was valued in 0%, 5%, 10%, 15%, 20%} and every model was trained three times with 10-fold cross validation. The average testing results are given in Figure 10.

Figure 10 (a) illustrates the trend of acc0 of the framework with or without msf in different noise rates. As the noise rate increases, the acc0 for both types of frameworks shows a decreasing trend. And the acc0 of framework with msf is always the higher, except when γ reaches 20%. This result

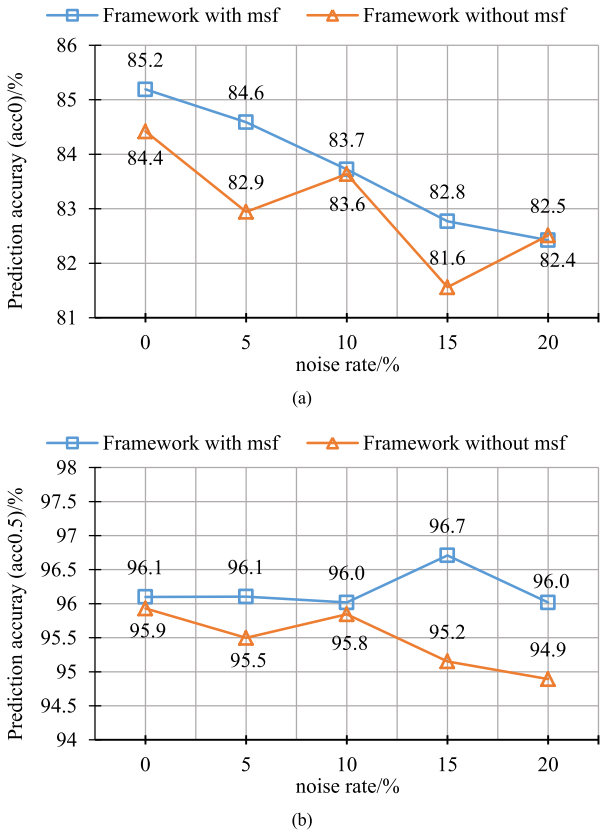

FIGURE 10. Testing prediction accuracies of the proposed framework with mislabeled sample filtering (msf) and framework without msf on different noise rates: (a) acc0, (b) acc0.5.

is consistent with our expectation, that is, the msf module can effectively reduce the interference of label noise on the model performance, but when the noise rate is too high, the effectiveness of the msf module may be reduced.

On the other hand, as in Figure 10 (b), the acc0.5 of the framework with msf is very stable to the increasing of noise rate. And the acc0.5 of framework without msf does not show the same trend, which decreases with the increase of noise rate. This result reveals that although the msf module cannot completely preven the model performance degradation caused by the label noise, it can effectively control the error within a small range.

G. DISCUSSION ON FEATURE FUSION FRAMEWORK

In the experiment described earlier, we compared the proposed framework with a series of methods for fabric smoothness appearance evaluation and some general image classification models, which demonstrates the its superior performance in the task. As the proposed framework with the fusion of hand-crafted and CNN features, in this experiment, we compared it with other feature fusion frameworks to further verify its structural superiority.

As introduced in section II, for hand-crafted and CNN features, the feature fusion frameworks mainly include the following categories: feature level fusion [32]–[34], [37], [38],

TABLE 5. The comparison results of different feature fusion frameworks.

Feature fusion framework	acc0 (%)	acc0.5 (%)	acc1 (%)
Hand-crafted feature only	82.6	95.8	100
CNN feature only	75.6	93.3	99.74
Feature level fusion	82.1	95.3	100
Fusion by ensemble	77.1	**96.1**	100
Multi-level fusion	83.9	95.8	100
The proposed framework (without msf)	**84.4**±0.56	95.9±0.24	100

msf: mislabeled sample filtering.

fusion by ensemble [35], multi-level fusion (that uses both feature level fusion and ensemble) [42], [67], and joint level fusion. As a representative of joint level fusion, the proposed framework was compared with the other three. In the compared frameworks, we used SVM and PCA as classifier and feature reducer for they are widely used in the field. The comparison results are given in table 5. In the table, as control groups, the first row gives the testing accuracies of the classifier trained with the hand-crafted features, and the second row shows the prediction accuracies of the classifier trained with the CNN feature reduced by PCA.

From the perspective of acc0, except the proposed framework, the multi-level fusion framework, which uses an ensemble of the classifiers in first three rows, showed a better performance than the others. In addition, the framework with feature fusion by ensemble achieves the highest acc0.5 with using an ensemble of the classifiers in the two control groups. These results further prove the complementarity between the hand-crafted and CNN features. In general, comparing with the other framework, the proposed framework showed the best overall performance. Although its acc0.5 was not the highest, it was only 0.17% lower than the highest. This result demonstrates the structure superior proposed framework.

VI. CONCLUSION

In this research, an effective image classification framework with a multi-level feature fusion module and a mislabeled sample filtering module was proposed for objective fabric smoothness appearance assessment. The proposed framework contains a mislabeled sample filter module and a multi-level feature fusion module. In the comparison experiments, the proposed framework achieved 85.2%, 96.1%, and 100% average accuracies under the errors of 0 degree, 0.5 degree, and 1 degree respectively, outperformed the stat-of-the-art models in existing research. Moreover, the training stability of the framework has also proved to be the best among various CNN methods. The ablation experiment demonstrated the important contributions of the multi-level feature fusion module and the mislabeled sample filtering module in the framework. And the proposed feature fusion module showed superior performance than the other existing feature fusion frameworks. More specifically, when the label error rate in the training set increases, the sample filtering module limited the prediction error to a smaller range in the experiment. Promisingly, the proposed framework can promote the industrial application of fabric smoothness appearance evaluation.

However, the study still has the following limitations. (1) Although we have explored the method for multi-color fabric image decoloration in our previous study [25], the performance of the smoothness appearance assessment methods should be further discussed on the decolored multi-color fabric images. (2) In order to further improve the training stability of the proposed method, the fabric image sample size should be further extended.

REFERENCES

[1] J. Fan and L. Hunter, *Engineering Apparel Fabrics and Garments*, 1st ed. Cambridge, U.K.: Woodhead, 2009.

[2] *Smoothness Appearance of Fabrics After Repeated Home Laundering*, Standard AATCC 124-2018, AATCC, Triangle Park, NC, USA, 2018.

[3] *Textiles—Test Method for Assessing the Smoothness Appearance of Fabrics After Cleansing*, Standard ISO 7768-2009, ISO, Cham, Switzerland, 2009.

[4] P. Xu, X. Ding, X. Wu, and R. Wang, "Characterization and assessment of fabric smoothness appearance based on sparse coding," *Textile Res. J.*, vol. 88, no. 4, pp. 367–378, Feb. 2018.

[5] B. Xu and J. A. Reed, "Instrumental evaluation of fabric wrinkle recovery," *J. Textile Inst.*, vol. 86, no. 1, pp. 129–135, Jan. 1995.

[6] R. Zaouali, S. Msahli, B. El Abed, and F. Sakli, "Objective evaluation of multidirectional fabric wrinkling using image analysis," *J. Textile Inst.*, vol. 98, no. 5, pp. 443–451, Sep. 2007.

[7] T. Mori and J. Komiyama, "Evaluating wrinkled fabrics with image analysis and neural networks," *Textile Res. J.*, vol. 72, no. 5, pp. 417–422, May 2002.

[8] X. Wang and M. Yao, "Grading the crease recovery with twisting of fabric by using image identification technique," *Int. J. Clothing Sci. Technol.*, vol. 12, no. 2, pp. 114–123, May 2000.

[9] H. C. C. Abril, E. Valencia, and M. A. S. Millán, "Objective assessment of wrinkled fabrics by optical and digital image processing," presented at the AIP Conf., 2008.

[10] H. C. Abril, M. S. Millán, and E. Valencia, "Influence of the wrinkle perception with distance in the objective evaluation of fabric smoothness," *J. Opt. A, Pure Appl. Opt.*, vol. 10, no. 10, Oct. 2008, Art. no. 104030.

[11] C. J. Choi, H. J. Kim, Y. C. Jin, and H. S. Kim, "Objective wrinkle evaluation system of fabrics based on 2D FFT," *Fibers Polym.*, vol. 10, no. 2, pp. 260–265, Apr. 2009.

[12] J. Sun, M. Yao, B. Xu, and P. Bel, "Fabric wrinkle characterization and classification using modified wavelet coefficients and support-vector-machine classifiers," *Textile Res. J.*, vol. 81, no. 9, pp. 902–913, Jun. 2011.

[13] J. Amirbayat and M. J. Alagha, "Objective assessment of wrinkle recovery by means of laser triangulation," *J. Textile Inst.*, vol. 87, no. 2, pp. 349–355, Jan. 1996.

[14] J. Su and B. Xu, "Fabric wrinkle evaluation using laser triangulation and neural network classifier," *Opt. Eng.*, vol. 38, pp. 1688–1693, Oct. 1999.

[15] N. Abidi, E. Hequet, C. Turner, and H. Sari-Sarraf, "Objective evaluation of durable press treatments and fabric smoothness ratings," *Textile Res. J.*, vol. 75, no. 1, pp. 19–29, Jan. 2005.

[16] C. Turner, H. Sari-Sarraf, E. F. Hequet, N. Abidi, and S. Lee, "Preliminary validation of a fabric smoothness assessment system," *J. Electron. Imag.*, vol. 13, pp. 418–428, 2004.

[17] M. Yao and B. Xu, "Evaluating wrinkles on laminated plastic sheets using 3D laser scanning," *Meas. Sci. Technol.*, vol. 18, no. 12, pp. 3724–3730, Dec. 2007.

[18] X. B. Yang and X. B. Huang, "Evaluating fabric wrinkle degree with a photometric stereo method," *Textile Res. J.*, vol. 73, no. 5, pp. 451–454, May 2003.

[19] J. Hu, B. Xin, and H. Yan, "Measuring and modeling 3D wrinkles in fabrics," *Textile Res. J.*, vol. 72, no. 10, pp. 863–869, Oct. 2002.

[20] T. J. Kang, D. H. Cho, and S. M. Kim, "New objective evaluation of fabric smoothness appearance," *Textile Res. J.*, vol. 71, no. 5, pp. 446–453, May 2001.

[21] W. Yu, M. Yao, and B. Xu, "3-D surface reconstruction and evaluation of wrinkled fabrics by stereo vision," *Textile Res. J.*, vol. 79, no. 1, pp. 36–46, Jan. 2009.

[22] W. Yu and B. Xu, "A sub-pixel stereo matching algorithm and its applications in fabric imaging," *Mach. Vis. Appl.*, vol. 20, no. 4, pp. 261–270, Jun. 2009.
[23] P. Xu, X. Ding, R. Wang, and X. Wu, "Feature-based 3D reconstruction of fabric by binocular stereo-vision," *J. Textile Inst.*, vol. 107, no. 1, pp. 12–22, Jan. 2016.
[24] J. Wang, K. Shi, L. Wang, R. Pan, and W. Gao, "A computer vision system for objective fabric smoothness appearance assessment with an ensemble classifier," *Textile Res. J.*, vol. 90, nos. 3–4, pp. 333–343, Feb. 2020.
[25] J. Wang, K. Shi, L. Wang, Z. Li, R. Pan, and W. Gao, "Decoloration of multi-color fabric images for fabric appearance smoothness evaluation by supervised image-to-image translation," *IEEE Access*, vol. 7, pp. 181284–181294, 2019.
[26] M. S. Hesarian, "Evaluation of fabric wrinkle by projected profile light line method," *J. Textile Inst.*, vol. 101, no. 5, pp. 463–470, Apr. 2010.
[27] W. Ouyang, "Evaluating fabric pilling/wrinkling appearance using 3D images," M.S. thesis, Textile Apparel Technol., Univ. Texas Austin, Austin, TX, USA, 2013. [Online]. Available: https://repositories.lib.utexas.edu/handle/2152/23656
[28] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1794–1801.
[29] J. Wang, K. Shi, L. Wang, Z. Li, R. Pan, and W. Gao, "An objective fabric smoothness assessment method based on a multi-scale spatial masking model," *IEEE Access*, vol. 7, pp. 73830–73840, 2019.
[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[31] J. Wang, K. Shi, L. Wang, Z. Li, F. Sun, R. Pan, and W. Gao, "Automatic assessment of fabric smoothness appearance based on a compact convolutional neural network with label smoothing," *IEEE Access*, vol. 8, pp. 26966–26974, 2020.
[32] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, "Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features," *IEEE Access*, vol. 7, pp. 105146–105158, 2019.
[33] Y. Lei, X. Chen, M. Min, and Y. Xie, "A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification," *Neurocomputing*, vol. 381, pp. 186–195, Mar. 2020.
[34] Q. Shi, W. Li, F. Zhang, W. Hu, X. Sun, and L. Gao, "Deep CNN with multi-scale rotation invariance features for ship classification," *IEEE Access*, vol. 6, pp. 38656–38668, 2018.
[35] H. Kuang, X. Zhang, Y.-J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 927–936, Apr. 2017.
[36] S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S. Zheng, "An enhanced deep feature representation for person re-identification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–8.
[37] R. Su, T. Liu, C. Sun, Q. Jin, R. Jennane, and L. Wei, "Fusing convolutional neural network features with hand-crafted features for osteoporosis diagnoses," *Neurocomputing*, vol. 385, pp. 300–309, Apr. 2020.
[38] C. Wang, A. Elazab, J. Wu, and Q. Hu, "Lung nodule classification using deep feature fusion in chest radiography," *Comput. Med. Imag. Graph.*, vol. 57, pp. 10–18, Apr. 2017.
[39] W. Ouyang, B. Xu, J. Hou, and X. Yuan, "Fabric defect detection using activation layer embedded convolutional neural network," *IEEE Access*, vol. 7, pp. 70130–70140, 2019.
[40] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, Dec. 2016.
[41] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. Int. Conf. Artif. Neural Netw.*, 2018, pp. 270–279.
[42] M. A. Khan, M. Sharif, T. Akram, M. Raza, T. Saba, and A. Rehman, "Hand-crafted and deep convolutional neural network features fusion and selection strategy: An application to intelligent human action recognition," *Appl. Soft Comput.*, vol. 87, Feb. 2020, Art. no. 105986.
[43] C. Liu, Y. Fu, and N. Wu, "Novel testing equipment for fabric wrinkle resistance simulating actual wear," *Textile Res. J.*, vol. 84, no. 10, pp. 1059–1069, Jun. 2014.
[44] C. Liu, "Investigation on the novel measurement for fabric wrinkle simulating actual wear," *J. Textile Inst.*, vol. 108, pp. 279–286, 2016.
[45] J. Silvestre-Blanes, J. Berenguer-Sebastiá, R. Pérez-Lloréns, I. Miralles, and J. Moreno, "Garment smoothness appearance evaluation through computer vision," *Textile Res. J.*, vol. 82, no. 3, pp. 299–309, Feb. 2012.
[46] L. Wang and W. Gao, "Comprehensive evaluation of the fabric crease recovery property by the whole recovery process," *Textile Res. J.*, to be published. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/0040517519899177, doi: 10.1177/0040517519899177.
[47] L. Wang, J. Liu, R. Pan, and W. Gao, "Dynamic measurement of fabric wrinkle recovery angle by video sequence processing," *Textile Res. J.*, vol. 84, no. 7, pp. 694–703, May 2014.
[48] F. Sun, M. Guo, X. Hu, L. Wang, and W. Gao, "Analysis of curve parameters to characterize multidirectional fabric wrinkling by a double extraction method," *Textile Res. J.*, vol. 89, no. 15, pp. 2973–2982, Aug. 2019.
[49] X. Hu, F. Sun, Q. Wang, and W. Gao, "*In situ* characterization of the morphological wrinkling of woven fibrous materials by a mechanical test," *Textile Res. J.*, to be published. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/0040517520910709, doi: 10.1177/0040517520910709.
[50] Y. Lu, X. Hu, F. Sun, F. Peng, and W. Gao, "Determination of optimal system parameters to characterize the wrinkle recovery of fabrics by an integrated shape retention evaluation system," *Textile Res. J.*, vol. 90, no. 1, pp. 91–100, Jan. 2020.
[51] J. L. Crowley, O. Riff, and J. H. Piater, "Fast computation of characteristic scale using a half octave pyramid," presented at the Int. Conf. Scale-Space Theories Comput. Vis., 2002.
[52] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[53] Y. Wang and S.-C. Zhu, "Perceptual scale-space and its applications," *Int. J. Comput. Vis.*, vol. 80, no. 1, pp. 143–165, Oct. 2008.
[54] R. X. Lukac, *Perceptual Digital Imaging: Methods and Applications*. Boca Raton, FL, USA: CRC Press, 2012.
[55] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, and C.-C.-J. Kuo, "Enhanced just noticeable difference model for images with pattern complexity," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2682–2693, Jun. 2017.
[56] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 14, no. 9, pp. 2379–2391, Sep. 1997.
[57] C.-H. Chou and Y.-C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, Dec. 1995.
[58] J. Wu, W. Lin, G. Shi, X. Wang, and F. Li, "Pattern masking estimation in image with structural uncertainty," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4892–4904, Dec. 2013.
[59] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[61] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.
[62] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *Proc. 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 1–4.
[63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[65] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[66] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[67] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *J. Med. Imag.*, vol. 1, no. 3, Oct. 2014, Art. no. 034003.