Multimed Tools Appl (2018) 77:29829–29846
https://doi.org/10.1007/s11042-018-5805-z

Blind visual quality assessment for image super-resolution by convolutional neural network

Yuming Fang[1] · Chi Zhang[1] · Wenhan Yang[2] · Jiaying Liu[2] · Zongming Guo[2]

Received: 11 September 2017 / Revised: 29 December 2017 / Accepted: 15 February 2018 /
Published online: 15 March 2018

Abstract Image super-resolution aims to increase the resolution of images with good visual experience. Over the past decades, there have been many image super-resolution algorithms proposed for various multimedia processing applications. However, how to evaluate the visual quality of high-resolution images generated by image super-resolution methods is still challenging. In this paper, a Convolutional Neural Network is designed to predict the visual quality of image super-resolution. The proposed network consists of two convolutional layers, two pooling layers including average, min and max pooling, three fully connected layers and one regression layer. The contribution of the proposed method is twofold. The first one is that we propose a the deep convolutional neural network to extract the high-level intrinsic features more effectively than the hand-crafted features for super-resolution images, which can be used to estimate the image quality accurately. The other is that we divide the super-resolution image into small patches, to consider the local information for the visual quality assessment of super-resolution image as well as increase the number of training data for the deep neural network. Experimental results show that the proposed metric can obtain better performance than other existing ones in visual quality assessment of image super-resolution.

Keywords Visual image quality assessment · Image super-resolution · Deep neural network

✉ Zongming Guo
guozongming@pku.edu.cn

Yuming Fang
fa0001ng@e.ntu.edu.sg

[1] School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China

[2] Institute of Computer Science and Technology, Peking University, Beijing, China

1 Introduction

Image super-resolution (SR) algorithms aim to provide effective solutions to the limitation from some specific imaging sensors such as mobile devices and surveillance cameras. By using image SR, low-resolution (LR) images can be displayed well on high-resolution (HR) displays with good visual experiences for observers. During the past decades, there have been many image SR algorithms proposed for various multimedia processing applications, including medical image processing, face/isis recognition, image editing, etc. [50, 78], retrieval [16, 39], sparse representation [89]. However, how to evaluate the visual quality of SR images effectively is still challenging at present.

The most common and reliable method for visual quality evaluation of SR images is the subjective test, in which participants are invited to provide their ratings for the visual quality of SR images. One problem with the subjective test is that it is time-consuming and expensive. The other problem is that the subjective test requires subjects to be involved and thus it cannot be used in practice. Currently, much less have been done to evaluate the visual quality of SR images objectively. Traditional visual quality assessment methods are mainly designed for distorted images degraded by certain specific distortion types such as Gaussian noise, compression, blurring, and so on. In general, they can be classified into three categories [38]: full-reference (FR) metrics which require complete reference information [64], reduced-reference (RR) metrics which require part of reference information, and no-reference (NR) metrics which do not need any reference information [9, 83]. For image super-resolution, the complete reference information is always unavailable in real applications. Also, the sizes of LR and HR images are different greatly. As a result, most existing RR metrics might not be used for IQA of image super-resolution. Thus, it is highly desired to design effective NR image quality assessment (IQA) metrics to evaluate the visual quality of HR images generated from LR images.

In the past years, there have been many NR-IQA metrics proposed for IQA by using different prior knowledge which are represented as various statistical properties of images [17, 19, 38, 56]. Some NR-IQA metrics are built based on the prior knowledge of specific distortion existing in images, such as blurring [36], compression distortion [23], *etc*. Another commonly used prior knowledge in NR-IQA is the natural scene statistics (NSS) of undistorted images [9, 14, 48, 52]. The third type of prior knowledge used in NR-IQA is the mechanism of the human visual system (HVS), which is derived by visual physiological and psychophysical experiments [30].

Deep neural network has gained researcher's attention and achieved great successes on various computer vision task recently [15, 20, 22, 40, 88]. Unsurprisingly, the convolutional neural network, one of the most representative deep neural network, can also be applied for the super-resolution image quality assessment. In the proposed CNN model, a mixture of max pooling, average pooling and min pooling is employed by the pooling layer. Moreover, batch normalization is applied for the convolutional layers and fully connected layers to prevent the gradient diffusion, and we use the method of Xavier to initialize the weights. Then the network can learn super-resolution image quality features more effectively and estimate the image quality more accurately. One contribution of our work is that we propose a deep convolutional neural network which can extract the high-level intrinsic features more effectively than the hand-crafted features for super-resolution images and can estimate the image quality more accurately. Another contribution of our work is that we divide the super-resolution image into 160*160 patches for the proposed method, which not only increases the number of training data but also considers the local information for the super-resolution image quality. The proposed method can obtain promising performance of quality prediction of SR images, as shown by the experimental results.

2 Related work

2.1 Image super-resolution

Image super-resolution (SR) aims to recover a high resolution (HR) image from one or more low resolution (LR) images. The quality degradations inherent to image acquisition, saving, and storage causes LR images to lose high frequency detail, which leads to image SR recovery being an ill-posed problem. To solve this problem, a *priori* knowledge is imposed. Thus, one important issue of image SR is to constrain SR recovery with proper priors.

Since 1984 [61], studies on image super-resolution have been investigated sequentially. Single image SR can be classified into three categories: interpolation-based, reconstruction-based and example learning-based. Interpolation-based methods [32, 34, 80, 85] utilize the correlation between pixels to construct a prediction function to estimate the missing pixels. Reconstruction-based methods adopt a maximum a *posteriori* probability (MAP) framework in which various regularization terms are imposed as prior knowledge to describe some desirable properties of natural images to constrain the solution of the ill-posed SR recovery problem. Typical regularization terms include gradient [55, 93], nonlocal [6, 27, 44] and total variation (TV) [1, 45]. For both interpolation-based and reconstruction-based methods, prior knowledge is typically achieved in a rather fixed or heuristic way. Thus, it is insufficient to represent the diversified patterns of natural images.

Example-based methods learn the mappings between LR and HR image patches from large training sets. Given an LR patch, its corresponding HR patch is estimated based on these learned mappings. In these methods, prior knowledge is dynamically learned rather than provided heuristically. Thus, the modeling capacity of example-based methods depends largely on the training data source. There are usually two kinds of training data sources: the LR data and external images, further dividing the example-based methods into two subclasses: internal and external SR methods.

Internal SR methods [6, 11, 12, 75, 77, 87, 91], learn priors from a training set cropped from the LR image itself. Based on the self-similarity property (that some salient features repeat across different scales within an image), the coupled LR/HR patches extracted from a hierarchical pyramid of LR images provide an effective prior for building the inverse recovery mapping. In [91], a fast single image super-resolution method is proposed by combining self-example learning and sparse representation. In [6], nonlocal similarity, one important kind of self-similarity, is incorporated into the sparse representation model to constrain and improve the estimation of sparse coefficients. To add more diversified and abundant patterns to the internal dictionary, Huang et al. [21] proposed to expand the internal patch search space by localizing planes with detected perspective geometry variations in the LR image. In these methods, the patch priors are selected and learned from the LR images; thus they are good at reconstructing the repeated patterns in the LR image. However, the internal patch priors fail to cover the diversified patterns of natural images and are poor at reconstructing the distinct patterns. Moreover, the degraded LR image loses high-frequency details, limiting the modeling capacity of internal priors.

In contrast to the internal methods, external methods present complementary and desirable properties. These methods utilize the general redundancy among natural images and learn the LR-HR mappings from large training datasets containing representative coupled external patches from an external dataset. Some external SR methods apply the learned priors to SR estimation directly, without any online auxiliary adaptation, thus they are categorized into *fixed external methods*, including neighbor embedding [4, 35, 57, 58, 79], kernel ridge regression [28], factor graph [71], kernel PCA [3], locality-constrained representation [25], coupled dictionary [18, 65, 74, 76] and the recently proposed deep learning [7, 66, 81, 82]. Compared with the internal methods, when the training set containing a variety of reference images, the priors extracted are more representative and general. However, the fixed prior may not succeed in modeling some image patterns because of the limited numbers of model parameters and training images.

Another branch of methods - *adaptive external methods* adjust the learned prior based on the information in LR images, to make the external prior more adaptive. In [90], the patch prior is modeled as a flexible deformation flow rather than a fixed vector. These deformable patches are more similar to the given LR patch in the LR feature space. Thus, HR patches estimated based on the fusion of these deformable patches present more similar HR features. However, image degradation can make the LR information ambiguous; thus, the deformation estimated in the LR feature space may be imprecise. Rather than adjusting the dictionary or the training set to the LR image, some works perform online compensation, which selects and imports correlated external information to update the training set and models. In [53], an Internet-scale scene matching performs searches for ideal example textures to constrain image upsampling. In [54], with the help of a database containing HR/LR image segment pairs, high-resolution pixels are ”hallucinated” from their texturally similar segments. In [60], the semantic information from parsing is used to choose the corresponding anchor points adaptively to benefit anchor regression-based image SR. In [84], Yue et al. proposed a cloud-based landmark SR method that searches for similar patches in registered and aligned correlated images and utilizes these patches to compensate the lost HR details. In [41], Liu et al. utilized a group-structured sparse representation to further use the nonlocal dependency information of the external HR references.

Due to the obvious strengths and weaknesses of these two kinds of priors, as well as their strong complementary properties, recent works have attempted to utilize both internal and external priors for image denoising and image SR. In [49, 92], the advantages of internal and external denoising methods are measured; then, these two kinds of methods are combined by balancing the error between noise-fitting and signal-fitting. In [2], Burger et al. proposed a learning method to adaptively combine internal and external denoising results. Timofte et al. [59] explored seven ways to benefit image SR, one of which is to create an internal dictionary containing internal anchor points for further joint anchor regression with the external dictionary. Wang et al. [67] proposed a joint SR method to adaptively fuse the results of sparse coding for external examples and those of epitomic matching for internal examples.

2.2 Image quality assessment

As indicated previously, IQA methods can be classified into FR [10], RR and NR methods [9] and can be used for image retargeting [8]. Since we design a NR-IQA metric for image super-resolution in this study, we only review existing NR-IQA methods here. Generally, NR-IQA methods assume that the statistics of the distorted images are different from those of the original images [62]. NR-IQA methods can be designed based on NSS models built in both spatial and transform domains.

In the study [31], the author proposed a NR-IQA metric to evaluate the visual quality of image blurring based on edge spread in the spatial domain. Xue et al. modeled the image gradient and Laplacian of Gaussian operators jointly for statistical naturalness destruction in images [73]. A general purpose NR-IQA metric was designed by Mittal et al. based on the distribution of locally normalized luminance and products of locally normalized luminance [46]. In the study [9], the authors built NSS models of entropy and intensity for visual quality assessment of contrast-distorted images. The features of discrete orthogonal moments are extracted in the spatial domain for visual quality assessment of image blur [36]. Wu et al. proposed a blind image quality assessment (BIQA) algorithm which is characterized by a feature fusion scheme and k-nearest-neighbor-based quality prediction model [68]. Ma et al. established a large-scale database named the Waterloo Exploration Database and present three alternative test criteria to evaluate the performance of IQA models [43]. Wu et al. proposed a BIQA method that introduce a pairwise rank-order constraint into the maximum margin regression framework [70]. A NR-IQA model was proposed by developing a local image representation which extracts the structural image information from both the spatial-frequency and spatial domains [69].

Besides NR-IQA metrics in the spatial domain, there have also been various NR-IQA methods. No-reference (NR) IQA measures try to estimate the human perceptual quality be extracting discriminative features from distorted images. In general, current NR IQA algorithms mainly can be divided into two trends. The traditional methods [17, 56] design features based on Natural Scene Statistic (NSS) which based approaches process image with certain type of filters and then the responses are used to extract features. Some typical domain and filters include DCT domain [52] and Wavelet domain [48]. Another trend is based on machine learning or deep learning technique [51]. Kang et al. [26] propose a Convolution Neural Network (CNN) to predict image quality without a reference image, feature extracting and regression are integrated into one optimization process within the network structure, their network extracts discriminative features from 32*32 patches with as a single convolutional layer and a pooling layer, and then estimates image quality score of each patch, finally they average the scores of all the 32*32 patches to obtain a quality estimation for the whole image. Chen et al. [5] focus on the relative quality ranking between enhanced image rather than giving an absolute quality score for a single enhanced image, the rank function is trained to fit the subjective assessment results, and can be used to predict ranks of new image which indicate the relative quality of enhancement algorithms. Li et al. [33] applied a general regression neural network that takes as input perceptual features including phase congruency, entropy and the image gradients. And Chetouani et al. used a neural network to combine multiple distortion-specific NR-IQA measures. Most existing methods regard the image quality assessment as a classification problem, and require pre-extracted handcrafted features and only use the neural network for learning the regression function. In contrast, the proposed method does not require any handcrafted features and directly learns intrinsic features from the deep neural network to get much better performance of visual quality assessment of super-resolution images.

3 The proposed CNN for SR-IQA

3.1 Network architecture

The proposed deep convolutional neural network consists of five layers, as shown in Fig. 1. The detailed configurations of the proposed network structure are shown in Table 1. For these two convolutional layers: the ”filter” parameter specifies the number and the size of convolutional kernels is set as num*size*size; the ”st.” and the ”pad” parameters specify the

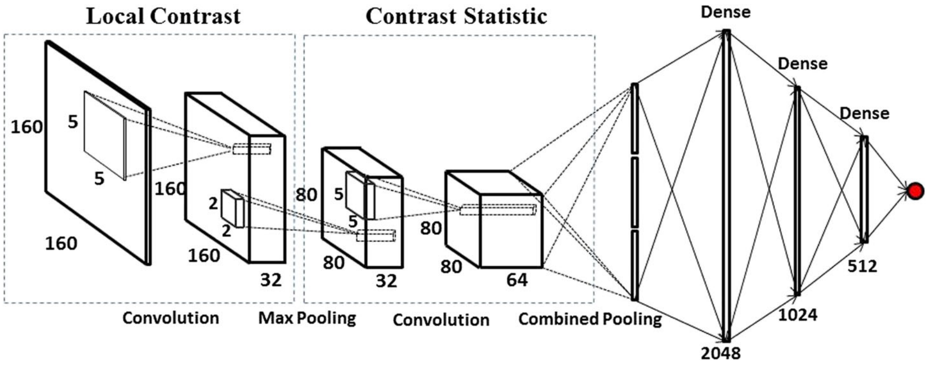

Fig. 1 The architecture of the proposed deep neural network

convolution stride and spatial padding respectively; the ”pool” indicates whether to apply the max-pooling window size by size*size. For these three full connected layers, we specify their dimensionality as 2048, 1024 and 520. The ”dropout” indicates whether the fully-connected layer is regularized by dropout to prevent overfitting, and the output layer has one dimension which indicates the predicted quality score.

3.2 Convolution

In the convolutional layers, the input image patches are convolved with 32 filters and each filter generates a feature map by Rectified Linear Units (ReLU). In these layers, the kth output feature map m_k can be calculated as follows:

$$m_k = g(\omega_k * n) \tag{1}$$

where n denotes the input image patches with the size 160*160; w_k stands for the convolutional filter associated with the kth feature map; * indicates the 2D convolution operation; g denotes the activation function. We use the Rectified Linear Units (ReLUs) instead of sigmoid or tanh activation function in the full connected layers. Formally ReLUs can be expressed as $g = max(0, \sum_i \omega_i \alpha_i)$, where g represents the output of ReLU, ω_i denotes the weight of ReLU, and α_i is the input of the previous layer. The study [29] demonstrates that ReLUs are robust to the input range and enable the network to train several times faster compared to using tanh units in a deep convolutional neural network. We provide the feature maps generate by the convolutional layers as shown in Fig. 2. The proposed DCNN method can extract high-level intrinsic features more effectively compared with the hand-crafted

Table 1 Detailed network configurations of our proposed for high-resolution image quality assessment

Layers	Configuration
conv1	filter 32*5*5, st. 1*1, pad 0, pooling 2*2
conv2	filter 32*5*5, st. 1*1, pad 0, pooling 2*2
fc3	2048, Relu, dropout
fc4	1024, Relu, dropout
fc5	512, Relu, dropout
fc6	1-dimensional

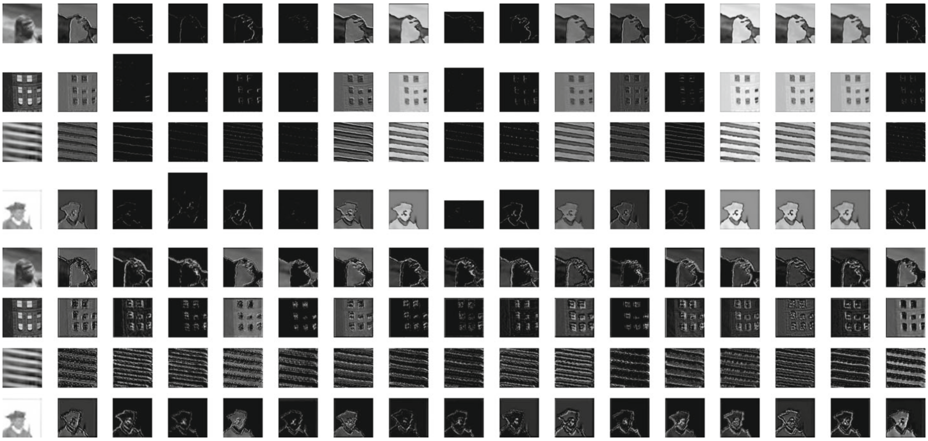

Fig. 2 Row1-Row4: Column 1 are the original super-resolution image, and the rest are the feature maps of the first convolutional layer; Row5-Row8: Column 1 are the same original super-resolution image, and the rest are the feature maps of the second convolutional layer

features used in existing related methods. As shown in Fig. 2, the feature map calculated by forward propagation can be used to represent the local contrast information and the edge information in the super-resolution image which are essential to predict the visual quality for super-resolution images. Additionally, we use some tricks during DCNN training processing which are important to predict the perceptual scores of super-resolution images. For instance, the combined pooling operation including max, min and mean can get more effective information than a single pooling operation. Also, we divide the super-resolution image into 160*160 patches in the proposed method, which not only increases the number of training data but also considers the local information for the super-resolution image quality.

3.3 Xavier initialization

Inspire by [13], we use Xavier to initialize the network for forward propagation and prevent the top hidden layer into saturation. We adopt a properly scaled uniform of Gaussian distribution for initialization.

$$Var(W) = \frac{1}{n_{in} + n_{out}} \tag{2}$$

where W is the initialization distribution for the neuron; n_{in} is the number of neurons feeding into it, n_{out} is the number of neurons where the result is fed to.

3.4 Batch normalization

Batch normalization is applied to accelerate deep network training by reducing internal covariate shift [24]. For a layer with d-dimensional input $x = (x^{(1)}...x^{(2)})$, we will normalize each dimension as follows:

$$\hat{x}^{(k)} = \frac{x^{(k)-E(x^{(k)})}}{\sqrt{Var(x^{(k)})}} \tag{3}$$

where the expectation and variance are computed over the training dataset. Note that simply normalizing each input of a layer may change what the layer can represent. To address this problem, we introduce, for each activation $x^{(k)}$, a pair of parameters $\gamma^{(k)}$, $\beta^{(k)}$, which scale and shift the normalized value:

$$y^{(k)} = \gamma^{(k)}\hat{x}^{(k)} + \beta^{(k)} \tag{4}$$

where the parameters can learned along with the original model parameters.

3.5 Pooling

In the two convolution layers, the input image patches are convolved with many filters and each filter generates a feature map, then we apply pooling operation on each feature map to reduce it to a low dimension. Specifically, the feature map of the last convolution layers is pooled into one max value, min value and mean value as follows:

$$p1_k = max_{i,j} R^k_{i,j} \tag{5}$$

$$p2_k = min_{i,j} R^k_{i,j} \tag{6}$$

$$p3_k = mean_{i,j} R^k_{i,j} \tag{7}$$

where $m \in \{1, 2, 3, 4, ..., M\}$, M is the number of kernels. $N(i, j)^m$ denotes the response at location (i, j) of the feature map. Therefore, each node of the next fully connected layer takes an input of size $3 * M$. Pooling operation is typically performed on every $2 * 2$ cell in the case and it can keep some location information and the image intrinsic structure while achieving robustness to translation.

3.6 Learning process

We train our network on non-overlapping 160*160 patches taken from super-resolution images and their perceptual scores from 50 subjects(the mean of the median 40 subject scores is used as perceptual score). By taking small patches as input and the geometric transform for the original image, we have a much larger number of training samples. Formally, We provide a set of super-resolution image patches and the label scores for the convolutional neural network. The training objective function of the network is to minimize the Minimum Squared-Error:

$$c = \frac{1}{N}\sum_{i=1}^{N} ||y_i - y_i^{'}||^2 \tag{8}$$

$$\delta' = \min_{\delta} c \tag{9}$$

where y_i represents the label of the input super-resolution image patch, $y_i \in (y_1, y_2, ..., y_N)$, N is the total numbers of the input data. $y_i^{'}$ denotes the score computed through the deep convolutional neural network. The optimizer we used for minimizing the loss function is Adaptive Moment Estimation which is a method for Stochastic Optimization, Adam optimizer adjust the learning rate for each parameters dynamical by using the estimation for moments of order 1 and 2 of the gradients. In our experiment we boost the performance by using the dropout which is a technique that prevents overfitting in training neural network and an efficient approximation of training many different networks with shared weights by masking out the neurons randomly. Specifically, the outputs of neurons are set to zero with a probability of 0.5 in the training stage and divided by 2 in the test stage. Unlike L1 and L2 regularization, dropout doesn't rely on modifying the cost function.

4 Experimental results

In this section, we provide the experimental results for the performance evaluation of the proposed method. The evaluation methodology for the comparison experiments, including the used database, evaluation methods, and existing IQA metrics are first introduced. Then the comparison results are given to demonstrate the performance of the proposed method. We conduct the comparison experiments based on the database including 1440 HR images generated from 180 LR images using 8 different image super-resolution algorithms [42]. In that database, they conducted an user study by collecting subjective scores from 50 participants (the mean of the median 40 subject scores is used as the MOS (Mean Opinion Score)). For the proposed deep convolutional neural network, the weights of the convolutional layers are initialized from zero mean Gaussian with a standard deviation of 0.01 and the bias is set to 0. The fully-connected layers are initialized by Xavier Initialization. The proposed network is trained on the dataset whose images are divided into small patches (160*160) with 40K iterations for super-resolution image quality assessment. The learning rate is set to 10^{-5}.

4.1 Evaluation

We evaluate the performance of the proposed method by the correlation between subjective scores and objective scores predicted by IQA metrics. In this study, we use three commonly used methods to calculate the correlation between the subjective and objective scores: Spearman's Rank-order Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC) and Kendall Rank Correlation Coefficient (KRCC). PLCC is used to measure the linear dependence between subjective scores and the predicted quality. Given the $i\text{-}th$ image in the database with size N, its subjective and objective scores are s_i and o_i, respectively. We use a nonlinear function o_i to map the raw predicted scores to MOS scale as follows [62]:

$$o_i^{'} = \beta_1[\frac{1}{2} - \frac{1}{1 + exp(\beta_2(o_i - \beta_3))}] + \beta_4 o_i + \beta_5 \tag{10}$$

where β_1, β_2, β_3, β_4, β_5 are parameters. PLCC can be calculated as follows:

$$PLCC = \frac{\sum_i(o_i^{'} - \overline{o^{'}})(s_i - \overline{s})}{\sqrt{\sum_i(o_i^{'} - \overline{o^{'}})\sum_i(s_i - \overline{s})}} \tag{11}$$

SRCC and KRCC measure the strength of association between the predicted scores and subjective scores in the aspect of monotonic relationship. They can be computed as follows:

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N(N^2 - 1)} \tag{12}$$

where e_i is the difference between the $i\text{-}th$ image's ranks in subjective and objective results.

$$KRCC = 1 - \frac{N_c - N_d}{\frac{1}{2}N(N-1)} \tag{13}$$

where N_c and N_d denote the numbers of concordant and discordant pairs in the dataset, respectively.

To demonstrate the performance of the proposed method, we have conducted the comparison experiments by using some existing IQA metrics including four state-of-the-art no-reference methods (BRISQUE [46], ILNIQE [86], NIQE [47], QAC [72], NRSL [37]) and three full-reference metrics (PSNR, SSIM [64], MSSSIM(multi-scale SSIM) [63]). The source code of these metrics were obtained from the corresponding authors or their public Websites. In Fig. 3, we provide the scatter plots of objective quality scores against subjective scores for different IQA metrics. From these scatter plots, we can observe that the points by the proposed method are more centralized than other existing NR-IQA and FR-IQA methods, which demonstrates that the objective quality scores obtained by the proposed method are more consistent with the subjective scores. The experimental results of PLCC, SRCC, RMSE and KRCC values are shown in Table 2. From Table 2, we can observe that: (1) the proposed method can get higher values of PLCC, SRCC and KRCC than other existing methods; (2) the proposed method can obtain lower RMSE value than other existing methods. These results denote that the scores predicted by the proposed method can obtain higher correlation with subjective data than those from other existing IQA methods.

Overall, the proposed metric can get perfect performance against the state-of-the-art methods, which can be attributed to: most existing NR-IQA methods include two stages of feature extraction and model learning for visual quality prediction. However, the features used in existing NR-IQA approaches are extracted for specific distortions, such as blurring, compression artifact, and so on. However, the visual distortion generated from image

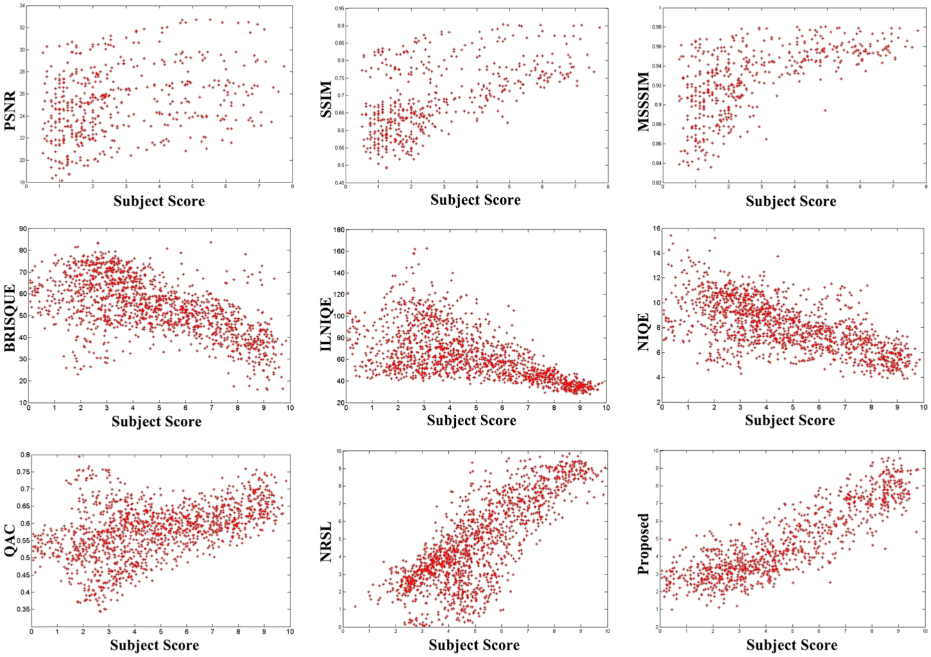

Fig. 3 Quality indices generated by different methods to perceptual scores

Table 2 Performance evaluation of the proposed method

Models	PSNR	SSIM	MSSSIM	BRISQUE	ILNIQE	NIQE	QAC	NRSL	Ours
PLCC	0.3335	0.5726	0.6218	0.6176	0.6198	0.6364	0.4704	0.8011	0.9156
SRCC	0.3110	0.5562	0.6452	0.5723	0.6282	0.6254	0.4938	0.6874	0.8394
RMSE	2.9383	1.798	1.0272	10.0747	18.3748	1.5582	0.0695	0.9149	1.2527
KRCC	0.2181	0.4012	0.4666	0.4181	0.4557	0.4573	0.3455	0.5112	0.6435

super-resolution is strongly correlated with the image content and there might be several different distortions in HR images. The features used in existing NR-IQA approaches cannot cover various visual content, and thus, traditional NR-IQA methods cannot be used for IQA of image super-resolution. And the deep convolutional neural network can get the high-level intrinsic feature to predict the scores for the super-resolution image.

4.2 Effects of parameters

Several parameters are defined in the proposed deep neural network. In this section, we will analysis the effects of parameters to the experiment performance on the dataset. There are some advantages that we divide the whole super-resolution images into small image patches. Not only it can extremely increase our training and testing data but also can learn and prediction of image quality on local image regions. Local quality calculation is important for the image denoising or reconstruction problems, applying enhancement only where required. The problem we face is that which size of the image patch is the best for the experiment results. Extensive experiments have been conduct for which size we choose to the super-resolution image patches, from the Fig. 4 which provides the relationship between

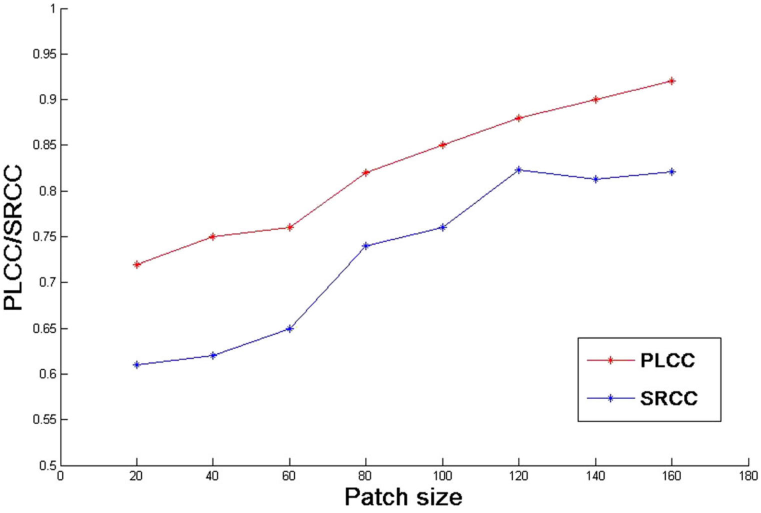

Fig. 4 PLCC and SROCC under different on different image patch

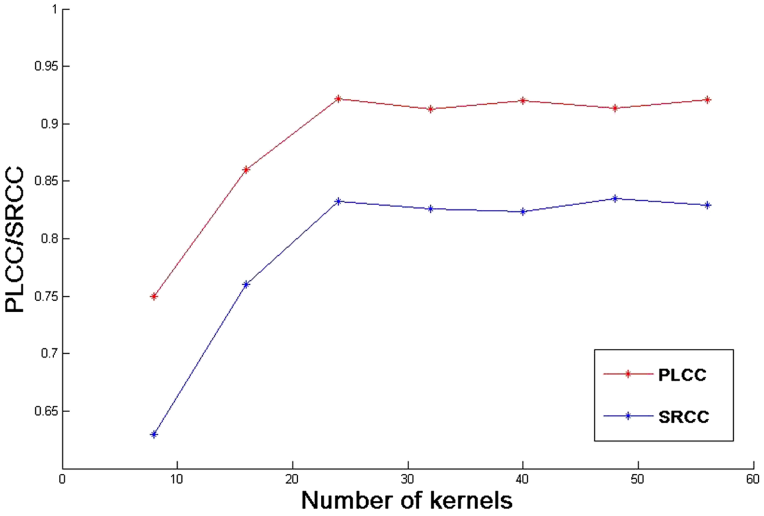

Fig. 5 PLCC and SROCC under different number of kernels

PLCC, SRCC values and the size of super-resolution image patches, it is obvious that we can get better performance with larger patches and get the best results when the patch size is 160*160.

In the convolutional layers, different size of convolve kernel may lead different performance. Convolutional neural network further achieve to imitate human 's perceptual characteristics compare to normal neural network. The observation of human for the natural scenes are from local to global, it's not based on every pixel but a local region information which can extract global feature by integrating themselves. The kernel size which control the size of perceptual region is crucial in the convolutional neural network, and Table 3 shows the performance with different kernel size in the proposed framework. It is clearly that the size of 5*5 for the kernel can get the better performance in the experiment.

Another effects of parameter is the number of convolutional kernels. The convolutional neural network can get better performance with a larger number of kernels on the condition that we have big enough training and testing data to conduct the experiment. And more convolutional kernels bring extensive training parameters and have higher cost on the computation. How the performance varies with the number of convolutional kernels as shown in Fig. 5. From the Fig. 5, it can be seen that the number of kernels influence the experiment performance significantly. As the number of filters increase, the experiment performance have been improved accordingly. But little performance boost is required when the number of filters exceeds 24.

Table 3 Performance evaluation of the proposed method

Size	3x3	5x5	7x7	9x9	11x11
PLCC	0.9012	0.9156	0.9149	0.8856	0.8712
SRCC	0.8034	0.8394	0.8325	0.8123	0.7921

5 Conclusion

In this study, we build an effective NR-IQA metric based on deep convolutional neural network for super-resolution images. Our algorithm combines high-level intrinsic feature extracting and regression as a complete optimization process, which enables us to apply modern training trick to obtain favorable neural network. We divide the whole super-resolution image into small image patches. It can not only increase the training and testing data but also can predict and prediction of image quality on local image regions. The proposed method for super-resolution image quality assessment by using convolutional neural network can get better performance compare to the other existing state-of-the-art IQA methods.

Acknowledgments This work was partially funded by the Natural Science Foundation of China under Grant 61571212, and by Natural Science Foundation of Jiangxi Province in China under Grant 20071BBE50068, 20171BCB23048, 20161ACB21014 and Grant GJJ160420.

References

1. Aly HA, Dubois E (2005) Image up-sampling using total-variation regularization with a new observation model. IEEE Trans Image Process 14(10):1647–1659
2. Burger HC, Schuler C, Harmeling S (2013) Learning How to Combine Internal and External Denoising Methods. Pattern Recognition 24(11):121–130
3. Chakrabarti A, Rajagopalan AN, Chellappa R (2007) Super-resolution of face images using kernel PCA-based prior. IEEE Trans Multimed 9(4):888–892
4. Chang H, Yeung DY, Xiong Y (2004) Super-resolution through neighbor embedding. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 275–282
5. Chen Z, Jiang T, Tian Y (2014) Quality assessment for comparing image enhancement algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 3003–3010
6. Dong W, Zhang L, Shi G (2013) Nonlocally centralized sparse representation for image restoration. IEEE Trans Image Process 22(4):1620–30
7. Dong C, Chen CL, He K (2014) Learning a Deep Convolutional Network for Image Super-Resolution. In: European Conference on Computer Vision, pp 184–199
8. Fang Y, Zeng K, Wang Z, Lin W, Fang Z, Lin CW (2014) Objective quality assessment for image retargeting based on structural similarity. IEEE J Emerging Sel Top Circ Syst 4(1):95–105
9. Fang Y, Ma K, Wang Z, Lin W, Fang Z, Zhai G (2015) No-reference quality assessment of contrast-distorted images based on natural scene statistics. IEEE Signal Process Lett 22(7):838–842
10. Fang Y, Yan J, Liu J, Wang S, Li Q, Guo Z (2017) Objective quality assessment of screen content images by uncertainty weighting. IEEE Trans Image Process 26(4):2016–2027
11. Freedman G, Fattal R (2010) Image and video upscaling from local self-examples. Acm Trans Graph 30(2):474–484
12. Glasner D, Bagon S, Irani M (2009) Super-resolution from a single image. In: IEEE International Conference on Computer Vision, pp 349–356
13. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. J Mach Learn Res 9:249–256
14. Gu K, Wang S, Zhai G, Ma S, Yang X, Lin W, Zhang W, Gao W (2016) Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure. IEEE Transactions on Multimedia 18(3):432–443
15. Guo Y, Ding G, Han J et al (2017) Zero-shot learning with transferred samples. IEEE Trans Image Process 26(7):3277–3290
16. Guo Y, Ding G, Liu L et al (2017) Learning to hash with optimized anchor embedding for scalable retrieval. IEEE Trans Image Process PP(99):1–1
17. He L, Tao D, Li X, Gao X (2012) Sparse representation for blind image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1146–1153

18. He L, Qi H, Zaretzki R (2013) Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. IEEE Conf Comput Vis Pattern Recogn 9(4):345–352
19. Hou W, Gao X, Tao D (2015) Blind image quality assessment via deep learning. IEEE Trans Neural Netw Learn Syst 26(6):1275–1286
20. Huang W (2016) A novel disease severity prediction scheme via big pair-wise ranking and learning techniques using image-based personal clinical data. Signal Process 124:233–245
21. Huang J, Singh A, Ahuja N (2015) Single image super-resolution from transformed self-exemplars. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 5197–5206
22. Huang W, Ding H, Chen G (2018) A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance. Signal Process 142:104–113
23. Ichigaya A, Nishida Y, Nakasu E (2008) Non reference method for estimating PSNR of MPEG-2 coded video by using DCT coefficients and picture energy. IEEE Trans Circ Syst Video Technol 18(6):817–826
24. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift, Computer Science
25. Jiang J, Hu R, Wang Z (2014) Noise robust face hallucination via locality-constrained representation. IEEE Trans Multimed 16(5):1268–1281
26. Kang L, Ye P, Li Y, Doermann D (2014) Convolution neural network for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1733–1740
27. Katkovnik V, Foiand A, Egiazarian K, Astola J (2010) From local kernel to nonlocal multiple-model image denoising. Int J Comput Vis 86(1):1–32
28. Kim KI, Kwon Y (2010) Single-image super-resolution using sparse regression and natural image prior. IEEE Trans Pattern Anal Mach Intell 32(6):1127–1133
29. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Conference on Neural Information Processing Systems
30. Legge GE, Foley JM (1980) Contrast masking in human vision. Journal of the Optical Society of America 70(12):1458–1471
31. Li X (2002) Blind image quality assessment. In: IEEE International Conference on Image Processing
32. Li X, Orchard MT (2001) New edge-directed interpolation. IEEE Transactions on Image Processing 10(10):1521–7
33. Li C, Bovik A, Wu X (2011) Blind image quality assessment using a general regression neural network. IEEE Trans Neural Netw 22(5):793–799
34. Li M, Liu J, Ren J, Guo Z (2015) Adaptive general scale interpolation based on weighted autoregressive models. IEEE Trans Circ Syst Video Technol 25(2):200–211
35. Li Y, Liu J, Yang W, Guo Z (2015) Neighborhood regression for edge-preserving image super-resolution. In: IEEE International Conference on Acoustics, Speech and Signal Processing
36. Li L, Lin W, Wang X, Yang G, Bahrami K, Kot AC (2016) No-reference image blur assessment based on discrete orthogonal moments. IEEE Trans Cybern 46(1):39–50
37. Li Q, Lin W, Xu J, Fang Y (2016) Blind image quality assessment using statistical structural and luminance features. Transactions on Multimedia 18(12):2457–2469
38. Lin W, Jay Kuo C-C (2011) Perceptual visual quality metrics: a survey. J Vis Commun Image Represent 22(4):297–312
39. Lin Z, Ding G, Han J et al (2016) Cross-view retrieval via probability-based semantics-preserving hashing. IEEE Trans Cybern PP(99):1–14
40. Lin Z, Ding G, Han J, Shao L (2017) End-to-end feature-aware label space encoding for multilabel classification with many classes. IEEE Trans Neural Netw Learn Syst 99:1–16
41. Liu J, Yang W, Zhang X, Guo Z (2017) Retrieval compensated group structured sparsity for image super-resolution. IEEE Trans Multimed 19(2):302–316
42. Ma C, Yang C, Yang X, Yang M (2016) Learning a No-reference Quality Metric for Single-Image Super-Resolution. The Korea-Japan joint workshop on Frontiers of Computer Vision
43. Ma K, Duanmu Z, Wu Q et al (2017) Waterloo exploration database: new challenges for image quality assessment models. IEEE Trans Image Process 26(2):1004–1016
44. Mairal J, Bach F, Ponce J (2009) Non-local sparse models for image restoration. IEEE Int Conf Comput Vis 30(2):2272–2279
45. Marquina A, Osher SJ (2008) Image super-resolution by TV-regularization and bregman iteration. J Sci Comput 37(3):367–382
46. Mittal A, Moorthy AK, Bovik A (2012) No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 24(12):4695–4708
47. Mittal A, Soundararajan R, Bovik A (2013) Making a ’completely blind’ image quality analyzer. IEEE Signal Process Lett 20(3):209–212

48. Moorthy AK, Bovik A (2011) Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Trans Image Process 20(12):3350–3364
49. Mosseri I, Zontak M, Irani M (2013) Combining the power of Internal and External denoising. IEEE International Conference on Computational Photography 8772(18):1–9
50. Nasrollahi K, Moeslund TB (2014) Super-resolution: a comprehensive survey. Mach Vis Appl 25(6):1423–1468
51. Pei S, Chen L (2015) Image quality assessment using human visual DOG model fused with random forest. IEEE Trans Image Process 24(11):3282–3892
52. Saad MA, Bovik A, Charrier C (2012) Blind image quality assessment: A natural scene statistics approach in the dct domain. IEEE Trans Image Process 21(8):3339–3352
53. Sun L, Hays J (2012) Super-resolution from internet-scale scene matching. In: IEEE International Conference on Computational Photography, pp 1–12
54. Sun J, Zhu J, Tappen MF (2010) Context-constrained hallucination for image super-resolution. IEEE Conf Comput Vis Pattern Recogn 26(2):231–238
55. Sun J, Sun J, Xu Z, Shum HY (2011) Gradient profile prior and its applications in image super-resolution and enhancement. IEEE Transactions on Image Processing
56. Tang H, Joshi N, Kapoor A (2011) Learning a blind measure of perceptual image quality. In: IEEE Conference on Computer Vision and Pattern Recognition
57. Timofte R, De V, Gool LV (2013) Anchored neighborhood regression for fast example-based super-resolution. In: IEEE International Conference on Computer Vision, pp 1920–1927
58. Timofte R, Smet VD, Gool LV (2014) A+: adjusted anchored neighborhood regression for fast super-resolution. In: Asian Conference on Computer Vision, pp 111-126
59. Timofte R, Rothe R, Gool LV (2015) Seven ways to improve example-based single image super resolution. Computer Science
60. Timofte R, Smet VD, Gool LV (2016) Semantic super-resolution: When and where is it useful? Comput Vis Image Underst 142:1–12
61. Tsai RY, Huang TS (1984) Multipleframe image restoration and registration. In: Advances in Computer Vision and Image Processing
62. Wang Z, Li Q (2011) Information content weighting for perceptual image quality assessment. IEEE Trans Image Process 20(5):1185–1198
63. Wang Z, Simoncelli E, Bovik A (2003) Multi-scale structual similarity for image quality assessment. In: IEEE Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers
64. Wang Z, Bovik A, Sheikh HR, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
65. Wang S, Zhang L, Liang Y (2012) Semi-Coupled Dictionary Learning with Applications to Image Super-Resolution and Photo-Sketch Synthesis. IEEE Conf Comput Vis Pattern Recogn 157(10):2216–2223
66. Wang Z, Liu D, Yang J (2015) Deep networks for image super-resolution with sparse prior. In: IEEE International Conference on Computer Vision, pp 370–378
67. Wang Z, Yang Y, Wang Z (2015) Learning super-resolution jointly from external and internal examples. IEEE Trans Image Process 24(11):4359–71
68. Wu Q, Li H, Meng F et al (2016) Blind image quality assessment based on multichannel feature fusion and label transfer. IEEE Trans Circ Syst Video Technol 26(3):425–440
69. Wu Q, Li H, Meng F et al (2016) No reference image quality assessment metric via multi-domain structural information and piecewise regression. J Vis Commun Image Represent 32(C):205–216
70. Wu Q, Li H, Wang Z et al (2017) Blind image quality assessment based on rank-order regularized regression. IEEE Trans Multimed PP(99):1–1
71. Xiong Z, Xu D, Sun X (2013) Example-based super-resolution with soft information and decision. IEEE Trans Multimed 15(6):1458–1465
72. Xue W, Zhang L, Mou X (2013) Learning without Human Scores for Blind Image Quality Assessment. In: IEEE Conference on Computer Vision and Pattern Recognition
73. Xue W, Mou X, Zhang L, Bovik A, Feng X (2014) Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. IEEE Trans Image Process 23(11):4850–4862
74. Yang J, Wright J, Huang TS (2010) Image super-resolution via sparse representation. IEEE Trans Image Process 19(11):2861–2873
75. Yang C, Huang J, Yang M (2011) Exploiting self-similarities for single frame super-resolution. In: Asian Conference on Computer Vision, pp 497–510
76. Yang J, Wang Z, Lin Z (2012) Coupled dictionary training for image super-resolution. IEEE Trans Image Process 21(8):3467–78
77. Yang M, Wang Y (2013) A self-learning approach to single image super-resolution. IEEE Trans Multimed 15(3):498–508

78. Yang CMC-Y, Yang M (2014) Single-image super-resolution: a benchmark. In: European Conference on Computer Vision
79. Yang S, Liu J, Fang Y, Guo Z (2016) Joint-feature guided depth map super-resolution with face priors. IEEE Transactions on Cybernetics
80. Yang W, Liu J, Li M, Guo Z (2016) Isophote-constrained autoregressive model with adaptive window extension for image interpolation, IEEE Transactions on Circuit System for Video Technology
81. Yang W, Deng S, Hu Y, Xing J, Liu J (2017) Real-Time Deep Video SpaTial Resolution UpConversion SysTem (STRUCT++ Demo). In: ACM Multimedia
82. Yang W, Feng J, Yang J, Zhao F, Liu J, Guo Z, Yan S (2017) Deep Edge Guided Recurrent Residual Learning for Image Super-Resolution. IEEE Transaction on Image Processing
83. Ye P, Kumar J, Kang L, Doermann DS (2012) Unsupervised feature learning framework for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition
84. Yue H, Sun X, Yang J (2013) Landmark image super-resolution by retrieving web images. IEEE Trans Image Process 22(12):4865–4878
85. Zhang L, Wu X (2006) An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans Image Process 15(8):2226–2238
86. Zhang L, Zhang L, Bovik A (2015) A feature-enriched completely blind image quality evaluator. IEEE Trans Image Process 24(8):2579–2591
87. Zhang Y, Liu J, Yang W, Guo Z (2015) Image super-resolution based on structure-modulated sparse representation. IEEE Trans Image Process 24(9):2797–2810
88. Zhang P, Zhuo T, Huang W, Chen K, Kankanhalli M (2017) Online object tracking based on CNN with spatial-temporal saliency guided sampling. Neurocomputing 257:115–127
89. Zhang Q, Liu Y, Blum S et al (2017) Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review. Information Fusion
90. Zhu Y, Zhang Y, Yuille AL (2014) Single image super-resolution using deformable patches. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 2917–2924
91. Zhu Z, Guo F, Yu H (2014) Fast single image super-resolution via self-example learning and sparse representation. IEEE Trans Multimed 16(8):2178–2190
92. Zontak M, Irani M (2011) Internal statistics of a single natural image. IEEE Conf Comput Vis Pattern Recogn 2(7):977–984
93. Zuo W, Zhang L, Song C, Zhang D (2013) Texture enhanced image denoising via gradient histogram preservation. In: IEEE Conference on Computer Vision and Pattern Recognition

Yuming Fang received his Ph.D. degree from Nanyang Technological University in Singapore, M.S. degree from Beijing University of Technology in Beijing, China, and B.E. degree from Sichuan University in Chengdu, China. Currently, he is a Professor in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He serves as an Associate Editor of IEEE Access and is on the editorial board of Signal Processing : Image Communication. He have authored and co-authored more than 90 academic papers in international journals and conferences in the areas of multimedia processing. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, 3D image/video processing, *etc.*

Chi Zhang is currently a graduate student in the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include saliency detection, computer vision, and machine learning.

Wenhan Yang received the B.S degree in computer science from Peking University, Beijing, China, in 2012, and is currently working toward the Ph.D. degree at the Institute of Computer Science and Technology, Peking University. He was a Visiting Scholar with the National University of Singapore, Singapore, from 2015 to 2016. His current research interests include image processing, sparse representation, image restoration and deep learning-based image processing.

Jiaying Liu received the B.E. degree in computer science from Northwestern Polytechnic University, Xi'an, China, and the Ph.D. degree with the Best Graduate Honor in computer science from Peking University, Beijing, China, in 2005 and 2010, respectively. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored over 80 technical articles in refereed journals and proceedings, and holds 15 granted patents. Her current research interests include image/video processing, compression, and computer vision. Dr. Liu was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She was a Visiting Researcher at Microsoft Research Asia (MSRA) in 2015 supported by "Star Track for Young Faculties". She has also served as TC member in APSIPA IVM since 2015, and APSIPA distinguished lecture in 2016-2017.

Zongming Guo received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively. He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication. Dr. Guo is the Executive Member of the China-Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008, the Government Allowance granted by the State Council in 2009, and the Distinguished Doctoral Dissertation Advisor Award of Peking University in 2012 and 2013.