



# Blind Image Quality Assessment via Deep Recursive Convolutional Network with Skip Connection

Qingsen Yan<sup>1,3</sup>, Jinqiu Sun<sup>2(✉)</sup>, Shaolin Su<sup>1</sup>, Yu Zhu<sup>1</sup>, Haisen Li<sup>1</sup>,  
and Yanning Zhang<sup>1</sup>

<sup>1</sup> School of Computer Science and Engineering,  
Northwestern Polytechnical University, Xi'an, China

<sup>2</sup> School of Astronautics, Northwestern Polytechnical University, Xi'an, China  
sunjinqiu@nwpu.edu.cn

<sup>3</sup> School of Computer Science, The University of Adelaide,  
South Australia, Australia

**Abstract.** The performance of traditional image quality assessment (IQA) methods are not robust, due to those methods exploit shallow hand-designed features. It has been demonstrated that deep neural network can learn more effective features compared with the traditional methods. In this paper we propose a multi-scale recursive deep neural network to accurately predict image quality. In order to learn more effective feature representations for IQA, many deep learning based works focus on using more layers and deeper network structure. However, deeper network layers introduce large numbers of parameters, which causes huge difficulty in training. The proposed recursive convolution layer ensures both the depth of the network and the light of parameters, which guarantees the convergence of training procedure. Moreover, extracting multi-scale features is the most prevalent approach in IQA. Based on this criteria, we using skip connection to combine information among layers, and it further enriches the coarse and fine features for quality assessment. The experimental results on the LIVE, CISQ and TID2013 databases show that the proposed algorithm outperforms all of the state-of-the-art methods, which verifies the effectiveness of our network architecture.

**Keywords:** Image quality assessment (IQA) · Feature extraction  
Deep learning · Convolutional neural networks (CNN) · Skip layer  
No-reference (NR)

## 1 Introduction

Image quality assessment (IQA) aims to evaluate image quality in various types of distortion during image acquisition, compression, transmission and restoration [23, 29, 34]. Typical ways of IQA include subjective quality assessment and

objective quality assessment. The former requires manual intervention, which is usually time consuming. In this case, image quality should be automatically generated and consistent with human perception. Therefore, objective image quality assessment obtains great attention in research. However, objective IQA has been a challenging issue in computer vision due to the variety of image distortion types and the difficulty in understanding the visual mechanisms of human perception.

Generally, objective IQA methods could be divided into three categories: full-reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA. FR-IQA methods are based on the full accessibility of raw image, and use this information to evaluate how much the distorted image has deviated from the origin one. The state-of-the-art FR-IQA methods include SSIM [24], MS-SSIM [26], FSIM [32], VIF [19] and GMSD [28]. The RR-IQA methods, including [25] and [22], extract only partial information of the reference image to predict the target image quality. However, in most cases, raw image is not available, therefore the NR-IQA methods that do not require a reference image becomes very necessary. Therefore, the NR-IQA method becomes very attractive in practical applications. Nevertheless, the lack of prior information has forced NR-IQA methods to work in different manners, making it the most challenging one among the three categories.

Early NR-IQA researches focus on extracting features from distorted images [11–14, 17], based on the observation that some features are distinguishable from the distortion free images. Although huge efforts have been devoted in designing feature forms, the performance improved rather slow, which shows the limitations and drawbacks of these hand-crafted features. On the other hand, deep learning methods have shown great ability in many computer vision [3, 7, 18, 30, 33], it also can be applied in NR-IQA research field, being expected to achieve better performances. Deep learning methods can use convolutional layer and pooling layer to extract features for IQA, and fully connected layers are used to mapping features to quality score. Based on the superiority that image features could be trained automatically instead of manually designed, deep learning frameworks are supposed to extract more capable features with higher efficiency. Not surprisingly, many deep learning based IQA works have achieved good performances following [6].

The motivation of our method is that learning the complex relationship between visual content and the perceived quality via a novel recursive convolutional neural network. In [21], it has been demonstrated that deep convolutional neural networks (CNNs) with more layers outperform shallow network architectures. Based on this, we train a deep neural network with 7 convolutional layers (including recursive layer) and 3 pooling layers for IQA feature extraction. Since the learned features are based on a data-driven approach, they are able to describe the changes in local image which are relevant to quality perception.

In this paper we propose a new framework for IQA. The contributions of this work are summarized as follows. First, we propose a deep convolutional neural network to obtain effective features from different distortions types for the esti-

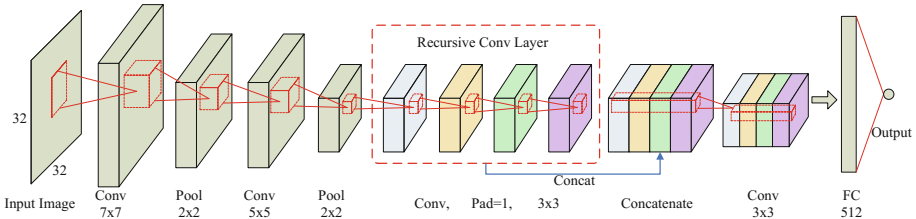
mation image quality based on training samples. Second, our network repeatedly applies the same convolutional layer as many times as desired. The convolution layers have drawback which will introduce more parameters and the pooling layers discard too much information. Since the parameters are shared in our network, the number of parameters does not increase. Third, we employ a skip-connection between layers to combines coarse and fine information. Extracting multi-scale feature is the most prevalent approach in IQA. How to fusion different scale information is the problem should be considered in quality assessment. The experimental results show that the proposed network is accurate compared with the existing IQA methods.

## 2 Related Work

NR-IQA methods can be generally classified into two groups: natural scene statistic (NSS) approaches and learning based approaches. NSS approaches are based on the observation that statistic features of an image changes with the presence of distortion. These approaches first extract the features from the query image, and then a regression model, which is learned previously to map the features with corresponding subjective perception scores, is used to predict the final quality score. In [14], BIQI was proposed to first estimate the distortion type, then a distortion-associated metric is applied to evaluate image quality. Later, Moorthy *et al.* improved BIQI into DIIVINE [13] by extracting features in wavelet domain. However, these distortion-specific methods may not perform well dealing with a generalized problem, since only certain types of distortions are considered. In [17], Saad *et al.* came up with an approach, called BLINDS-II, solve the problem by combining contrast and structure information in DCT. After realizing the potential of spatial features, BRISQUE [11] was proposed to capture the statistics of locally normalized illumination coefficients. NIQE [12] works in spatial domain as well, but uses a multivariate model (MVG) to fit the local features.

For learning approaches, image features are learned to map with subjective scores directly. To capture the relevant features, lots of training samples are needed. In [31], spatial features of training images are extracted to construct a codebook, and the raw image quality is estimated via encoding and pooling. In [27], FR-IQA method was used to build a training database, where image patches of similar quality are clustered to evaluate the target image quality. In [10], a generalized regression neural network was deployed to train the IQA model. Inspired by the recent success of CNNs for classification and detection tasks, Kang [6] proposed a shallow CNN consisting of a convolutional layer with max and min pooling and contrasting the normalized image patches as input. Gu [4] introduced sparse autoencoder based Image Quality Index (DIQI) for blind quality assessment. Bianco [1] estimated the image quality by average-pooling the scores predicted on multiple sub-regions of the original image. But those networks cannot make full use the information of different layers. Hou [5] learned qualitative evaluations directly and outputs numerical scores for general

utilization. Actually, images are represented by natural scene statistics features, some information is lost in this method. Bosse [2] constructed a network consisting of 10 convolutional layers and 5 pooling layers for feature extraction, and 2 fully connected layers for regression. Because it is more deeper than other networks, more parameters are introduced. In contrast, the proposed method considers taking the various advantages of different layer features while reducing the number of parameters.



**Fig. 1.** The framework of our recursive convolutional neural network. Features are extracted from the distorted patch image by a convolutional neural network in order to generate the score of distortion image. The dashed-boxed represents a recursive convolution layer which share the same parameters. The layers with different colors capture different information of distorted patch images.

### 3 Deep Neural Network for NR-IQA

The proposed network takes an RGB image as input. Given a distortion image, our goal is to get the quality score by estimating the mapping from the images to numerical ratings. The framework of our convolutional neural network is shown in Fig. 1. We sample non-overlapping patches from a given image, then the quality score of each patch can be estimated by multi-scale network with skip connection. The score of full size image is calculated by average the patch scores.

#### 3.1 Network Architecture for NR-IQA

The proposed network consists of 12 layers as shown in Fig. 1. The layers are organized as conv7-32, max pool, conv5-32, max pool, conv3-32, conv3-32, conv3-32, conv3-32 (four layers have same parameters), concatenate, conv3-128, FC-512, FC-1. This makes about 34 thousand trainable parameters in the network.

The convolution layers consist of filter banks. The response of each convolution layer is given by  $f_n^{l+1} = \sum_m (f_m^l * k_{m,n}^{l+1} + b_n^{l+1})$ , where  $k_{m,n}^{l+1}$  is the convolution kernel of the  $l$  layer  $m$ -th feature map to the  $l+1$  layer  $n$ -th feature map,  $f_m^l$  denotes the  $m$ -th feature of the  $l$  layer, and similar for  $f_n^{l+1}$ . The network use convolutional to learn effective feature representations. The first part of network is conv7-32, max pooling, conv5-32 and max pooling to capture the effective semantic information, then further refined by the recursive convolution

layer. In order to obtain an output of the same size as the previous input in recursive convolution layer, padding is used for convolutions layers.

Instead of traditional sigmoid or tanh neurons, all convolutional layers are activated through the rectified linear unit (ReLU) activation function:

$$g = \max(0, \sum_{n=1} w_n f_n) \quad (1)$$

where  $g$ ,  $w_n$ ,  $f_n$  denote the output of the ReLU, the weights of the ReLU and the output of the previous layer, respectively [15]. ReLUs enable the network to train several times faster compared tanh units. The input is  $32 \times 32$  image patches. All the max pooling layers have  $2 \times 2$  pixel-sized kernels in network. The network is trained end-to-end, the last layer is a linear regression with one output which is the quality score of the image patches. More detail will be explained in the following subsection.

### 3.2 Recursive Convolution Layer

Recursive convolution layer [8] takes the input matrix  $R_0$  (after conv7-32, max pool, conv5-32 and max pool layer) and computes the matrix output  $R_1, R_2, R_3, R_4$ . The same weight  $W_r$  and bias  $b_r$  are used for all operations in this step. For example,  $R_1$  is calculated by

$$R_1 = \max(0, W_r * R_0 + b_r) \quad (2)$$

Similar operation are performed in the following layers. The recurrence relation is

$$R_d = \max(0, W_r * R_{d-1} + b_r) \quad (3)$$

where,  $d = 1, 2, 3, 4$ . Then, we will get four feature matrices with different kinds of information. The *concat* layer is used to concatenate  $R_1, R_2, R_3, R_4$  with skip structure in order to fuse coarse and fine information. As we can see, recursive convolution layer increases the depth of network and reduces the number of parameters simultaneously.

### 3.3 Training

Due to training need more samples for network, We train our network on non-overlapping  $32 \times 32$  patches taken from large images, thus we have numbers of patches for training. However it may cause one problem that we only have the ground truth score of full image. Fortunately, the dataset of training images in our experiments have homogeneous distortions, we put the source images quality score to each patch of image. In the process of testing, the full size images quality score calculated by average the predicted patch scores.

$$q = \frac{1}{N_p} \sum_{i=1}^{N_p} f(x_i; w) \quad (4)$$

where,  $x_i$  denotes the input patch,  $f(x_i; w)$  represents the predicted score of patch  $x_i$  with parameters  $w$  and  $N_p$  is the number of patches sampled from the image.

Learning the mapping between distortion images and scores are achieved by minimizing the loss between the predicted score  $f(x_i; w)$  and the corresponding ground truth  $y_i$ . We adopt a similar objective function as [6]:

$$\min_w \frac{1}{N} \sum_{i=1}^N \|f(x_i; w) - y_i\|_{l_1} \quad (5)$$

where  $N$  is the number of images in the training set. We optimize the regression objective using the mini-batches gradient descent method based on the back-propagation learning rule. We implement our model using the Chainer package.

**Table 1.** LCC on LIVE dataset. The best two results are presented with bold face and italic fonts.

| Method     | JP2K         | JPEG         | WN           | GBLur        | FF           | ALL          |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| PSNR       | 0.896        | 0.860        | 0.986        | 0.783        | 0.890        | 0.824        |
| SSIM       | 0.937        | 0.928        | 0.970        | 0.874        | 0.943        | 0.863        |
| IFC        | 0.903        | 0.905        | 0.958        | 0.961        | 0.961        | 0.911        |
| VIF        | 0.962        | 0.943        | 0.984        | 0.974        | 0.962        | 0.950        |
| NSS        | 0.929        | 0.427        | 0.835        | 0.597        | 0.895        | 0.504        |
| BIQI       | 0.942        | 0.922        | 0.945        | 0.941        | 0.856        | 0.902        |
| BLIINDS-II | <i>0.963</i> | 0.979        | 0.985        | 0.948        | 0.944        | 0.923        |
| DIIVINE    | 0.922        | 0.921        | 0.988        | 0.923        | 0.888        | 0.917        |
| SRNSS      | 0.936        | 0.939        | 0.940        | 0.936        | 0.947        | 0.932        |
| BRISQUE    | 0.936        | 0.937        | 0.958        | 0.935        | 0.898        | 0.917        |
| CORNIA     | 0.915        | 0.902        | 0.952        | 0.940        | 0.913        | 0.903        |
| DLIQA      | 0.953        | 0.948        | 0.961        | 0.950        | 0.892        | 0.934        |
| CNN        | 0.953        | <b>0.981</b> | 0.984        | <i>0.953</i> | 0.933        | 0.953        |
| SOM        | 0.952        | 0.961        | <i>0.991</i> | <b>0.974</b> | <i>0.954</i> | <i>0.962</i> |
| Proposed   | <b>0.977</b> | <i>0.980</i> | <b>0.994</b> | 0.949        | <b>0.972</b> | <b>0.971</b> |

## 4 Experiments and Results

### 4.1 Datasets and Evaluation

**Datasets:** The following image quality datasets LIVE [20], TID2013 [16] and CSIQ [9] are used in our experiments. The LIVE dataset comprises 779 distorted images with five different types of distortions JPEG 2000 compression

(JP2K), JPEG compression (JPEG), White Gaussian Noise (WN), Gaussian blur (GBlur) and Fast Fading (FF) based on 29 source reference images. Distortion types have 7–8 degradation levels. Quality ratings were captured use a single-stimulus methodology. Differential Mean Opinion Scores (DMOS) of each image quality score is in a range of [0 100], where a higher DMOS means lower quality of the image.

The TID2013 image quality dataset include 3000 distorted images by 24 different distortions which derived from 25 reference images at 5 degradation levels each. The distortion types cover a wide range for real world, which makes TID2013 to be a challenging database. Each image is associated with a Mean Opinion Score (MOS) values lie in the range [0, 9], where a lower MOS denotes bad visual quality.

The CSIQ image quality dataset also contains of a corresponding set of 866 distorted images based on 30 reference images from 35 different observers reported in the form of DMOS. For the proposed Deep Quality system, the DMOS scores are mapped into 5 different levels of image quality for evaluation purposes. After alignment and normalization the DMOS values range in [0 1], where a higher DMOS presents lower quality.

**Evaluation:** Random segmentation of the data set was repeated 10 times to eliminate bias from individual data. For each repetition we calculate the Linear Correlation Coefficient (LCC), Root Mean Square Error (RMSE) and Spearman Rank Order Correlation Coefficient (SROCC) between the predicted quality score and the ground truth, then compute the average of metrics. The value of correlation metrics close to 1, or RMSE close to 0 indicates high performance.

**Table 2.** RMSE on LIVE dataset.

| Method    | JP2K         | JPEG         | WN           | GBlur        | FF           | ALL          |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|
| PSNR      | 7.187        | 8.170        | 2.680        | 9.772        | 7.516        | 9.124        |
| SSIM      | 5.671        | 5.947        | 3.916        | 7.639        | 5.485        | 8.126        |
| IFC       | 6.972        | 6.813        | 4.574        | 4.360        | 4.528        | 6.656        |
| VIF       | 4.449        | 5.321        | 2.851        | 3.533        | 4.502        | 5.024        |
| NSS       | 8.911        | 21.25        | 12.13        | 17.22        | 9.821        | 19.69        |
| BIQI      | 8.213        | 9.233        | 7.005        | <b>6.566</b> | 11.38        | 9.849        |
| BLINDS-II | 7.257        | 9.103        | 6.825        | 7.894        | 9.709        | 8.800        |
| DIIVINE   | 9.660        | 12.25        | <i>5.310</i> | 7.070        | 12.93        | 10.90        |
| SRNSS     | 7.892        | 7.948        | 7.971        | 7.591        | <i>7.157</i> | <i>7.618</i> |
| BRISQUE   | 8.150        | 9.230        | 7.273        | 7.516        | 9.536        | 9.538        |
| CORNIA    | 9.666        | 10.32        | 6.541        | 7.689        | 8.917        | 9.935        |
| DLIQA     | <i>7.250</i> | <i>7.596</i> | 5.881        | <i>6.570</i> | 9.540        | 8.149        |
| Proposed  | <b>4.785</b> | <b>4.793</b> | <b>2.288</b> | 6.749        | <b>5.129</b> | <b>5.514</b> |

**Table 3.** SROCC on LIVE dataset.

| Method     | JP2K         | JPEG         | WN           | GBlur        | FF           | ALL          |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|
| PSNR       | 0.890        | 0.841        | 0.985        | 0.782        | 0.890        | 0.820        |
| SSIM       | 0.932        | 0.903        | 0.963        | 0.894        | 0.941        | 0.851        |
| IFC        | 0.892        | 0.866        | 0.938        | 0.959        | 0.963        | 0.913        |
| VIF        | 0.953        | 0.913        | 0.986        | 0.973        | 0.965        | 0.953        |
| NSS        | 0.882        | 0.247        | 0.852        | 0.644        | 0.859        | 0.339        |
| BIQI       | 0.940        | 0.915        | 0.971        | 0.947        | 0.831        | 0.903        |
| BLIINDS-II | 0.951        | 0.942        | 0.978        | 0.944        | 0.927        | 0.920        |
| DIIVINE    | 0.913        | 0.910        | <i>0.984</i> | 0.921        | 0.863        | 0.916        |
| SRNSS      | 0.928        | 0.931        | 0.938        | 0.933        | <i>0.941</i> | 0.930        |
| BRISQUE    | 0.910        | 0.919        | 0.955        | 0.941        | 0.874        | 0.920        |
| CORNIA     | 0.903        | 0.889        | 0.958        | 0.946        | 0.915        | 0.906        |
| DLIQA      | 0.933        | 0.914        | 0.968        | 0.947        | 0.857        | 0.929        |
| CNN        | <i>0.952</i> | <b>0.977</b> | 0.978        | 0.962        | 0.908        | 0.956        |
| SOM        | 0.947        | <i>0.952</i> | <i>0.984</i> | <b>0.976</b> | 0.937        | <i>0.964</i> |
| Proposed   | <b>0.966</b> | 0.949        | <b>0.989</b> | <i>0.963</i> | <b>0.973</b> | <b>0.965</b> |

## 4.2 Consistency Experiment

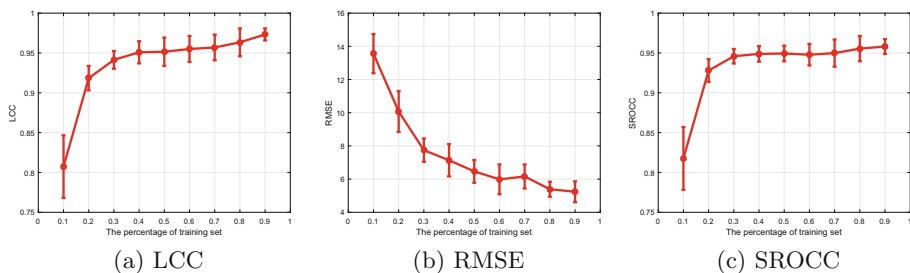
In this subsection, we consider how the proposed network corresponds to human assessment on the LIVE database. We train and test on images of all five distortions (JP2K, JPEG, WN, BLUR and FF) together without providing a distortion type. Since machine learning requires training samples, we randomly divide into several groups and the rest are used as test sets. To eliminate effects from the separated data, the random division of the data set was repeated 10 times. Other learning-based BIQA approaches are all executed in this way.

We employ four traditional full-reference IQA methods as the benchmarks, including PSNR, SSIM, IFC, and VIF. In addition, there are 10 kinds of BIQA methods for comparison: (1) NSS; (2) BIQI; (3) BLIINDS-II; (4) DIIVINE; (5) SRNSS; (6) BRISQUE; (7) CORNIA; (8) DLIQA; (9) CNN; (10) SOM. All of these methods are based on machine learning and can be found in [5] and [2]. The results are evaluated by using 90% of the data for training, then testing on the other 10% of the data.

Tables 1, 2 and 3 show the experimental results of different methods on the LIVE dataset with different distortion types. The best two results are shown in bold face and italic fonts. Our proposed network outperforms all previous NR-IQA methods. From LCC and SROCC we can see that the proposed network works well on the entire database, especially on JP2K, WN and FF. SOM method ranks second in the entire database. In particular, our RMSE is significantly reduced compared to other methods. This phenomenon comes from recursive



convolution and skip structure. We have reason to believe that the proposed network can obtain more useful features for describing image quality.



**Fig. 2.** Performance of the proposed network versus the percentage of training sets

Figure 2 shows the relationship between the percentage of training sets and the performance of proposed network. The random split of the LIVE II dataset is repeated 10 times, each group include training and testing data. The average score of LCC, RMSE and SROCC are calculated according this datasets. As can be seen, the RMSE curve decreases slowly with the increases of the training set. The trend of LCC and SROCC curves is consistent. The new network can get better results even when the training sets are fewer.

### 4.3 Extensibility Experiment

To evaluate the performance of generalization we perform a cross-dataset evaluation as shown in Table 4. The subset of CSIQ and TID2013 includes only four types of distortions that are shared with LIVE dataset. Unfortunately, no results are available for the other methods. All models are trained on the full LIVE dataset and evaluated on subset of CSIQ and TID2013 or full set. We can see that our network is superior to previous state of the art methods on full dataset.

**Table 4.** SROCC results of the cross-dataset evaluations.

| Method     | Subset       |              | Full         |              |
|------------|--------------|--------------|--------------|--------------|
|            | CSIQ         | TID2013      | CSIQ         | TID2013      |
| DIIVINE    | -            | -            | 0.596        | 0.355        |
| DLIINDS-II | -            | -            | 0.577        | 0.393        |
| BRISQUE    | 0.899        | 0.882        | 0.557        | 0.367        |
| CORNIA     | <b>0.899</b> | 0.892        | 0.603        | 0.429        |
| CNN        | -            | <b>0.920</b> | -            | -            |
| Proposed   | 0.897        | 0.889        | <b>0.609</b> | <b>0.412</b> |

## 5 Conclusion

This paper develops a CNN for no-reference image quality assessment. Our approach describes a deep recursive neural network which predict image quality accurately by learning the mapping between images and their corresponding scores. Recursive convolution layer increases the depth of net and reduces the number of parameters simultaneously. The experimental results prove its efficiency and robustness to different standard IQA datasets, and verifies the high consistency between the designed network and human perception.

## References

1. Bianco, S., Celona, L., Napoletano, P., Schettini, R.: On the use of deep learning for blind image quality assessment. *Signal Image Video Process.* **3**, 1–8 (2016)
2. Bosse, S., Maniry, D., Müller, K.R., Wiegand, T., Samek, W.: Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Process.* **27**(1), 206–219 (2018)
3. Gong, D., et al.: From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In: *Computer Vision and Pattern Recognition*, pp. 1827–1836 (2017)
4. Gu, K., Zhai, G., Yang, X., Zhang, W.: Deep learning network for blind image quality assessment. In: *IEEE International Conference on Image Processing*, pp. 511–515 (2014)
5. Hou, W., Gao, X., Tao, D., Liu, W.: Blind image quality assessment via deep learning **26**(6), 1275–1286 (2015)
6. Kang, L., Ye, P., Li, Y., Doermann, D.: Convolutional neural networks for no-reference image quality assessment. In: *Computer Vision and Pattern Recognition*, pp. 1733–1740 (2014)
7. Kavukcuoglu, K., Boureau, Y.L., Gregor, K., Lecun, Y.: Learning convolutional feature hierarchies for visual recognition. In: *International Conference on Neural Information Processing Systems*, pp. 1090–1098 (2010)
8. Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1637–1645 (2016)
9. Larson, E.C., Chandler, D.M.: Most apparent distortion: full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* **19**(1), 6–11 (2010)
10. Li, C., Bovik, A.C., Wu, X.: Blind image quality assessment using a general regression neural network. *IEEE Trans. Neural Netw.* **22**(5), 793–799 (2011)
11. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **21**(12), 4695–4708 (2012)
12. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013)
13. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **20**(12), 3350 (2011)
14. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. *IEEE Signal Process. Lett.* **17**(5), 513–516 (2010)

15. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: International Conference on International Conference on Machine Learning, pp. 807–814 (2010)
16. Ponomarenko, N., et al.: Color image database TID2013: peculiarities and preliminary results. In: European Workshop on Visual Information Processing, pp. 106–111 (2013)
17. Saad, M.A., Bovik, A.C., Charrier, C.: Blind image quality assessment: a natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **21**(8), 3339–3352 (2012)
18. Schmidhuber, J.: Multi-column deep neural networks for image classification. In: *Computer Vision and Pattern Recognition*, pp. 3642–3649 (2012)
19. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **15**(2), 430 (2006)
20. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **15**(11), 34–51 (2006)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations* (2015)
22. Soundararajan, R., Bovik, A.C.: RRED indices: reduced reference entropic differencing for image quality assessment. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **21**(2), 517–526 (2012)
23. Wang, Z., Bovik, A.: *Modern Image Quality Assessment*. Morgan and Claypool, San Rafael (2006)
24. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
25. Wang, Z., Simoncelli, E.P.: Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. *Proc. SPIE* **56**, 149–159 (2005)
26. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale structural similarity for image quality assessment (2004)
27. Xue, W., Zhang, L., Mou, X.: Learning without human scores for blind image quality assessment. In: *Computer Vision and Pattern Recognition*, pp. 995–1002 (2013)
28. Xue, W., Zhang, L., Mou, X., Bovik, A.C.: Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **23**(2), 684–695 (2013)
29. Yan, Q., Sun, J., Li, H., Zhu, Y., Zhang, Y.: High dynamic range imaging by sparse representation. *Neurocomputing* **269**, 160–169 (2017)
30. Yang, J., Gong, D., Liu, L., Shi, Q.: Seeing deeply and bidirectionally: a deep learning approach for single image reflection removal. In: *European Conference on Computer Vision* (2018)
31. Ye, P., Kumar, J., Kang, L., Doermann, D.: Unsupervised feature learning framework for no-reference image quality assessment. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1098–1105 (2012)
32. Zhang, D.: FSIM: a feature similarity index for image quality assessment. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **20**(8), 2378–2386 (2011)
33. Zhang, L., et al.: Adaptive importance learning for improving lightweight image super-resolution network. *arXiv preprint [arXiv:1806.01576](https://arxiv.org/abs/1806.01576)* (2018)
34. Zhang, L., Wei, W., Zhang, Y., Shen, C., van den Hengel, A., Shi, Q.: Cluster sparsity field: an internal hyperspectral imagery prior for reconstruction. *Int. J. Comput. Vis.* **126**(8), 797–821 (2018)