IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 14, 2017, accepted January 24, 2018, date of publication February 5, 2018, date of current version March 13, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2802498

No Reference Image Quality Assessment based on Multi-Expert Convolutional Neural Networks

CHUNLING FAN[1,2], YUN ZHANG[1], (Senior Member, IEEE), LIANGBING FENG[1], AND QINGSHAN JIANG[1]

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China

Corresponding author: Yun Zhang (yun.zhang@siat.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471348, in part by the Shenzhen Key Technologies Program under Grant JSGG20160229123657040, in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2016A030306022, in part by the Key Project for Guangdong Provincial Science and Technology Development under Grant 2017B010110014, in part by the Shenzhen International Collaborative Research Project under Grant GJHZ20170314155404913, and in part by the Shenzhen Science and Technology Foundation under Grant JCYJ20150630114942277.

ABSTRACT No Reference (NR) Image Quality Assessment (IQA) algorithm is capable of measuring the quality of distorted images without referencing the original images. This property is of great importance in image processing, compression, and transmission. However, due to the diversity of the distortion types and image contents, it is difficult for the existing NR IQA algorithms to be applied and maintain the best performance for all cases. To address this problem, we develop a novel NR IQA algorithm based on multi-expert convolutional neural networks (CNNs), which consists of distortion type classification, CNN based IQA algorithms and fusion algorithm. First, we present a distortion type classifier to identify the distortion type of the input image. Then, we propose a multi-expert CNN based IQA algorithms for each type of distortion. Finally, a fusion algorithm is adopted to aggregate the classification result of distortion types and multi-expert CNN based image quality predictions. The proposed algorithm has been tested on commonly used LIVE II database and a cross-dataset evaluation was carried on CSIQ database. The experimental results show that the proposed algorithm provides effective improvements for NR IQA.

INDEX TERMS Image quality assessment, no reference image quality assessment, distortion type classification, multi-expert CNN.

I. INTRODUCTION

With the rapid development of digital technologies, multimedia applications have been widely applied to facilitate human daily life, such as video surveillance, high definition digital TV, distant education, video-on-demand system, and other applications. These applications produce a large number of digital images and videos every day. However, due to distortions or artifacts in acquisition, processing, and display such as capturing defects, processing noise, transmission error, and compression distortion, the quality of acquired and processed images may be reduced, which may hence degrade human visual experience. How distortions in natural images effect human visual experience is an open problem and a hot topic. Based on the availability of the reference image, algorithms on objective Image Quality Assessment (IQA) can be divided into three classes: Full Reference (FR), Reduced Reference (RR), and No Reference (NR). FR requires the reference image to evaluate distorted images. RR requires partial information of the reference images and NR assesses image quality score without using any information of the reference images. In many real applications, the information of the reference images is often unavailable or difficult to be acquired, so NR IQA algorithms are highly desirable and practically more challenging.

A. RELATED WORKS

In general, existing NR IQA algorithms can be classified into two categories: distortion-specific NR IQA algorithms and general purpose NR IQA algorithms. The former is used to predict the image quality in case of specific distortions, such as blur [1], [2], JP2K [3], JPEG [4], noise [5], and contrast [6]. The latter is used to predict the image quality score across different distortion types. However, in real applications, the information of image distortion type is often unavailable beforehand, so the latter algorithm is more practical and highly demanding. In detail, the general purpose NR IQA algorithms can be further categorized into two classes: statistical characteristics based and learning based. With the

analyses on statistical characteristics of images, hand-crafted feature extractors are designed according to researchers' professional knowledge, such as Natural Scene Statistics (NSS), wavelet coefficients statistics, Discrete Cosine Transform (DCT) coefficients [7], and Gabor-filter-based local features [8]. Saad *et al.* [7] proposed an NR IQA algorithm named BLIINDS-II based on an NSS model of DCT coefficients. Ye and Doermann [8] introduced an NR IQA algorithm using visual codebooks, which consists of Gabor-filter-based local image features. Xie *et al.* [9] introduced an NR IQA approach using bag-of-words model based on local quantized pattern features. These statistical characteristics are usually represented by traditional probabilistic models, such as Generalized Gaussian Distribution (GGD) and Weibull distribution [10]. The performance of these algorithms highly rely on the feature extractors.

Learning-based NR IQA algorithms can be further divided into learning algorithms based on hand-crafted features and automatic learning characteristics. Hand-crafted features based algorithms usually have two steps: hand-crafted feature extraction and quality prediction. These algorithms firstly extract image features and then learn a regression model from image features to image quality score, such as Support Vector Regression (SVR). Xue *et al.* [11] jointed the gradient magnitude map and the Laplacian of Gaussian response, and then learned a regression function using SVR. Liu *et al.* [12] utilized multiple kernel learning to learn the mapping function between the image features and image quality score. These algorithms utilized hand-crafted feature extractors, while the understanding of human visual system is still limited, the evaluation of these algorithms can not accurately reflect the subjective perception of human visual characteristics. Automatic learning characteristics based algorithms try to learn quality aware image features from raw images. Ye *et al.* [13] introduced an unsupervised learning algorithm which learned a dictionary from a set of unlabeled image patches. Hou *et al.* [14] developed a deep learning network for NR IQA, which could convert the learned qualitative labels into numerical quality scores using a quality pooling. Kang *et al.* [15] proposed a shallow Convolutional Neural Network (CNN) based NR IQA. With one convolutional layer, two pooling algorithms, and two fully connected layers, it can learn a nonlinear mapping from normalized image patches and its quality score. Wang *et al.* [16] developed a CNN based approach, which could identify the distortion type of an image and predict its quality score with only one general CNN network. These algorithms are end-to-end and can learn complex mapping from raw image to its quality score. However, they all designed one general network for different kinds of distortion, which can hardly outperform other algorithms across all distortion types.

B. MOTIVATION AND ANALYSES

In practical applications, there are many different types of distortions, such as JPEG, JPEG2000 (JP2K), White Noise (WN), Gaussian Blur (GBLUR), Fast Fading (FF) and so on. Fig.1 shows an example of distorted images with similar Mean Squared Error (MSE), in which (a) is the reference image and (b)-(f) are five distorted versions (GBLUR, JPEG, WN, JP2K, and FF) in LIVE II database [17]. DCT and wavelet transform are utilized in JPEG and JP2K respectively. WN adds white Gaussian noise to the images. GBLUR is the result of filtering an image using a Gaussian kernel. We can observe that the distorted images are significantly different from each other, though they are from the same content and MSE. Because of the significant differences among these distortions, it is extremely challenging to find a universally applicable features set which are sensitive to various distortion types for quality assessment.

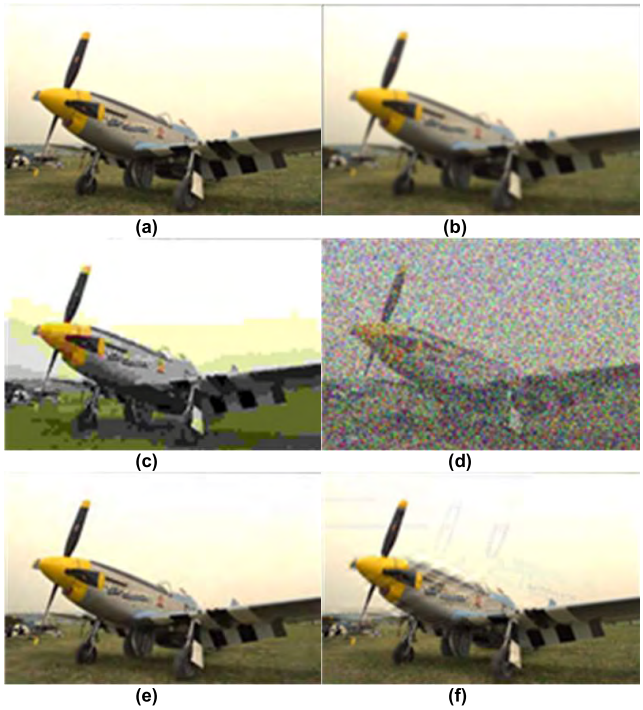

FIGURE 1. One reference image and its five distorted versions. (a) Reference image. (b) GBLUR. (c) JPEG. (d) WN. (e) JP2K. (f) FF.

To analyze the IQA performance over different distortion types more precisely, we compare the performance of some of state-of-the-art IQA schemes, including BRISQUE [18], BLIINDS-II [6], DIIVINE [19], [20], NSS-TS [21], CORNIA [13], TCLT-Gray [22], and NIQE [23]. Fig. 2 shows schemes with top-3 the Spearman Rank Order Correlation Coefficient (SROCC) values for each distortion type in LIVE II database. While BRISQUE (drawn in red legend in Fig.2) performs the best for JPEG distortion, it is in the second place for GBLUR, and the third place for WN. DIIVINE (drawn in blue legend in Fig.2) performs the best for WN, but not in the top three for other four distortion types. We can find that there is no one single scheme that outperforms others across all the five distortion types in LIVE II database, which means different distortion types have different characteristics and shall be treated specifically. Similar findings are also reported

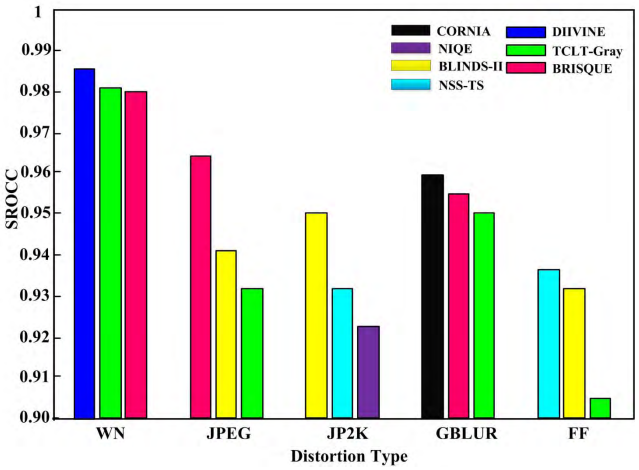

FIGURE 2. Top three quality schemes for five image distortion types in LIVE II database.

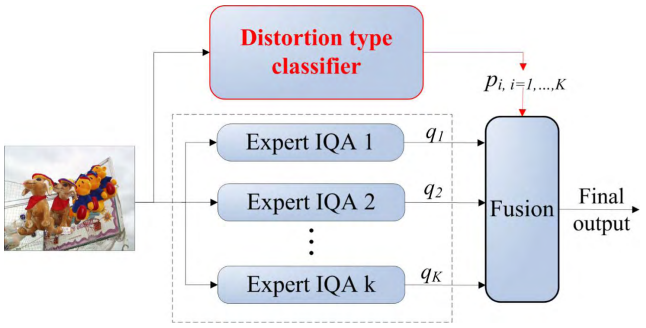

FIGURE 3. Flowchart of the proposed algorithm.

in recent works [24] and [25]. Wu *et al.* [24] found that the two statistics including normalized Histogram of Oriented Gradient (HOG) in the wavelet sub band and the marginal distribution of wavelet coefficients in a frequency band present obvious differences under different distortion degrees for the five distortion types (WN, GBLUR, JPEG, JP2K, and FF). Zhang *et al.* [25] analyzed the statistical distributions of the reference image and its distortion versions under WN, GBLUR, and JPEG. They found that the distribution of Man Subtracted Contrast Normalized (MSCN) coefficients of the reference image can be well analyzed with GGD regression, but the distorted images can not. More importantly, fitting errors are effected by distortion type and distortion degree. These findings further demonstrate that images under different types of distortion should be assessed with various statistical characteristics. Generally, a single algorithm can hardly always being the best for all distortion types.

To address this problem, a novel NR IQA algorithm based on Multi-expert CNN is proposed in this paper, in which identifying the distortion types is modeled as a multi-class classification task and then multi-expert CNN is designed for the IQA of each distortion type. The remainder of this paper is organized as follows. Section II presents the proposed NR IQA algorithm and the implementation details. In section III, the experimental results and analyses are illustrated. Finally, section IV draws the conclusions of this work.

II. THE PROPOSED IQA-MCNN

A. FRAMEWORK OF THE PROPOSED IQA-MCNN

Fig. 3 shows the framework of the proposed IQA-MCNN, which includes three major components. The first part is Distortion Type Classifier (DTC), which gives the probabilities of the input image belonging to each distortion type as P=$\{p_i, i = 1, \ldots, K\}$. The second part is composed of a group of IQA experts for multiple distortion types. The input of each expert is the distorted image to be assessed, and the output is its predicted quality score q_i. If the number of distortion types is large, clustering method can be applied on distortion types and we can design an expert for each group of distortion types. Any distortion-specific traditional algorithm can be used as an expert, such as JP2K-oriented algorithm [3] for JP2K distortion, and blur-oriented algorithm [1] for GBLUR distortion. But in this paper, we design a CNN as an expert for each distortion type, because CNN performs excellent in automatically feature learning. The third part is a fusion algorithm, which aggregates the output of the DTC and the multi-expert CNN and then give the final quality score. With our proposed architecture, a complicated IQA problem of multiple distortion types can be divided into IQA of fewer groups of similar distortion types, or even several single distortion types.

B. IMAGE PREPROCESSING

Natural images vary significantly in intensity, local normalization are required in IQA for its robustness to intensity and contrast change. For a given gray-scale image, we first perform a local contrast normalization which is the same as BRISQUE [18]. Given an image, let the intensity of the pixel at location (i, j) be $I(i, j)$, we compute its normalized intensity value as

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}, \tag{1}$$

where C is a small positive constant, $\mu(i,j)$ and $\sigma(i,j)$ are mean and variance, respectively, which are defined as

$$\mu(i,j) = \sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} I(i+p, j+q), \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} (I(i+p, j+q) - \mu(i,j))^2}, \tag{3}$$

where P and Q are the width and height of normalization window, respectively. Local contrast normalization is important for improving the performance.

In order to prevent overfitting, data augmentation is commonly used [14]. Here, we crop a set of non-overlapping $B \times B$ image patches from each preprocessed image, and label each image patch with the same distortion type of its source image.

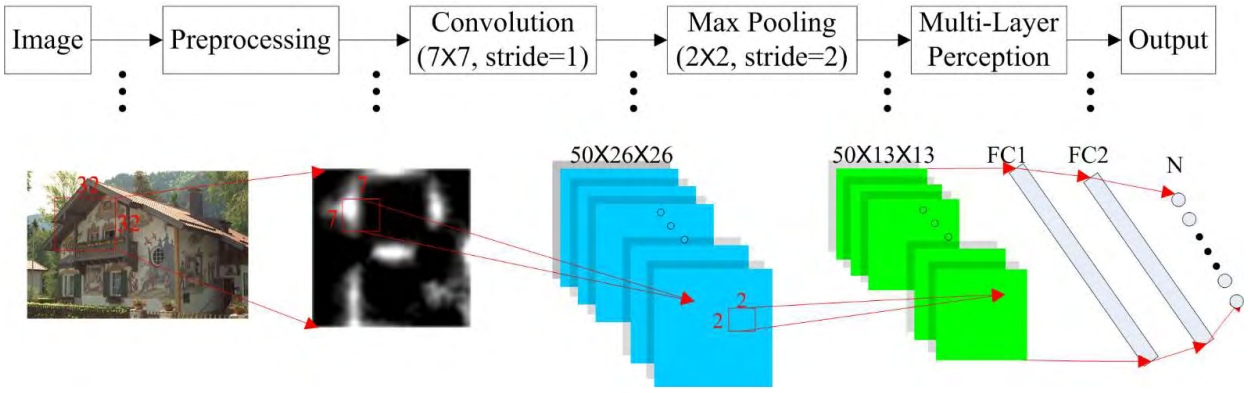

FIGURE 4. CNN Structure for DTC and IQA expert.

Also, we assign a quality score of each image patch with the ground-truth of its source image because the distortion is homogeneously.

C. DESIGN FOR CNN BASED DTC

It is reported that the human visual system is a hierarchical structure, in which the low-level image features and more complex features are extracted from the early visual area and higher visual areas respectively [22]. Fortunately, CNN is developed to fit the properties of human visual system. Specifically, the first CNN layer extracts low-level image features, e.g. edges, luminance, and contrast. The later CNN layers learn high-level image features, e.g., object parts, image patterns, and content, which is important and useful for image recognition. Since CNN has excellent performance in image pattern recognition and image distortion type classification in [16], [26] in recent years, CNN is adopted here for DTC. We choose a shallow CNN with one convolution layer in our architecture, as shown in Fig.4. In consideration of the size of receptive field and the amount of parameters, we will compare different configurations in convolutional layer in section III. Pooling is needed to reduce the dimension of the learned feature maps and improve the position invariance. The following is a max pooling layer with kernel size of 2×2 and stride of 2. Three Fully Connected (FC) layers comes after the pooling, which are used to summarize the representation and give a quality score. In order to introduce non-linearity into the system, we use Rectified Linear Units (ReLUs) as the neurons in the first two FC layers. The ReLUs can be represented as $f(x) = max(0, x)$, where x denotes the input. Compared with the traditional tanh units, the CNN networks with ReLUs train is much faster [27]. The dimension of the output layer is equal to the number of distortion types and we use Soft-max in the output layer. As (4) shows, for an input image patch x_n, Soft-max produces a probability for each distortion type

$$p(y_n = k|x_n) = \exp\left(\alpha_n^k\right) / \sum_{i=1}^{K} \exp\left(\alpha_n^i\right), \qquad (4)$$

where y_n is the label of the distortion type, K is the number of distortion types, α_n^k is the output of the kth neuron in the layer, $k \in \{1, 2, \cdots, K\}$. These probabilities will be used as weights for different distortion types in fusion algorithm.

D. DESIGN FOR CNN BASED IQA EXPERT

For each distortion type, we design a specific IQA expert and train it using distortion-specific samples. Each expert should extract distortion-specific aware features and learn a nonlinear mapping from the image features to image quality score for each distortion type. Take JPEG as an example, the expert should learn quality aware features which can represent the characteristic of JPEG the best. It is worth mentioning that we can use any traditional distortion-specific IQA algorithm as an expert network. In this paper, we will develop CNN based IQA expert. Each expert is a nonlinear regression problem for IQA, and we choose the popular CNN [15], [16] as an expert for two reasons. Firstly, CNN can fit any nonlinear function with the supervised learning framework, even if we know nothing about the nonlinear function. Secondly, since the understanding of the human visual system is limited, it is hard to choose hand-crafted features which can perfectly reflect the human visual system in IQA. We need a CNN based IQA expert to learn the features from the raw images automatically.

In traditional IQA algorithms, low-level features are often used. For instance, luminance, contrast, and structure information are used in SSIM [28]. In [29] and [30], they both used low-level features learned via shallow CNN in IQA. Kang *et al.* [15] used CNN for IQA with only one network for predicting quality score of multi-type distortions which is different from our expert system, but it gives us some information about the network structure. Kang analyzed that larger patch size increases IQA performance slightly but costs more processing time. To have a good balance between the prediction accuracy and complexity, we choose 32×32 image patches as input. Similar to the DTC, the first layer is a convolutional layer with 50 kernels and the kernel size is 7×7, the stride is 1 pixel. The first convolutional layer generates

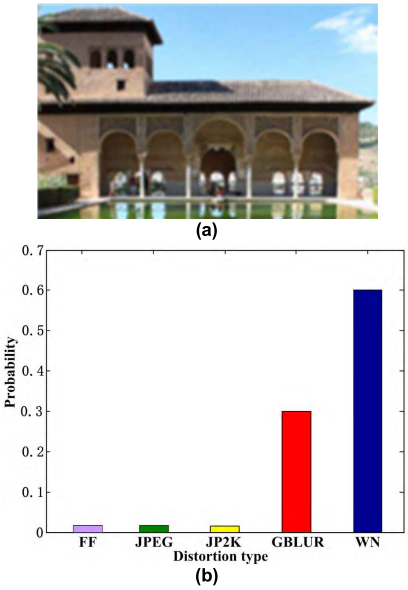

FIGURE 5. One image with multiple distortion types and the probability distribution of multiple distortion types. (a) Distorted image with multiple distortion types. (b) The probability distribution of multiple distortion types.

50 feature maps each of size 26×26. The following is a max pooling layer with kernel size of 2×2 and stride of 2. Generally, two FC layers are often used for mapping the two dimensional feature maps to a feature vector. Here, we use two FC layers each of 400 nodes after the max pooling layer.

Different from DTC, the last FC layer with $N = 1$ node is a simple regression, which gives the quality q_i of the input image I. We use a sigmoid neuron in the last FC layer, which can be represented as

$$q_i = \frac{1}{1 + e^{-x}}, \tag{5}$$

where x denotes the input of the neuron in the last layer. Given an image patch x_i and its quality score y_i, the objective function we used is

$$L = \frac{1}{N} \sum_{i=1}^{N} \|y_i - \widehat{y_i}\|_{l_2}, \tag{6}$$

where $\widehat{y_i}$ is the prediction quality score.

E. FUSION ALGORITHM

For those distorted images with multiple distortion types, the quality score should be assessed by multiple distortion-specific IQA experts together. Fig.5a shows a distorted image, which is firstly blurred and then filtered by white Gaussian noise. Fig.5b shows the probability of the image shown in Fig.5a. As Fig.5b shows, the horizontal axis denotes different distortion types and the vertical axis denotes the probability of different distortion types. The probabilities for GBLUR and WN are large, and the probabilities for JPEG, JP2K and FF are almost close to zero, because the image is firstly blurred and then filtered by white Gaussian noise.

For an image I, after multi-expert CNN, we actually get a set of random discrete variables, $Q = \{q_1, q_2, \ldots, q_K\}$, q_i is the prediction score of the i^{th} expert. The probability mass function of the random discrete variables is $p(q_i) = p_i$, where p_i is the probability belonging to the i^{th} distortion type, and satisfies

$$\sum_{q_i \in Q} p\,(q_i) = 1. \tag{7}$$

There are many kinds of fusion strategies, such as max probability pooling, maximal score pooling, and weighted average. Max probability pooling outputs the score with maximal probability from all experts, maximal score pooling gives the largest score from all experts. Actually, Max probability pooling and maximal score pooling are two special cases of weighted average. In order to make full use of multi-expert CNN networks and DTC, we choose weighted average, which is consistent with the fact that most distortions tend to be additive.

From the probabilistic perspective, since the output score is random variable, the best fusion result is the expectation of the random variable over the distortion types, which is

$$E\,(q) = \sum_{q_i \in Q} q_i \cdot p\,(q_i). \tag{8}$$

III. EXPERIMENTAL RESULTS AND ANALYSES

In this section, we train and test the proposed architecture on Caffe [31]. We firstly test the performance of DTC, and then evaluate the performance of the proposed algorithm IQA-MCNN and make comparison with the state-of-the art algorithms.

Dataset:

1) LIVE II: The LIVE II dataset is composed of 29 reference images and their 799 distorted versions with five typical distortion types: JP2K, JPEG, WN, GBLUR, and FF. In our experiment the dataset is split into three non-overlapping subsets. In order to ensure sample proportion consistency in training, validation, and test set. Among all the images, 19 of 29 reference images and their distorted versions of each distortion type are used as training set, 5 of 29 reference images and their distorted versions are used as validation set, and the rest 5 reference images and their distorted versions are used to test the performance of the proposed algorithm.

2) CSIQ [32]: The CSIQ dataset consists of 30 reference images and their 866 distorted ones with six distortion types at five different distortion levels. We mainly focus on the four common types (JP2K, JPEG, WN, and BLUR) with the LIVE II dataset to evaluate the performance of database independence.

Training, Validation and Test Settings:

Inspired by Kang *et al.* [15], we train the DTC and multi-expert CNN on locally normalized non-overlapping 32×32 image patches from each image. We update the network with momentum strategy. During training stage, base learning rate is 0.01 and gradually decrease with the number

TABLE 1. CNN based DTCs configurations and their prediction accuracies.

	DTC-A	DTC-B	DTC-C	DTC-D
CNN based DTCs Configurations	Conv7-50	Conv7-50	Conv3-50	Conv3-50
	Max pool	Max pool	Max pool	Max pool
	FC-400	FC-800	Conv3-50	Conv3-50
	FC-400	FC-800	FC-400	FC-800
	FC-5	FC-5	FC-400	FC-800
			FC-5	FC-5
Prediction Accuracies	94.78%	99.25%	94.78%	99.25%

of epochs, weight decay is 0.0005. Stochastic gradient decent and back propagation are the most commonly used in CNN network. Here we also choose this algorithm to learn the parameters in all networks.

Protocols:

Three measures are used to evaluate the performance of the proposed algorithm with others, i.e., the Pearson's Linear Correlation Coefficient (PLCC), SROCC, and the Root Mean Squared Error (RMSE). The values of PLCC and SROCC are both in the range of [0, 1], and the larger PLCC and SROCC values means the better performance. For RMSE, the smaller value means the better performance. Before PLCC and RMSE were calculated, the prediction score x was fitted to DMOSp using the logistic function [33]

$$DMOSp = \beta_1(\frac{1}{2} - \frac{1}{1+e^{\beta_2(x-\beta_3)}}) + \beta_4 x + \beta_5, \qquad (9)$$

where β_1 to β_5 are parameters found using a nonlinear regression.

A. PERFORMANCE EVALUATION FOR THE PROPOSED DTC

To testify the performance of the CNN based DTC, we design and compare four different DTCs. As listed in Table 1, the base network is DTC-A with one convolutional layer with 50 kernels each of size 7×7, one max pooling layer, two FC layers (each with 400 nodes), and the last layer with 5 nodes standing for 5 distortion types in LIVE II dataset. Compared with DTC-A, DTC-B has double size (800 nodes) of neurons in the two FC layers. DTC-C adopts two convolutional layers with 50 kernels each of size 3×3, DTC-D has double size of neurons in the two FC layers, compared with DTC-C. During training, we only use distorted images of the training set to avoid label confusion. It is worth noting that reference images can be labeled as any distortion type among the five distortion types, so the test set only consists distorted images. As shown in Table 1, the top-1 accuracy for distorted images of DTC-B is 99.25%, which is higher than DTC-A, because DTC-B has more neurons in FC layers. Similar results can be observed from DTC-C and DTC-D. In summary, DTC-D and DTC-B have the same accuracy, but they give different probabilities for each distortion type. In subsection B, we will compare the performance of these DTCs.

TABLE 2. PLCC comparison between the proposed IQA-MCNN and the benchmark schemes.

Schemes	JPEG	JP2K	WN	GBLUR	FF	ALL
PSNR	0.9463	0.9542	0.9932	0.9211	0.9352	0.9292
SSIM	0.9849	0.9805	0.9428	0.9670	0.9077	0.9647
VIF	0.9527	0.9778	0.9558	0.9818	0.9736	0.9696
DIIVINE	0.8020	0.8724	0.9684	0.9486	0.9123	0.8152
OG-IQA	0.9445	0.9560	0.9762	0.9380	0.9369	0.9366
TCLT*	0.948	0.902	**0.989**	**0.955**	0.923	0.935
Kang	0.9563	0.9591	0.9673	0.9486	0.9838	0.9500
IQA-SCNN	0.9382	0.9366	0.5518	0.9137	0.9301	0.8439
IQA-MCNN1	**0.9570**	0.9642	0.9869	0.9457	0.9817	0.9569
IQA-MCNN2	**0.9570**	**0.9643**	0.9869	0.9459	**0.9843**	**0.9572**

B. PERFORMANCE EVALUATION FOR THE PROPOSED IQA-MCNN

The training set is used to learn the DTC and distortion-specific expert networks. We compare the proposed IQA-MCNN with three FR IQA algorithms (PSNR, SSIM [28], and VIF [31]) and some NR IQA algorithms (DIIVINE [19], [20], TCLT [22], OG-IQA [34] [35], and Kang [15]). In IQA-MCNN1, we adopt DTC-B. In IQA-MCNN2, we adopt DTC-D. In order to evaluate the effectiveness of our proposed IQA-MCNN1 and IQA-MCCN2 further, we also train a single CNN using images of all distortion types together, which is called single CNN based IQA (IQA-SCNN). The configuration of IQA-SCNN is the same as one of the expert network, as shown in Fig.4.

We tested the performances of 10 different algorithms on LIVE II database, and Table 2 shows the PLCC comparison between the proposed IQA-MCNN and the benchmark schemes. The best result in each column are marked in bold. The above three rows are FR IQA algorithms, and the below seven rows are NR IQA algorithms. Although the PLCC of SSIM and VIF are 0.9647 and 0.9696, which perform better than NR IQA algorithms, but they need the information of reference images. It should be noted that the results of TCLT* scheme is TCLT-Gray, which is quoted from the literature [22]. The PLCC of TCLT for WN and GBLUR are 0.989 and 0.955, which are slightly higher than IQA-MCNN. It was trained on 23 of 29 reference images and their distorted versions and tested on the rest 6 reference images and their distorted versions, which is slightly different from the settings in the paper. The PLCC of DIIVINE for all the test data is 0.8152 which is lower than our proposed algorithm. DIIVINE also identified distortion types firstly and then train SVR for distortion-specific image quality regression, but it used SVM based on natural scene statistics. As Table 2 shows, the PLCC of IQA-MCNN2 for all test data is 0.9572, which is higher than other NR IQA algorithms and IQA-MCNN2 performs better for most of the distortion types. That indicates

TABLE 3. SROCC comparison between the proposed IQA-MCNN and the benchmark schemes.

Schemes	JPEG	JP2K	WN	GBLUR	FF	ALL
PSNR	0.9273	0.9289	0.9835	0.8802	0.8404	0.9020
SSIM	0.9635	0.9745	0.9844	0.9790	0.9339	0.9582
VIF	0.9654	0.9759	0.9893	0.9839	0.9312	0.9740
DIIVINE	0.7818	0.8169	0.9764	**0.9674**	0.7797	0.8107
OG-IQA	0.9181	0.9402	0.9768	0.9148	0.9291	0.9389
TCLT*	0.932	0.898	0.980	0.947	0.903	0.934
Kang	**0.9407**	0.9408	0.9701	0.9509	0.9407	0.9440
IQA-SCNN	0.9129	0.9298	0.3299	0.9420	0.9010	0.8382
IQA-MCNN1	0.9381	**0.9508**	0.9813	0.9358	0.9402	0.9529
IQA-MCNN2	0.9390	0.9498	**0.9822**	0.9358	**0.9434**	**0.9531**

TABLE 4. RMSE comparison between the proposed IQA-MCNN and the benchmark schemes.

Schemes	JPEG	JP2K	WN	GBLUR	FF	ALL
PSNR	7.8558	7.3080	2.5342	7.7694	7.5260	8.3821
SSIM	4.2137	4.7937	7.2590	5.0855	8.9179	5.9732
VIF	4.4238	5.1170	6.4051	3.9870	7.2048	5.8360
DIIVINE	14.519	14.707	5.651	6.618	10.455	13.913
OG-IQA	7.1651	7.9861	4.7240	6.9195	7.4296	7.9447
TCLT*	**5.340**	7.721	**2.571**	**5.085**	6.964	**5.131**
Kang	7.0910	6.8883	5.5112	6.2984	3.7962	7.0807
IQA-SCNN	8.388	8.532	18.228	8.088	7.788	12.153
IQA-MCNN1	7.0333	6.4592	3.5168	6.4688	4.0399	6.5632
IQA-MCNN2	7.0311	**6.4506**	3.5007	6.4538	**3.7395**	6.5651

IQA-MCNN2 has much higher correlation with ground-truth DMOS under most of the distortion types. The PLCC of IQA-MCNN2 are all higher than IQA-MCNN1, which means IQA-MCNN2 performs better than IQA-MCNN1. Although DTC-B and DTC-D have almost the same accuracy in test data, they give different probabilities for each distortion type. If the classification accuracy can be improved further, then we can improve the quality prediction of the proposed algorithm. IQA-MCNN1 and IQA-MCNN2 have larger PLCC values than IQA-SCNN under five distortion types, which means the proposed IQA-MCNN gives great improvement compared with IQA-SCNN. PLCC for WN of IQA-SCNN is 0.5518, which means IQA-SCNN can work well for other four distortion types but not WN. While PLCC of IQA-MCNN2 under WN is 0.9869, which means it works well for WN. We believe this is due to the reason that we train an expert network for each distortion type, which can learn discriminative features for distortion type WN.

As Table 3 shows, the SROCC of VIF is 0.9740, which performs the best among all the 10 schemes, but VIF needs information of reference images during testing. The SROCC of IQA-MCNN1 and IQA-MCNN2 are 0.9529 and 0.9530, which indicates that IQA-MCNN performs much better than other NR IQA algorithms. SROCC of IQA-MCNN2 are all higher than IQA-MCNN1 except for JP2K, which means IQA-MCNN2 performs better than IQA-MCNN1 for most of the distortion types. IQA-MCNN1 and IQA-MCNN2 have larger SROCC values than IQA-SCNN for five distortion types, which means that the proposed algorithm gives great improvement compared with IQA-SCNN. SROCC for WN of IQA-SCNN is 0.3299, which means IQA-SCNN can work well for other four distortion types but not WN. While SROCC of IQA-MCNN2 for WN is 0.9822, which means it works well for WN.

As Table 4 shows, most of the RMSE of TCLT are smaller than other NR IQA algorithms, but note that the results of TCLT scheme is quoted from the literature [22]. It was trained on 23 of 29 reference images and their distorted versions and tested on the rest 6 reference images and their distorted images, which is slightly different from the settings in the paper. Except TCLT, the RMSE of IQA-MCNN2 is 6.5469, which means IQA-MCNN2 performs better than other NR IQA algorithms. RMSE for WN of IQA-SCNN is 18.2276, which means IQA-SCNN can work well for other four distortion types but not WN, which suggests that the kernels learned from five distortion types together performs not well for WN. While RMSE of IQA-MCNN2 for WN is 3.5007, which indicates that distortion-specific expert works much better for WN distortion type than IQA-SCNN.

The scatter plots of subjective quality score DMOS against predictive DMOS (DMOSp) by DIIVINE, OG-IQA, Kang, IQA-SCNN, and the proposed IQA-MCNN1 and IQA-MCNN2 are shown in Fig.6. It can be seen that scatter plots of Fig.6e and Fig.6f are nearly linear and the most compact among all the algorithms compared, which means the predictive scores of our algorithm has a much higher correlation with ground truth. It should be noted that any distortion-specific algorithm can be used as an expert network in our proposed algorithm.

For better observation, we visualize the learned filters in the first convolution layer by distortion-specific expert CNN and IQA-SCNN. The learned filters of expert CNN of FF, GBLUR, JP2K, JPEG, and WN are shown in Fig.7a-e. The learned filters of IQA-SCNN are shown in Fig.7f. Obviously, the learned filters of JP2K (Fig.7c) and JPEG (Fig.7d) are similar with the learned filters of IQA-SCNN (Fig.7f), which indicates IQA-SCNN performs better for JP2K and JPEG than other distortion types. We can also observe that the learned filters of WN (Fig.7e) are the most different from the learned filters of IQA-SCNN (Fig.7f), which demonstrates that compared with other distortion types the IQA-SCNN scheme does not perform well for distortion of WN. Different to IQA-SCNN, a specific expert-CNN more efficiently extracts the special feature which is related to specific distortion. In summary, it is reasonable to design an expert CNN for IQA for different distortion types and the proposed IQA-MCNN is efficient.

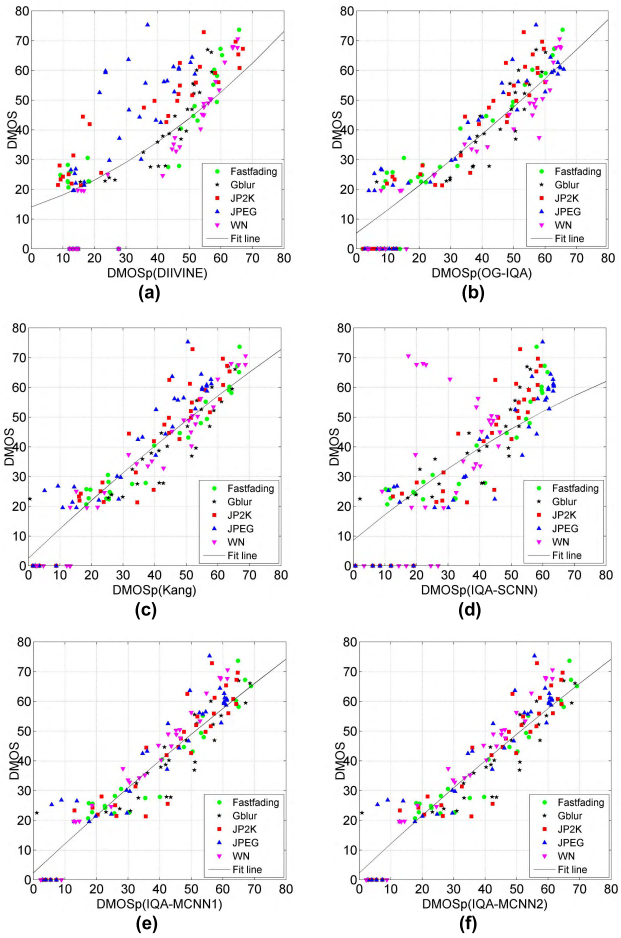

FIGURE 6. Scatter plot of DMOS vs Predicted DMOS (PDMOS) by various algorithms. (a) DIIVINE. (b) OG-IQA. (c) Kang. (d) IQA-SCNN. (e) IQA-MCNN1. (f) IQA-MCNN2.

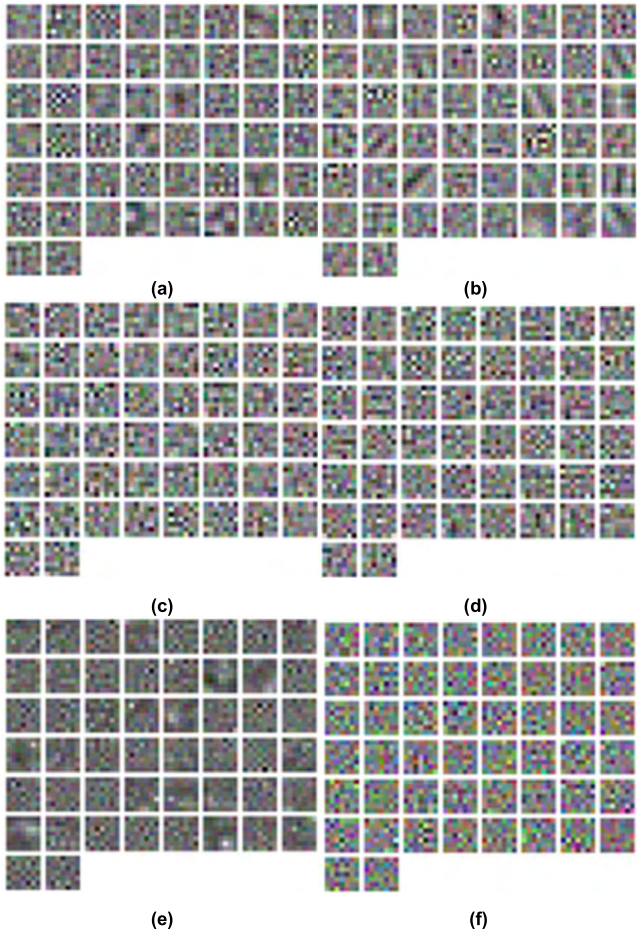

FIGURE 7. Learned kernels by expert networks of IQA-MCNN and IQA-SCNN (a) FF. (b) GBLUR. (c) JP2K. (d) JPEG. (e) WN. (f) IQA-SCNN.

C. CROSS-DATABASE EVALUATION

To verify the independency of our model, we also tested on CSIQ image database. We trained our model on the train set in LIVE release II image database and test on distortion images of the common four distortion types: JPEG, JP2K, additive pink Gaussian noise, and Gaussian blurring in CSIQ database.

Table 5 shows the PLCC comparison between the proposed IQA-MCNN and the benchmark schemes. The best result in each column are marked in bold. It should be noted that the results of TCLT* scheme is TCLT-Gray, which is quoted from the literature [22]. For JPEG, the PLCC of the proposed IQA-MCNN1 and IQA-MCNN2 are 0.9654, which is the best among all the NR-IQA algorithms compared. For all the test distorted images, the PLCC of IQA-MCNN2 is 0.8935, which is slightly lower than OG-IQA. Table 6 shows the SROCC between the proposed IQA-MCNN and the benchmarks. For JP2K, the SROCC of IQA-MCNN2 is 0.8925, which means it performs the best. For JPEG, the SROCC of the IQA-MCNN1 and IQA-MCNN2 is 0.9309, which is slightly lower than the best one. For all the four types and the whole test set, the SROCC of IQA-MCNN2 are all higher than 0.85. Table 7 shows the RMSE between the proposed IQA-MCNN and the benchmarks. For JPEG, the RMSE of the proposed IQA-MCNN1 and IQA-MCNN2 are 0.0798, which is the best among all the NR-IQA algorithms compared. For JP2K, IQA-MCNN2 gives the best performance. For all, the RMSE of IQA-MCNN2 is 0.1269, which is the second among all the

TABLE 5. PLCC trained on LIVE and testing on CSIQ comparison between the proposed IQA-MCNN and the benchmark schemes.

Schemes	JPEG	JP2K	WN	GBLUR	ALL
PSNR	0.8907	0.9468	0.9532	0.9252	0.9218
SSIM	0.9786	0.9694	0.8983	0.9496	0.9269
VIF	0.9884	0.9777	0.9607	0.9793	0.9671
DIIVINE	0.8044	0.8870	0.8802	0.8824	0.8454
OG-IQA	0.9621	0.8920	**0.9097**	0.9086	**0.9108**
TCLT*	-	-	-	-	-
Kang	0.9330	0.8106	0.7613	**0.9105**	0.8110
IQA-MCNN1	**0.9654**	0.8679	0.8593	0.8259	0.8231
IQA-MCNN2	**0.9654**	**0.9151**	0.8590	0.8882	0.8935

TABLE 6. SROCC trained on LIVE and testing on CSIQ comparison between the proposed IQA-MCNN and the benchmark schemes.

Schemes	JPEG	JP2K	WN	GBLUR	ALL
PSNR	0.8879	0.9361	0.9363	0.9291	0.9218
SSIM	0.9543	0.9605	0.8974	0.9608	0.9325
VIF	0.9703	0.9671	0.9575	0.9744	0.9587
DIIVINE	0.7998	0.8304	0.8662	0.8713	0.8284
OG-IQA	**0.9345**	0.8652	**0.8886**	**0.8960**	**0.8841**
TCLT*	0.8900	0.8760	0.8830	0.8950	0.8780
Kang	0.9114	0.7953	0.7534	0.8759	0.7909
IQA-MCNN1	0.9309	0.8495	0.8570	0.8167	0.8325
IQA-MCNN2	0.9309	**0.8925**	0.8538	0.8751	0.8766

TABLE 7. RMSE trained on LIVE and testing on CSIQ comparison between the proposed IQA-MCNN and the benchmark schemes.

Schemes	JPEG	JP2K	WN	GBLUR	ALL
PSNR	0.1391	0.1017	0.0507	0.1087	0.1185
SSIM	0.0629	0.0776	0.0737	0.0898	0.1061
VIF	0.0464	0.0663	0.0466	0.0581	0.0719
DIIVINE	0.1818	0.1459	0.0796	0.1384	0.1509
OG-IQA	0.0835	0.1428	**0.0697**	0.1197	**0.1167**
TCLT*	-	-	-	-	-
Kang	0.1101	0.1850	0.1088	**0.1185**	0.1653
IQA-MCNN1	**0.0798**	0.1570	0.0858	0.1616	0.1605
IQA-MCNN2	**0.0798**	**0.1274**	0.0859	0.1317	0.1269

algorithms compared. On one hand, the proposed algorithm is based on Multi-expert CNN, and deep learning is data driven, which can learn image features automatically from raw images, but it highly relies on data. On the other hand, there exists great difference between LIVE II and CSIQ database, specifically, image resolution, image category, rating method, and DMOS range. The PLCC, SROCC, and RMSE of IQA-MCNN2 are 0.8935, 0.8766, and 0.1269, which is effective and highly correlated with the human visual system. The performance can be improved by training on a large-scale image database.

IV. CONCLUSIONS

In this paper, we presented a general purpose NR IQA-MCNN. We firstly identify image distortion types by a trained CNN and train multi-expert CNN networks for distortion-specific quality predictions, then aggregate the classification result and the predicted qualities of multi-expert networks. The IQA-MCNN performs better than IQA-SCNN, which means the proposed algorithm based on multiple distortion-specific CNNs gives great improvement than IQA-SCNN for all distortion types. Experimental results show that the proposed IQA-MCNN algorithm performs better than state-of-the-art NR IQA algorithms on LIVE II database. Although images in CSIQ database are quite different from LIVE II, the proposed algorithm still has a high correlation with the human visual system. In the future, the performance can be improved by training on large-scale image database.

REFERENCES

[1] Q. Sang, H. Qi, X. Wu, C. Li, and A. Bovik, "No-reference image blur index based on singular value curve," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1625–1630, 2014.
[2] L. Li, Y. Yan, Z. Lu, J. Wu, K. Gu, and S. Wang, "No-reference quality assessment of deblurred images based on natural scene statistics," *IEEE Access*, vol. 5, pp. 2163–2171, Jul. 2017.
[3] J. Zhang, S. H. Ong, and T. M. Le, "Kurtosis-based no-reference quality assessment of JPEG2000 images," *Signal Process., Image Commun.*, vol. 26, no. 1, pp. 13–23, 2011.
[4] Y. Zhan and R. Zhang, "No-reference JPEG image quality assessment based on blockiness and luminance change," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 760–764, Jun. 2017.
[5] G. Yang, Y. Liao, Q. Zhang, D. Li, and W. Yang, "No-Reference quality assessment of noise-distorted images based on frequency mapping," *IEEE Access*, vol. 5, pp. 23146–23156, Oct. 2017.
[6] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.
[7] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[8] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.
[9] X. Xie, Y. Zhang, J. Wu, G. Shi, and W. Dong, "Bag-of-words feature representation for blind image quality assessment with local quantized pattern," *Neurocomputing*, vol. 266, pp. 176–187, May 2017.
[10] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.
[11] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
[12] L. Liu, H. Dong, H. Huang, and A. C. Bovik, "No-reference image quality assessment in curvelet domain," *Signal Process. Image Commun.*, vol. 29, no. 4, pp. 494–505, Apr. 2014.
[13] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. CVPR*, Jun. 2014, pp. 1098–1105.
[14] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.
[15] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. CVPR.*, Jun. 2014, pp. 1733–1740.
[16] H. Wang, L. Zuo, and J. Fu, "Distortion recognition for image quality assessment with convolutional neural network," in *Proc. ICME*, Jul. 2016, pp. 1–6.
[17] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik. (2005). *LIVE Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality
[18] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[19] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[20] A. Moorthy and A. Bovik. (2010). *DIIVINE Software Release*. [Online]. Available: http://live.ece.utexas.edu/research/quality/DIIVINE_release.zip
[21] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2013–2026, Dec. 2013.

[22] Q. Wu *et al.*, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.
[23] A. Mittal, R. Soundararajan, and A. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[24] Q. Wu, H. Li, F. Meng, K. N. Ngan, and S. Zhu, "No reference image quality assessment algorithm via multi-domain structural information and piecewise regression," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 205–216, Aug. 2013.
[25] Y. Zhang, J. Wu, X. Xie, L. Li, and G. Shi, "Blind image quality assessment with improved natural scene statistics model," *Digit. Signal Process.*, vol. 57, pp. 56–65, Oct. 2016.
[26] L. Kang, P. Ye, Y. Li, and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," in *Proc. ICIP*, Quebec City, QC, Canada, Dec. 2015, pp. 2791–2795.
[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1106–1114.
[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[29] J. Li, J. Yan, D. Deng, T. Qu, and G. Xie, "No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 609–616, Apr. 2016.
[30] Y. Liang, J. Wang, X. Wan, Y. Gong, and N. Zheng, "Image quality assessment using similar scene as reference," in *Proc. ECCV*, 2016, pp. 3–18.
[31] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. Multimedia*, 2014, pp. 675–678.
[32] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011006-1–011006-21, 2010.
[33] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Mar. 2006.
[34] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Image Commun.*, vol. 40, pp. 1–15, Jan. 2016.
[35] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. Bovik. (2015). *OG-IQA Software Release*. [Online]. Available: http://live.ece.utexas.edu/research/quality/og-iqa_release.zip

CHUNLING FAN received the M.S. degree from Nanjing Normal University, Nanjing, in 2011. She is currently pursuing the Ph.D. degree with the Shenzhen Institutes of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China. Her research interests include image processing and image quality assessment.

YUN ZHANG (M'12–SM'16) received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China. From 2009 to 2014, he was a Visiting Scholar with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, under the supervision of Prof. Sam Kwong, IEEE Fellow. Since 2010, he has been with the Shenzhen Institutes of Advanced Technology, CAS at Shenzhen, China. He is currently a Professor with the High Performance Computing Center, Shenzhen Institutes of Advanced Technology, CAS at Shenzhen. His research interests include video compression, 3-D video processing, and visual perception.

LIANGBING FENG received the Ph.D. degree in computer science and engineering from Yamaguchi University, Japan, in 2011. He is currently an Associate Professor with the High Performance Computing Center, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests include machine learning and computer vision.

QINGSHAN JIANG received the Ph.D. degree in mathematics from the Chiba Institute of Technology, Japan, in 1996, and the Ph.D. degree in computer science from the University of Sherbrooke, Canada, in 2002. In 1999, he was as a Post-Doctoral Fellow with The Fields Institute for Research in Mathematical Sciences, University of Toronto, Canada. He is currently a Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His research interests include data mining, information security, pattern recognition, massive data analysis, and database technology.