

RESEARCH ARTICLE

# A shallow convolutional neural network for blind image sharpness assessment

Shaode Yu<sup>1,2</sup>, Shibin Wu<sup>1,2</sup>, Lei Wang<sup>1</sup>, Fan Jiang<sup>3</sup>, Yaoqin Xie<sup>1\*</sup>, Leida Li<sup>4\*</sup>

**1** Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China, **2** Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong, China, **3** Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan, China, **4** School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu, China

\* [yq.xie@siat.ac.cn](mailto:yq.xie@siat.ac.cn) (YQX); [lileida@cumt.edu.cn](mailto:lileida@cumt.edu.cn) (LDL)



## Abstract

Blind image quality assessment can be modeled as feature extraction followed by score prediction. It necessitates considerable expertise and efforts to handcraft features for optimal representation of perceptual image quality. This paper addresses blind image sharpness assessment by using a shallow convolutional neural network (CNN). The network takes single feature layer to unearth intrinsic features for image sharpness representation and utilizes multilayer perceptron (MLP) to rate image quality. Different from traditional methods, CNN integrates feature extraction and score prediction into an optimization procedure and retrieves features automatically from raw images. Moreover, its prediction performance can be enhanced by replacing MLP with general regression neural network (GRNN) and support vector regression (SVR). Experiments on Gaussian blur images from LIVE-II, CSIQ, TID2008 and TID2013 demonstrate that CNN features with SVR achieves the best overall performance, indicating high correlation with human subjective judgment.

## OPEN ACCESS

**Citation:** Yu S, Wu S, Wang L, Jiang F, Xie Y, Li L (2017) A shallow convolutional neural network for blind image sharpness assessment. PLoS ONE 12(5): e0176632. <https://doi.org/10.1371/journal.pone.0176632>

**Editor:** You Yang, Huazhong University of Science and Technology, CHINA

**Received:** November 18, 2016

**Accepted:** April 13, 2017

**Published:** May 1, 2017

**Copyright:** © 2017 Yu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript.

**Funding:** This work is supported in part by grants from National Natural Science Foundation of China (Grant Nos. 81501463 and 61379143), Natural Science Foundation of Guangdong Province (Grant No. 2014A030310360), Major Project of Guangdong Province (Grant No. 2014B010111008), the Qing Lan Project of Jiangsu Province, National 863 Programs of China (Grant No. 2015AA043203), National Key Research Program of China (Grant No. 2016YFC0105102)

## Introduction

A picture wins a thousand words. With the rapid pace of modern life and the massive dissemination of smart phones, digital images have been a major source of information acquisition and distribution. Since an image is prone to various kinds of distortions from its capture to the final display on digital devices, a lot of attention has been paid to the assessment of perceptual image quality [1–8].

Subjective image quality assessment (IQA) is the most straightforward. However, it is laborious and may introduce bias and errors. Comparatively, objective evaluation of visual image quality with full- or reduced-reference based methods enables impartial judgment [9–22]. These algorithms have reached high-level performance, while in most possible situations, the reference messages are not easy or impossible to acquire. Thus, no-reference or blind IQA methods are more useful in real applications [23–34].

Blind image quality assessment (BIQA) mainly consists of two steps, feature extraction ( $T$ ) and score prediction ( $f$ ). Before rating an image,  $T$  and  $f$  should be prepared. The former aims

and Guangdong Innovative Research Team Program (Grant No. 2011S013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

to select optimal features for image quality representation, while the latter builds the functional relationship between the features and subjective scores. With considerable expertise and efforts, a BIQA system can be built. As such, a test image ( $I$ ) is input to the system and represented with features ( $T$ ). Finally, the function  $f$  will quantify the features and figure out a numerical score ( $s$ ) as the output, denoting the predicted quality of the test image. The procedure for score prediction can be formulated as follows,

$$s = f(T(I)). \quad (1)$$

Blind image sharpness assessment (BISA) is studied in this paper. Among various kinds of distortions, sharpness is commonly degraded by camera out-of-focus, relative target motion and lossy image compression. It is crucial to readability and content understanding. Sharpness is inversely related to blur which is typically determined by the spread of edges in the spatial domain, and accordingly the attenuation of high frequency components. Karam *et al.* [35] introduced the Just Noticeable Blur (JNB) model and integrated local contrast and edge width in each edge blocks into a probability summation model. Later, they improved the model with the cumulative probability of blur detection (CPBD) [36]. Ciancia *et al.* [37] selected blur-related features as the input of a neural network and realized no-reference blur assessment with multi-feature classifiers. Vu *et al.* [38] combined two features, the high frequency content with the slope of local magnitude spectrum and the local contrast with total variation, to form the spectral and spatial sharpness (S3) index. Vu *et al.* [39] defined a fast image sharpness (FISH) metric which weights the log-energies of wavelet coefficients. Hassen *et al.* [40] explored the strength of local phase coherence (LPC) based on the observation that blur disrupts image LPC structures. Sang *et al.* [41, 42] used the shape of singular value curve (SVC) to measure the extent of blur, because the extent of blur results in attenuation of singular values. Bahrami and Kot [43] took account of maximum local variation (MLV) of each pixel and utilized the standard deviation of ranking weighted MLVs as the sharpness score. Li *et al.* [44] proposed the sparse representation based image sharpness (SPARISH) model that utilizes dictionary learning of natural image patches. Gu *et al.* [45] designed an autoregressive based image sharpness metric (ARISM) via image analysis in the autoregressive parameter space. Li *et al.* [46] presented a blind image blur evaluation (BIBLE) index which characterizes blur with discrete moments, because noticeable blur affects the moment magnitudes of images.

Deep learning has revolutionized image representation and shed light on utilizing high-level features for BIQA [47, 48]. Li *et al.* [49] adapted Shearlet transform for spatial feature extraction and employed a deep network for image score regression. Hou and Gao [50] recast BIQA as a classification problem and used a saliency-guided deep framework for feature retrieval. Li *et al.* [51] took the Prewitt magnitudes of segmented images as the input of convolutional neural network (CNN). Lv *et al.* [52] explored the local normalized multi-scale difference of Gaussian response as features and designed a deep network for image quality rating. Hou *et al.* [53] designed a deep learning model trained by deep belief net and then fine-tuned it for image quality estimation. Yet it is found that some deep learning based methods need to handcraft features [49–52] or redundant operations [50, 52, 53].

This paper presents a shallow CNN to address BISA. On the one hand, several studies indicate that image sharpness is generally characterized by the spread of edge structures [35–38, 44, 46]. Interestingly, what CNN learns in the first layer are mainly edges [47, 48]. Thus, it is intuitive to design a single feature layer CNN for image sharpness estimation. On the other hand, small data sets make deep networks hard to converge which may increase the risk of over-fitting. Consequently, a shallow CNN can be well trained with limited samples [54]. To the best of our knowledge, the most similar work is Kang's CNN [55]. The network utilizes

two full-connection layers and obtains dense features by both maximum and minimum pooling before image scoring. Relatively, our network is much simpler in the architecture and more suitable for the analysis of small databases. Besides, our CNN is verified with Gaussian blurring images from four popular databases. After features are retrieved for representation of sharpness, the prediction performance of multilayer perceptron (MLP) is compared to both general regression neural network (GRNN) [56] and support vector regression (SVR) [57]. In the end, the effect of color information on our CNN and the running time are reported.

### A shallow CNN

The simplified CNN consists of one feature layer and the feature layer is made up of convolutional filtering and average pooling. As shown in Fig 1, a gray-scale image is pre-processed with local contrast normalization. Then, a number of image patches are randomly cropped for feature extraction. At last, the features are as input to MLP for score prediction. By supervised learning, parameters in the network are updated and fine-tuned with back-propagation.

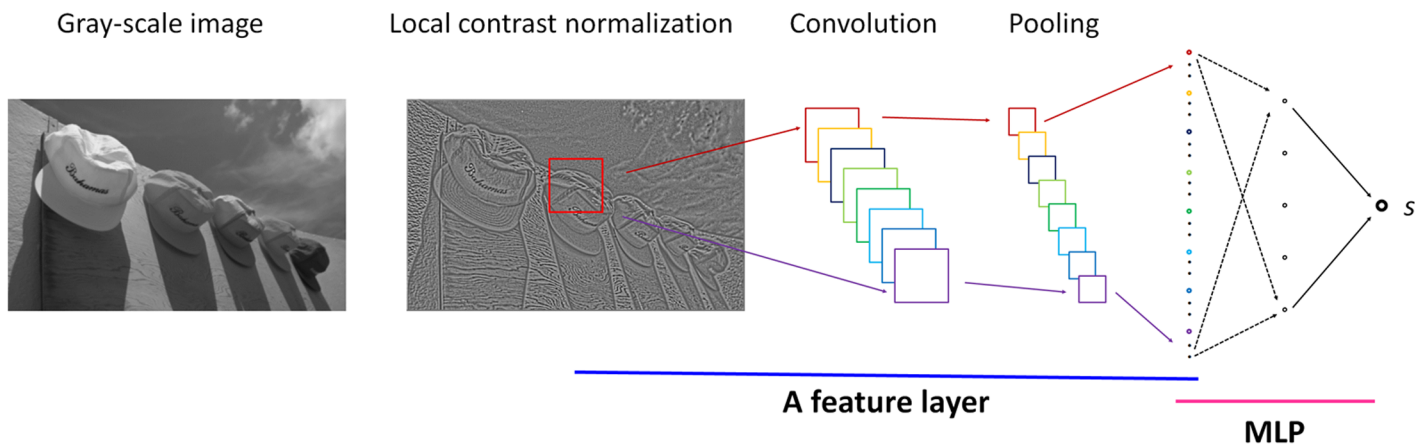
### Feature extraction

**Local contrast normalization.** It has a decorrelating effect in spatial image analysis by applying a local non-linear operation to remove local mean displacements and to normalize the local variance [25, 58]. As in [52, 55], the local normalization is formulated as following,

$$\tilde{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C}, \tag{2}$$

where,

$$\mu(i, j) = \frac{1}{(2P + 1)(2Q + 1)} \sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} I(i + p, j + q), \tag{3}$$



**Fig 1. The proposed BISA system.** A gray-scale image is pre-processed with local contrast normalization and then a number of image patches are randomly cropped for CNN training, validation and final testing.

<https://doi.org/10.1371/journal.pone.0176632.g001>

and

$$\sigma^2(i, j) = \sum_{p=-P}^{p=P} \sum_{q=-Q}^{q=Q} (I(i + p, j + q) - \mu(i, j))^2. \tag{4}$$

In the equations,  $I(i, j)$  is the pixel intensity value at  $(i, j)$ ,  $\tilde{I}(i, j)$  is its normalized value,  $\mu(i, j)$  is the mean value,  $\sigma(i, j)$  is the standard deviation and  $C$  is a positive constant ( $C = 10$ ). Besides,  $[2P + 1, 2Q + 1]$  is the window size and  $P = Q = 3$ .

**Feature representation.** Each patch randomly cropped in the pre-processed image is through convolutional filtering and pooling before full connection to MLP. A feature vector of an image patch is generated and formulated as,

$$X = T(I_p) = (x_1, \dots, x_l, \dots, x_n)', \tag{5}$$

where  $I_p$  is an image patch,  $n$  is the feature dimension and  $x_l$  is the  $l^{th}$  component of the feature vector  $X$ .

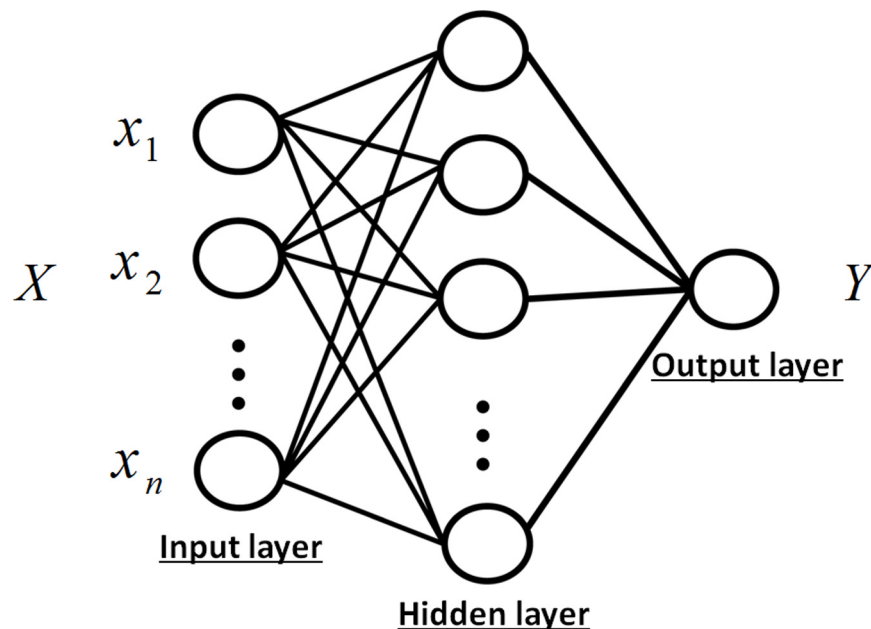
### Score prediction

**Multilayer perceptron (MLP).** Fig 2 illustrates an MLP with a hidden layer. The output  $f(X)$  with regard to the input feature  $X$  can be expressed as following,

$$f(X) = f_{mlp}(\mathbf{w}, \mathbf{b}; X), \tag{6}$$

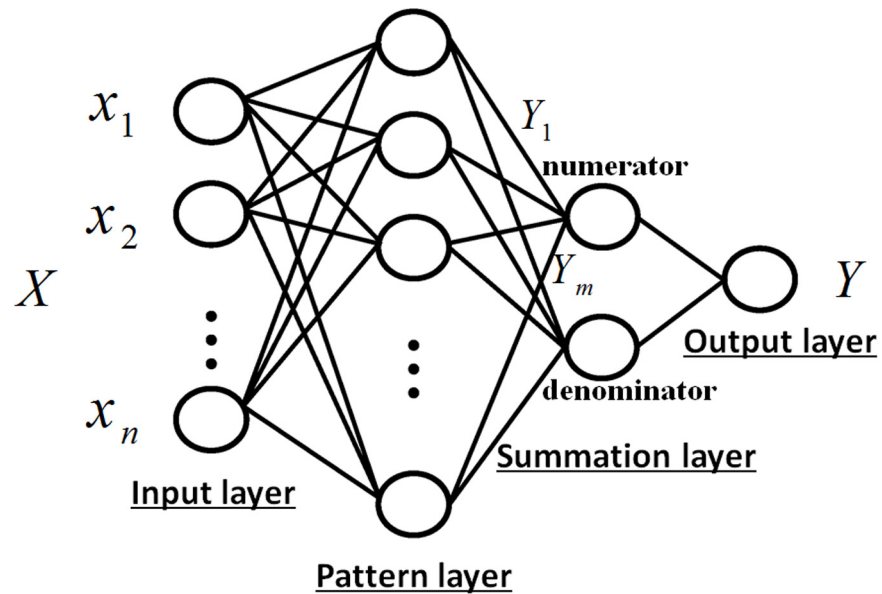
where  $f_{mlp}$  denotes an activation function, while  $\mathbf{w}$  and  $\mathbf{b}$  respectively stand for the weight vector and the bias vector.

**General regression neural network (GRNN).** GRNN is a powerful regression tool based on statistical principles [56]. It takes only a single pass through a set of feature instances and requires no iterative training. GRNN consists of four layers as shown in Fig 3. Assume that



**Fig 2. MLP with one hidden layer.** It consists of three layers, the input layer, the hidden layer and the output layer.

<https://doi.org/10.1371/journal.pone.0176632.g002>



**Fig 3. A semantic description of GRNN.** It consists of four layers, the input layer, the pattern layer, the summation layer and the output layer.

<https://doi.org/10.1371/journal.pone.0176632.g003>

$m$  samples  $\{X_i, Y_i\}_{i=1}^m$  have been used to train the GRNN. To an input feature vector  $X$ , its output  $f(X)$  can be described as below,

$$f(X) = f_{grnn}(X) = \frac{\sum_{i=1}^n Y_i e^{-(X-X_i)'(X-X_i)/2\sigma^2}}{\sum_{i=1}^n e^{-(X-X_i)'(X-X_i)/2\sigma^2}}, \quad (7)$$

where  $Y_i$  is the weight between the  $i^{th}$  neuron in the pattern layer and the numerator neuron in the summation layer, and  $\sigma$  is a spread parameter. In GRNN, only  $\sigma$  is tunable and a larger value leads to a smoother prediction.

**Support vector regression (SVR).** SVR is effective in handling numerical prediction in high dimension space [57, 59]. For an input  $X$ , the goal of  $\epsilon$ -SVR is to find a function  $f(X)$  that has the maximum deviation of  $\epsilon$  from the subjective score  $Y$  for all the training patches. The function is defined by

$$f(X) = f_{svr}(X) = \mathbf{w}'\varphi(X) + \gamma, \quad (8)$$

where  $\varphi(\cdot)$  is a nonlinear function,  $\mathbf{w}$  is a weight vector and  $\gamma$  is a bias. The aim is to find  $\mathbf{w}$  and  $\gamma$  from the training data such that the error is less than a predefined value of  $\epsilon$ . The radial basis function is used as the kernel function,  $K(X_i, X) = e^{-\rho\|X_i-X\|}$ , and  $\rho$  is a positive parameter that controls the radius and  $X_i$  is a training sample. By using a validation set to tradeoff the prediction error,  $\rho$  and  $\epsilon$  are determined [60].

### Network training

CNN is end-to-end trained by supervised learning with stochastic gradient descent. Assume there are a set of features  $\{X_i\}_{i=1}^n$  and corresponding scores  $\{Y_i\}_{i=1}^n$ . The training aims to

minimize the loss function  $L(\mathbf{w}, \mathbf{b})$ ,

$$\begin{aligned}
 L(\mathbf{w}, \mathbf{b}) &= \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \|Y_i - s_i\|^2 \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \|Y_i - f_{mlp}(\mathbf{w}, \mathbf{b}; X_i)\|^2 \right),
 \end{aligned}
 \tag{9}$$

which is the sum of square error between the predicted  $s_i$  and the subjective score  $Y_i$ .

Using gradient descent, the relationship between the  $l^{th}$  and the  $(l + 1)^{th}$  iteration to each weight component can be described as following,

$$w^{l+1} = \mu w^l - \eta \frac{\partial L(\mathbf{w}, \mathbf{b})}{\partial w^l}, \tag{10}$$

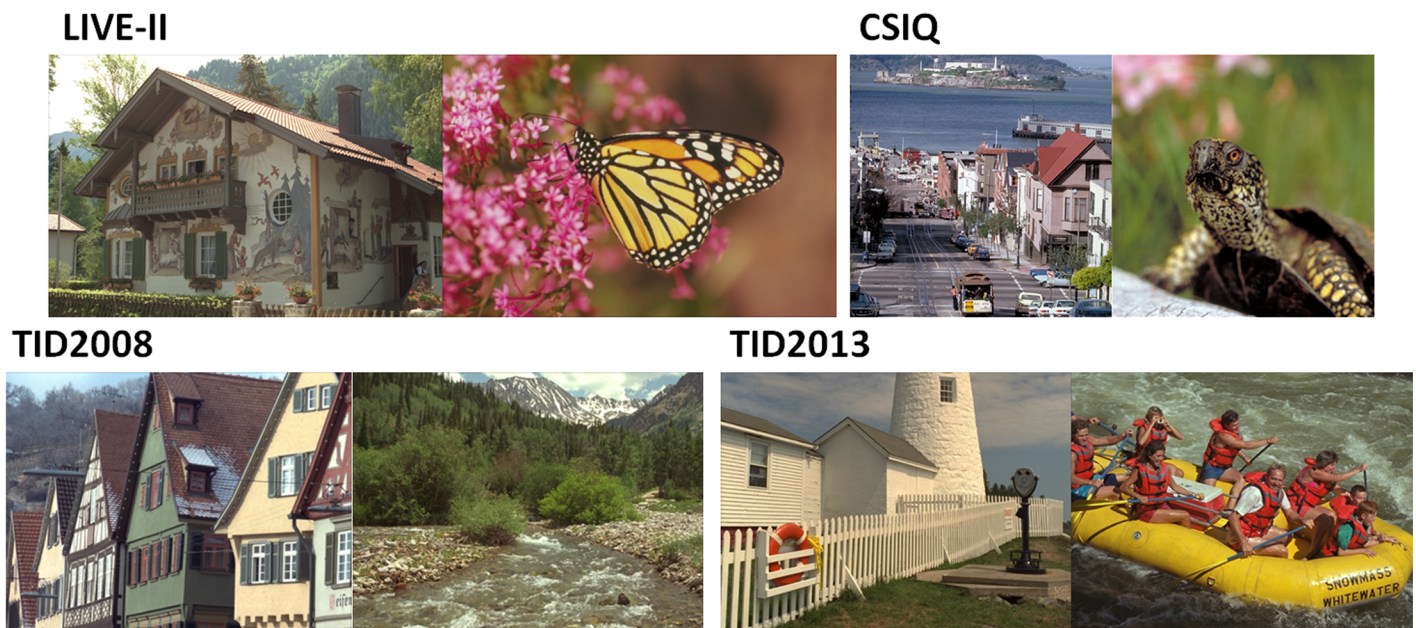
$$b^{l+1} = \mu b^l - \eta \frac{\partial L(\mathbf{w}, \mathbf{b})}{\partial b^l}, \tag{11}$$

where  $\mu$  is the momentum that indicates the contribution of the previous weight update in the current iteration, and  $\eta$  denotes the learning rate.

## Experiments

### Images for performance evaluation

Gaussian blurring images are collected from four popular databases. LIVE-II [10] and CSIQ [61] respectively contain 29 and 30 reference images which are distorted with 5 blur levels and scored by differential mean opinion scores (DMOS). Both TID2008 [62] and TID2013 [63] have 25 references and use mean opinion scores (MOS) for scoring. Each reference image in TID2008 and TID2013 is degraded with 4 and 5 different blur levels, respectively. Fig 4 shows some representative images.



**Fig 4. Example of Gaussian blurring images in four databases.**

<https://doi.org/10.1371/journal.pone.0176632.g004>

## Experiment design

LIVE-II is taken as the baseline database for tuning parameters in CNN, GRNN and SVR. Blurred images in LIVE-II are portioned into 20:4:5 for training, validation and test, respectively. After that, parameters in GRNN and SVR are optimized based on learned features from CNN. In the end, about 60%, 20% and 20% blurring images in each database are randomly selected for training, validation and test, respectively.

Besides Kang's CNN [55], ten state-of-the-art BISA methods are evaluated. These methods are JNB [35], CPBD [36], S3 [38], FISH [39], LPC [40], SVC [42], MLV [43], SPARISH [44], ARISM [45] and BIBLE [46]. In the end, the running time of involved algorithms and the effect of color information on our CNN are studied.

## Performance criteria

Two criteria are recommended for IQA performance evaluation by the video quality experts groups (VQEG, <http://www.vqeg.org>). Pearson linear correlation coefficient (PLCC) evaluates the prediction accuracy, while Spearman rank-order correlation coefficient (SROCC) measures the prediction monotonicity. Values of both criteria range in [0, 1] and higher value indicates better rating prediction.

A nonlinear regression is first applied to map the predicted scores to subjective human ratings using a five-parameter logistic function as follows,

$$Q(s) = q_1 \left( \frac{1}{2} - \frac{1}{1 + e^{q_2(s - q_3)}} \right) + q_4 s + q_5, \quad (12)$$

where  $s$  and  $Q(s)$  are the input score and the mapped score, and  $q_i$  ( $i = 1, 2, 3, 4, 5$ ) are determined during the curve fitting.

## Software and platform

Softwares are run on Linux system (Ubuntu 14.04). The system is embedded with 8 Intel Xeon (R) CPU (3.7GHz), 16GB DDR RAM and one GPU card (Nvidia 1070). Kang's CNN is implemented by us following the paper [55]. Both CNN models are realized with Theano 0.8.2 (Python 2.7.6) and accessible on GitHub at present for fair comparison (<https://github.com/Dakar-share/Plosone-IQA>). Other codes are realized with Matlab. Ten BISA methods are provided by authors and estimated without any modifications, GRNN is with the function *newgrnn* and SVR is from LIBSVM [59].

## Result

### Parameter tuning

Several parameters are experimentally determined, the patch number per image ( $P_n$ ), the kernel number ( $K_n$ ) and the kernel size ( $[K_x, K_y]$ ) in feature extraction, and the iteration number ( $N_i$ ) in network training. In addition, the spread parameter ( $\sigma$ ) in GRNN and cost function ( $c$ ) in  $\varepsilon$ -SVR are also studied. Note that in the network, we define the size of image patch [16 16], the learning rate  $\eta = 0.01$ , the bias  $\gamma = 0.1$  and the momentum  $\mu = 0.9$ , and other parameters are set by default.

**Parameters in CNN.** Fig 5 shows CNN performance when the iteration number ( $N_i$ ) varies from  $10^3$  to  $10^4$  and the patch number per image ( $P_n$ ) changes from  $10^2$  to  $10^3$ . No much change is found after  $N_i$  reaches 4000. On the other side,  $P_n = 400$  is a good point to tradeoff PLCC and SROCC. Therefore, we use  $N_i = 4000$  and  $P_n = 400$  hereafter.

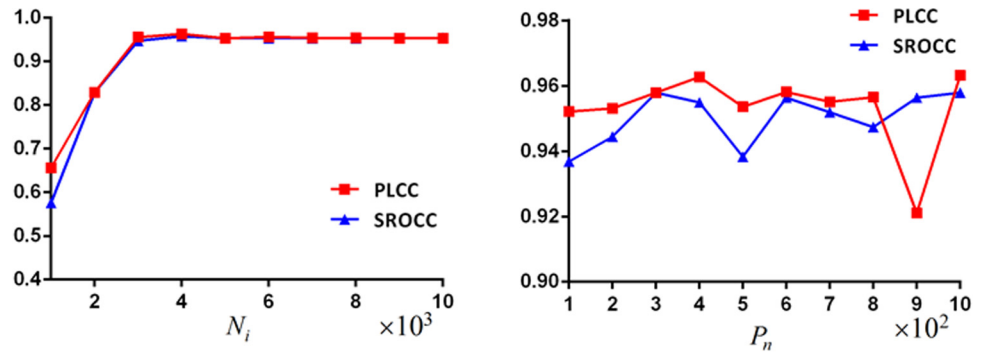


Fig 5. CNN prediction performance with  $N_i$  or  $P_n$  changes.

<https://doi.org/10.1371/journal.pone.0176632.g005>

Table 1. CNN performance with regard to kernel number and kernel size.

Kernel number	8	16	24	32
PLCC	0.9444	<b>0.9634</b>	0.9352	0.9298
SROCC	0.9519	<b>0.9543</b>	0.9504	0.9323
Kernel size	[3 3]	[5 5]	[7 7]	[9 9]
PLCC	0.9606	0.9508	<b>0.9632</b>	0.9319
SROCC	0.9669	<b>0.9684</b>	0.9579	0.9278

<https://doi.org/10.1371/journal.pone.0176632.t001>

Table 1 shows the CNN performance with regard to the kernel number ( $K_n$ ) and the kernel size ( $[K_x, K_y]$ ). When  $K_n = 16$ , CNN performs well, while it is unstable when  $K_n = 32$ . On the other hand, prediction performance of CNN is insensitive to kernel size  $[K_x, K_y]$  changes. So we define  $K_n = 16$  and  $K_x = K_y = 7$ .

**Parameters in GRNN and SVR.** The spread parameter ( $\sigma$ ) in GRNN and the cost function ( $c$ ) in  $\epsilon$ -SVR are studied with learned CNN features. Fig 6 shows PLCC and SROCC values when  $\sigma$  or  $c$  changes. The left plot indicates that when  $\sigma = 0.01$ , GRNN performs the best. The right shows that PLCC and SROCC increase when  $\log_{10}(c)$  increases, while when  $\log_{10}(c) > 1$ , SROCC keeps stable. Thus,  $\sigma = 0.01$  in GRNN and  $c = 50$  in  $\epsilon$ -SVR.

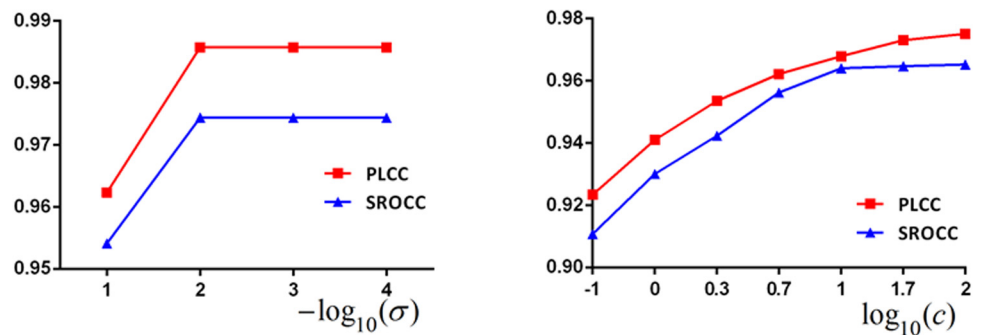
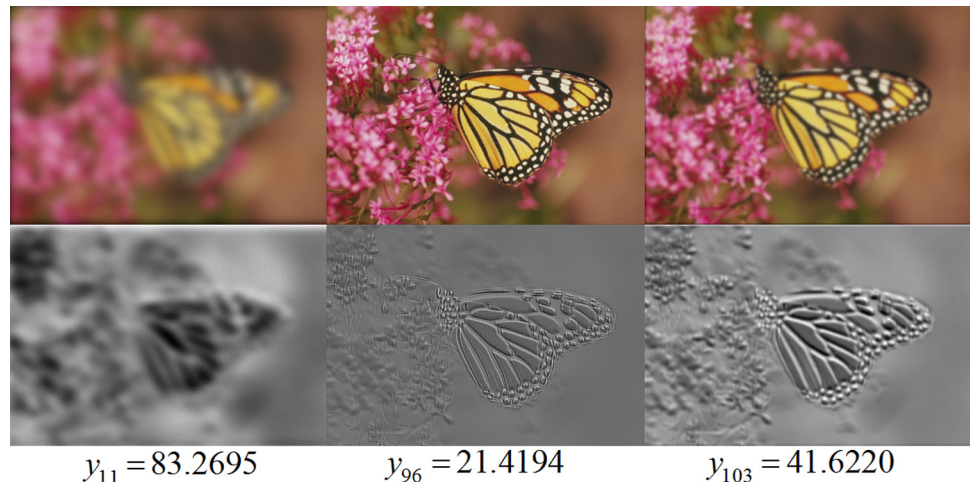


Fig 6. GRNN (left) and SVR (right) respectively perform when the spread parameter  $\sigma$  and the cost function  $c$  changes based on learned CNN features.

<https://doi.org/10.1371/journal.pone.0176632.g006>





**Fig 7. One trained kernel visualized by using “monarch.bmp”.** After convolutional filtering with the trained kernel, edge structures is hard to notice in heavily blurred images ( $y_{11}$ ), while fine structures can be seen in relatively high-quality images ( $y_{96}$ ).

<https://doi.org/10.1371/journal.pone.0176632.g007>

### Learned CNN features

One trained kernel is visualized by using “monarch.bmp” in LIVE-II. Blurred images and their filtered results are shown in Fig 7. The top row shows Gaussian blurring images and the bottom row are images after convolutional filtering with the trained kernel. Underneath the filtered results are subjective scores, where lower scores indicate better visual quality. Compared to the relatively high-quality image ( $y_{96}$ ), fine structures vanish in low-quality images ( $y_{11}$  and  $y_{103}$ ).

### Algorithm performance

Table 2 summarizes the PLCC values and the highest values are marked in bold face. With handcrafted features, BIBLE [46] predicts the best, followed by SPARISH [44]. For CNNs, Kang’s CNN is instable. It achieves the best performance on TID2013 and the lowest value on

**Table 2. Performance evaluation with PLCC on Gaussian blurring images.**

	LIVE-II	CSIQ	TID2008	TID2013	Overall
JNB [35]	0.8161	0.8061	0.6931	0.7115	0.7567
CPBD [36]	0.8955	0.8822	0.8236	0.8620	0.8658
S3 [38]	0.9434	0.9107	0.8542	0.8816	0.8975
FISH [39]	0.9043	0.9231	0.8079	0.8327	0.8670
LPC [40]	0.9181	0.9158	0.8573	0.8917	0.8957
SVC [42]	0.9416	0.9319	0.8556	0.8762	0.9013
MLV [43]	0.9429	0.9247	0.8583	0.8818	0.9019
SPARISH [44]	0.9595	0.9380	0.8891	0.9004	0.9217
ARISM [45]	0.9560	0.9410	0.8430	0.8954	0.9088
BIBLE [46]	0.9622	0.9403	0.8929	0.9051	0.9251
Kang’s CNN [55]	0.9625	0.7743	0.8803	<b>0.9308</b>	0.8875
Our CNN	0.9627	0.9255	0.8977	0.8875	0.9184
CNN features + GRNN	<b>0.9857</b>	<b>0.9473</b>	0.9059	0.9117	0.9377
CNN features + SVR	0.9730	0.9416	<b>0.9374</b>	0.9221	<b>0.9435</b>

<https://doi.org/10.1371/journal.pone.0176632.t002>

**Table 3. Performance evaluation of SROCC on Gaussian blurring images.**

	LIVE-II	CSIQ	TID2008	TID2013	Overall
JNB [35]	0.7872	0.7624	0.6667	0.6902	0.7266
CPBD [36]	0.9182	0.8853	0.8414	0.8518	0.8742
S3 [38]	0.9436	0.9059	0.8480	0.8609	0.8896
FISH [39]	0.8808	0.8941	0.7828	0.8024	0.8400
LPC [40]	0.9389	0.9071	0.8561	0.8888	0.8977
SVC [42]	0.9343	0.9055	0.8362	0.8589	0.8837
MLV [43]	0.9312	0.9247	0.8548	0.8787	0.8974
SPARISH [44]	0.9593	0.9141	0.8869	0.8927	0.9133
ARISM [45]	0.9511	0.9261	0.8505	0.8982	0.9065
BIBLE [46]	0.9607	0.9132	0.8915	0.8988	0.9160
Kang's CNN [55]	<b>0.9831</b>	0.7806	0.8496	<b>0.9215</b>	0.8837
Our CNN	0.9579	0.9048	0.8403	0.8376	0.8852
CNN features + GRNN	0.9744	0.9205	0.9163	0.9020	0.9283
CNN features + SVR	0.9646	<b>0.9253</b>	<b>0.9189</b>	0.9135	0.9310

<https://doi.org/10.1371/journal.pone.0176632.t003>

CSIQ. For the proposed methods, CNN features with GRNN or SVR shows advantage. In general, retrieved features with SVR reaches an average PLCC value of 0.9435, and CNN features with GRNN gets 0.9377, followed by BIBLE (0.9251) and SPARISH (0.9217). Our CNN achieves an average PLCC of 0.9184.

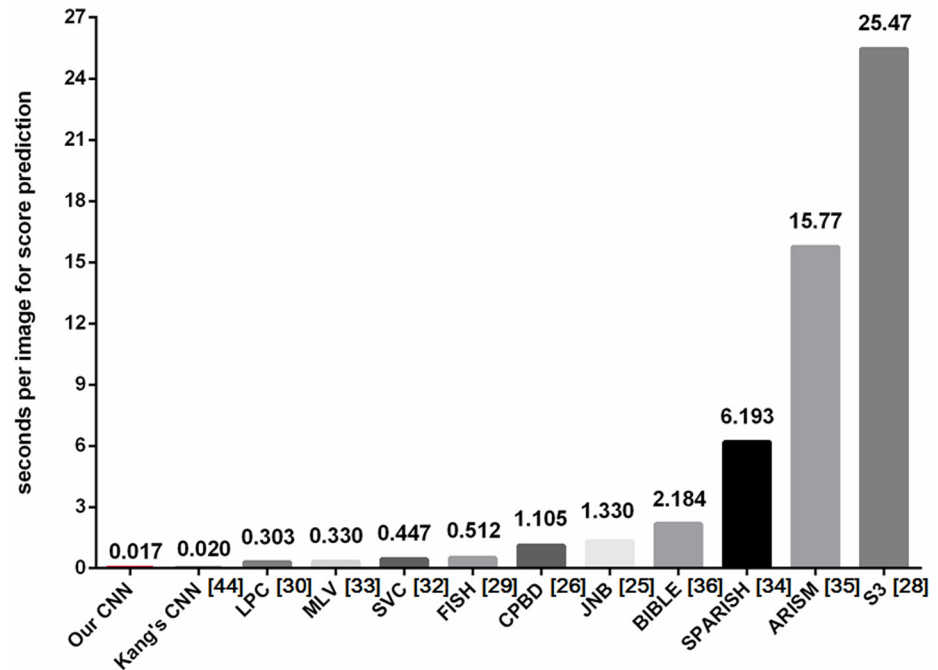
Table 3 shows SROCC and bolded values indicate best predication monotonicity. BIBLE [46] shows superiority over algorithms based on handcrafted features, followed by SPARISH [44] and ARISM [45]. Kang's CNN [55] achieves the highest SROCC on Gaussian blurring images from LIVE-II and TID2013, while it gets the second lowest SROCC on images from CSIQ among all metrics. On contrary, SROCC values from our CNN methods are robust on images from different databases. Particularly, CNN features with SVR outperforms other methods on CSIQ and TID2008. Furthermore, it ranks the second and the third place on TID2013 and LIVE-II, respectively. Generally, learned CNN features with SVR reaches an average SROCC of 0.9310, which is higher than CNN features with GRNN (0.9283), BIBLE (0.9160) and other methods.

### Time consumption

The time spent on score prediction of image sharpness is shown in Fig 8. Among traditional methods, several algorithms show promise in real-time image sharpness estimation, such as LPC, MLV, SVC and FISH which require less than 1 s. For CNN-based methods, both models take about 0.02 s to rate an image. It should be noted that the major time of CNN models is spent on local contrast normalization which costs about 8 s for an image. Moreover, GRNN and SVR need time after the model is well trained. Fortunately, with the help of code optimization and advanced hardware, it is feasible to accelerate these algorithms and to satisfy real time requirement.

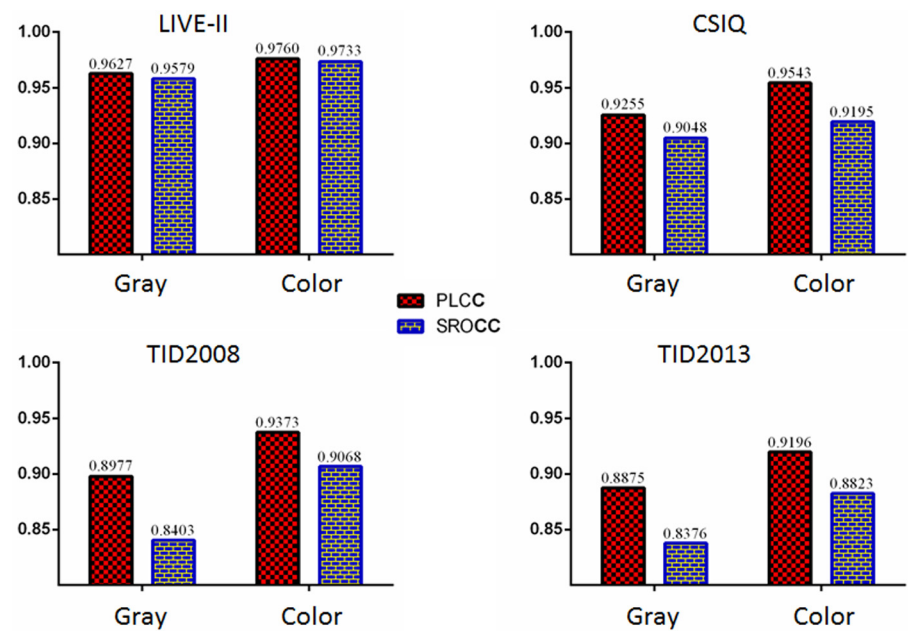
### Effect of color information

Chroma is an important underlying property of human vision system [64, 65] and it is highly correlated with image quality perception [30, 44]. Effect of color information on image sharpness estimation is studied with our CNN. The performance of CNN with gray and color inputs is shown in Fig 9. It is observed that chromatic information positively enhances CNN's



**Fig 8. The time spent on score prediction of image sharpness.** Several algorithms show promise in real-time image sharpness estimation.

<https://doi.org/10.1371/journal.pone.0176632.g008>



**Fig 9. Effect of color information on our CNN.** Compared to gray-scale input, color image input positively enhances our network's prediction metrics.

<https://doi.org/10.1371/journal.pone.0176632.g009>

performance on image sharpness estimation. The improved magnitude of PLCC ranges from 0.013 (LIVE-II) to 0.040 (TID2008). Meanwhile, the improved magnitude range of SROCC is from 0.014 (CSIQ) to 0.067 (TID2008).

## Future work

The proposed shallow CNN methods have achieved the state-of-the-art performance on simulated Gaussian blur images from four popular databases. Our future work will be to integrate handcrafted features and CNN features for improved prediction capacity. On the other hand, deeper networks will also be considered for representative features in image sharpness. In addition, with the public accessibility to the real-life blurring image databases of BID2011 [37] and CID2013 [66], it will be interesting to explore the proposed algorithm for more general and more practical applications [32, 67, 68].

## Conclusion

A shallow convolutional neural network is proposed to address blind image sharpness assessment. Its retrieved features with support vector regression achieves the best overall performance, indicating high correlation with subjective judgment. In addition, incorporating color information benefits image sharpness estimation with the shallow network.

## Acknowledgments

The authors would like to thank reviewers for their valuable advices that has helped to improve the paper quality. Thanks are also given to researchers who share their codes for fair comparison. This work is supported in part by grants from National Natural Science Foundation of China (Grant Nos. 81501463 and 61379143), Natural Science Foundation of Guangdong Province (Grant No. 2014A030310360), Major Project of Guangdong Province (Grant No. 2014B010111008), Guangdong Innovative Research Team Program (Grant No. 2011S013), the Qing Lan Project of Jiangsu Province, National 863 Programs of China (Grant No. 2015AA043203) and National Key Research Program of China (Grant No. 2016YFC0105102).

## Author Contributions

**Conceptualization:** SDY FJ YQX.

**Data curation:** SDY SBW FJ LDL.

**Formal analysis:** SDY LDL YQX LW.

**Funding acquisition:** SDY LDL YQX LW.

**Investigation:** LDL YQX.

**Methodology:** SDY FJ LDL YQX.

**Project administration:** SDY LDL.

**Resources:** SDY SBW LDL.

**Software:** SDY SBW FJ LDL.

**Supervision:** SDY LDL YQX.

**Validation:** SDY LDL.

**Visualization:** SDY FJ.

**Writing – original draft:** SDY SBW.

**Writing – review & editing:** LDL YQX LW.

## References

1. Lin W, Kuo CCJ. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*. 2011 Jan; 22(4): 297–312.
2. Manap RA, Shao L. Non-distortion-specific no-reference image quality assessment: A survey. *Information Sciences*. 2015 Dec; 301(1): 141–160.
3. Gao X, Gao F, Tao D, Li X. Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning. *IEEE Transactions on Neural Networks and Learning Systems*. 2013 Jul; 24(12): 2013–2026. <https://doi.org/10.1109/TNNLS.2013.2271356> PMID: 24805219
4. Li L, Lin W, Zhu H. Learning structural regularity for evaluating blocking artifacts in JPEG images. *IEEE Signal Processing Letters*. 2014 Aug; 21(8): 918–922. <https://doi.org/10.1109/LSP.2014.2320743>
5. Xue W, Mou X, Zhang L, Bovik AC, Feng X. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Transactions on Image Processing*. 2014 Nov; 23(11): 4850–4862. <https://doi.org/10.1109/TIP.2014.2355716> PMID: 25216482
6. Li L, Zhu H, Yang G, Qian J. Referenceless measure of blocking artifacts by Tchebichef kernel analysis. *IEEE Signal Processing Letters*. 2014 Jan; 21(1): 122–125. <https://doi.org/10.1109/LSP.2013.2294333>
7. Wu Q, Wang Z, Li H. A highly efficient method for blind image quality assessment. *IEEE Conference on Image Processing*. 2015 Sep; 1: 339–343.
8. Oszust M. Full-reference image quality assessment with linear combination of genetically selected quality measures. *PloS one*. 2016 Jun; 11(6):e0158333. <https://doi.org/10.1371/journal.pone.0158333> PMID: 27341493
9. Gu K, Li L, Lu H, Min X, Lin W. A fast computational metric for perceptual image quality assessment. *IEEE Transactions on Industrial Electronics*. 2017 Jan.
10. Sheikh HR, Sabir MF, Bovik AC. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*. 2006 Nov; 15(11):3440–3451. <https://doi.org/10.1109/TIP.2006.881959> PMID: 17076403
11. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004 Apr; 13(4):600–612. <https://doi.org/10.1109/TIP.2003.819861> PMID: 15376593
12. Zhang L, Zhang L, Mou X, Zhang D. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*. 2011 Aug; 20(8):2378–2386. <https://doi.org/10.1109/TIP.2011.2109730> PMID: 21292594
13. Qian J, Wu D, Li L, Cheng D, Wang X. Image quality assessment based on multi-scale representation of structure. *Digital Signal Processing*. 2014 Oct; 33:125–133. <https://doi.org/10.1016/j.dsp.2014.06.009>
14. Zhou F, Lu Z, Wang C, Sun W, Xia ST, Liao Q. Image quality assessment based on inter-patch and intra-patch similarity. *PloS one*. 2015 Mar; 10(3):e0116312. <https://doi.org/10.1371/journal.pone.0116312> PMID: 25793282
15. Yuan H, Kwong S, Wang X, Zhang Y, Li F. A virtual view PSNR estimation method for 3-D videos. *IEEE Transactions on Broadcasting*. 2016 Mar; 62(1):134–140. <https://doi.org/10.1109/TBC.2015.2492461>
16. Yang Y, Wang X, Liu Q, Xu ML, Wu W. User models of subjective image quality assessment of virtual viewpoint in free-viewpoint video system. *Multimedia Tools and Applications*. 2016 Oct; 75(20):12499–12519. <https://doi.org/10.1007/s11042-014-2321-7>
17. Chen L, Jiang F, Zhang H, Wu S, Yu S, Xie Y. Edge preservation ratio for image sharpness assessment. *IEEE World Congress on Intelligent Control and Automation*. 2016 Jun; 1:1377–1381.
18. Wang Z, Bovik AC. Reduced- and no-reference image quality assessment. *IEEE Signal Processing Magazine*. 2011 Nov; 28(6):29–40. <https://doi.org/10.1109/MSP.2011.942471>
19. Soundararajan R, Bovik AC. RRED indices: Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Image Processing*. 2012 Feb; 21(2):517–526. <https://doi.org/10.1109/TIP.2011.2166082> PMID: 21878414
20. Wu J, Lin W, Shi G. Reduced-reference image quality assessment with visual information fidelity. *IEEE Transactions on Multimedia*. 2013 Feb; 15(7):1700–1705. <https://doi.org/10.1109/TMM.2013.2266093>
21. Wang X, Liu Q, Wang R, Chen Z. Natural image statistics based 3D reduced reference image quality assessment in Contourlet domain. *Neurocomputing*. 2015 Mar; 151(2):683–691.

22. Ma L, Wang X, Liu Q, Ngan KN. Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment. *Neurocomputing*. 2016 Nov; 215:21–31. <https://doi.org/10.1016/j.neucom.2015.06.116>
23. Moorthy AK, Bovik AC. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*. 2011 Dec; 20(12): 3350–3364. <https://doi.org/10.1109/TIP.2011.2147325> PMID: 21521667
24. Saad MA, Bovik AC, Charrier C. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE transactions on Image Processing*. 2012 Aug; 21(8): 3339–3352. <https://doi.org/10.1109/TIP.2012.2191563> PMID: 22453635
25. Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*. 2012 Dec; 21(12): 4695–4708. <https://doi.org/10.1109/TIP.2012.2214050> PMID: 22910118
26. Gao F, Tao D, Gao X, Li X. Learning to rank for blind image quality assessment. *IEEE Transactions on Neural Networks and Learning Systems*. 2015 Oct; 26(10): 2275–2290. <https://doi.org/10.1109/TNNLS.2014.2377181> PMID: 25616080
27. Zhang L, Zhang L, Bovik AC. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*. 2015 Aug; 24(8): 2579–2591. <https://doi.org/10.1109/TIP.2015.2426416> PMID: 25915960
28. Wu Q, Li H, Meng F, Ngan KN, Zhu S. No reference image quality assessment metric via multi-domain structural information and piecewise regression. *Journal of Visual Communication and Image Representation*. 2015 Oct; 32: 205–216. <https://doi.org/10.1016/j.jvcir.2015.08.009>
29. Gu K, Zhai G, Yang X, Zhang W. Using free energy principle for blind image quality assessment. *IEEE Transactions on Multimedia*. 2015 Jan; 17(1): 50–63. <https://doi.org/10.1109/TMM.2014.2373812>
30. Wu Q, Li H, Meng F, Ngan KN, Luo B, Huang C, et al. Blind image quality assessment based on multi-channel feature fusion and label transfer. *IEEE Transactions on Circuits and Systems for Video Technology*. 2016 Mar; 26(3): 425–440. <https://doi.org/10.1109/TCSVT.2015.2412773>
31. Li L, Zhou Y, Lin W, Wu J, Zhang X, Chen B. No-reference quality assessment of deblocked images. *Neurocomputing*. 2016 Feb; 177: 572–584. <https://doi.org/10.1016/j.neucom.2015.11.063>
32. Gu K, Zhai G, Lin W, Liu M. The analysis of image contrast: From quality assessment to automatic enhancement. *IEEE Transactions on Cybernetics*. 2016 Jan; 46(1): 284–297. <https://doi.org/10.1109/TCYB.2015.2401732> PMID: 25775503
33. Zhang C, Pan J, Chen S, Wang T, Sun D. No reference image quality assessment using sparse feature representation in two dimensions spatial correlation. *Neurocomputing*. 2016 Jan; 173: 462–470. <https://doi.org/10.1016/j.neucom.2015.01.105>
34. Wang S, Deng C, Lin W, Huang G, Zhao B. NMF-based image quality assessment using extreme learning machine. *IEEE Transactions on Cybernetics*. 2017 Jan; 47(1): 232–243. <https://doi.org/10.1109/TCYB.2015.2512852> PMID: 26863686
35. Ferzli R, Karam LJ. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE Transactions on Image Processing*. 2009 Aug; 18(4):717–728. <https://doi.org/10.1109/TIP.2008.2011760> PMID: 19278916
36. Narvekar ND, Karam LJ. A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). *IEEE Transactions on Image Processing*. 2011 Sep; 20(9):2678–2683. <https://doi.org/10.1109/TIP.2011.2131660> PMID: 21447451
37. Ciancio A, da Costa ALANT, da Silva EAB, Said A, Samadani R, Obrador P. No-reference blur assessment of digital pictures based on multifeature classifiers. *IEEE Transactions on Image Processing*. 2011 Jan; 20(1):64–75. <https://doi.org/10.1109/TIP.2010.2053549> PMID: 21172744
38. Vu CT, Phan TD, Chandler DM. S3: A spectral and spatial measure of local perceived sharpness in natural images. *IEEE Transactions on Image Processing*. 2012 Jan; 21(3):934–945. <https://doi.org/10.1109/TIP.2011.2169974> PMID: 21965207
39. Vu PV, Chandler DM. A fast wavelet-based algorithm for global and local image sharpness estimation. *IEEE Signal Processing Letters*. 2012 Jul; 19(7):423–426. <https://doi.org/10.1109/LSP.2012.2199980>
40. Hassen R, Wang Z, Salama MM. Image sharpness assessment based on local phase coherence. *IEEE Transactions on Image Processing*. 2013 Jul; 22(7):2798–2810. <https://doi.org/10.1109/TIP.2013.2251643> PMID: 23481852
41. Sang QB, Wu XJ, Li CF, Lu Y. Blind image blur assessment using singular value similarity and blur comparisons. *PloS one*. 2014 Sep; 9(9): e108073. <https://doi.org/10.1371/journal.pone.0108073> PMID: 25247555

42. Sang Q, Qi H, Wu X, Li C, Bovik AC. No-reference image blur index based on singular value curve. *Journal of Visual Communication and Image Representation*. 2014 Oct; 25(7):1625–1630. <https://doi.org/10.1016/j.jvcir.2014.08.002>
43. Bahrami K, Kot AC. A fast approach for no-reference image sharpness assessment based on maximum local variation. *IEEE Signal Processing Letters*. 2014 Jun; 21(6):751–755. <https://doi.org/10.1109/LSP.2014.2314487>
44. Li L, Wu D, Wu J, Li H, Lin W, Kot AC. Image sharpness assessment by sparse representation. *IEEE Transactions on Multimedia*. 2016 Jun; 18(6):1085–1097. <https://doi.org/10.1109/TMM.2016.2545398>
45. Gu K, Zhai G, Lin W, Yang X, Zgagb W. No-reference image sharpness assessment in autoregressive parameter space. *IEEE Transactions on Image Processing*. 2015 Oct; 24(10):3218–3231. <https://doi.org/10.1109/TIP.2015.2439035> PMID: 26054063
46. Li L, Lin W, Wang X, Yang G, Bahrami K, Kot AC. No-reference image blur assessment based on discrete orthogonal moments. *IEEE Transactions on Cybernetics*. 2016 Jan; 46(1):39–50. <https://doi.org/10.1109/TCYB.2015.2392129> PMID: 25647763
47. Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013 Aug; 35(8): 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50> PMID: 23787338
48. LeCun Y, Bengio Y, Hinton GE. Deep learning. *Nature*. 2015 May; 521:436–444. <https://doi.org/10.1038/nature14539> PMID: 26017442
49. Li Y, Po L, Xu X, Feng L, Yuan F, Cheung C, et al. No-reference image quality assessment with shearlet transform and deep neural networks. *Neurocomputing*. 2015 Apr; 154: 94–109. <https://doi.org/10.1016/j.neucom.2014.12.015>
50. Hou W, Gao X. Saliency-guided deep framework for image quality assessment. *IEEE Multimedia*. 2015 Jul; 22(2): 46–55. <https://doi.org/10.1109/MMUL.2014.55>
51. Li J, Zou L, Yan J, Deng D, Qu T, Xie G. No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks. *Signal, Image and Video Processing*. 2016 Apr; 10(4): 609–616. <https://doi.org/10.1007/s11760-015-0784-2>
52. Lv Y, Jiang G, Yu M, Xu H, Shao F, Liu S. Difference of Gaussian statistical features based blind image quality assessment: A deep learning approach. *IEEE Conference on Image Processing*. 2015 Sep; 1: 2344–2348.
53. Hou W, Gao X, Tao D, Li X. Blind image quality assessment via deep learning. *IEEE Transactions on Neural Networks and Learning Systems*. 2015 Aug; 26(6): 1275–1286. <https://doi.org/10.1109/TNNLS.2014.2336852> PMID: 25122842
54. Yu S, Jiang F, Li L, Xie Y. CNN-GRNN for image sharpness assessment. *Asian Conference on Computer Vision*. 2016 Oct; 1: 50–61.
55. Kang L, Ye P, Li Y, Doermann D. Convolutional neural networks for no-reference image quality assessment. *IEEE Conference on Computer Vision and Pattern Recognition*. 2014 Jun; 1: 1733–1740.
56. Specht DF. A general regression neural network. *IEEE Transactions on Neural Networks*. 1991 Nov; 2(6):568–576. <https://doi.org/10.1109/72.97934> PMID: 18282872
57. Basak D, Pal S, Patranabis DC. Support vector regression. *Neural Information Processing—Letters and Reviews*. 2007 Oct; 11(10): 203–224.
58. Ruderman DL. The statistics of natural images. *Network: Computation in Neural Systems*. 1994 Jul; 5(4): 517–548. [https://doi.org/10.1088/0954-898X\\_5\\_4\\_006](https://doi.org/10.1088/0954-898X_5_4_006)
59. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011 Apr; 2(3): 27. <https://doi.org/10.1145/1961189.1961199>
60. Narwaria M, Lin W. Objective image quality assessment based on support vector regression. *IEEE Transactions on Neural Networks*. 2010 Mar; 21(3):515–519. <https://doi.org/10.1109/TNN.2010.2040192> PMID: 20100674
61. Larson EC, Chandler DM. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*. 2010 Jul; 19(1): 11006.
62. Ponomarenko N, Lukin V, Zelensky A, Egiazarian K, Astola J, Carli M, et al. TID2008—A database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*. 2009 Apr; 10(4): 30–45.
63. Ponomarenko N, Jin J, Ieremeiev O, Lukin V, Egiazarian K, Astola J, et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*. 2015 Jan; 20: 57–77. <https://doi.org/10.1016/j.image.2014.10.009>
64. Solomon SG, Lennie P. The machinery of colour vision. *Nature Reviews Neuroscience*. 2007 Apr; 8(4): 276–286. <https://doi.org/10.1038/nrn2094> PMID: 17375040

65. Van De Sande K, Gevers T, Snoek C. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010 Sep; 32(9): 1582–1596. <https://doi.org/10.1109/TPAMI.2009.154> PMID: 20634554
66. Virtanen T, Nuutinen M, Vaahteranoksa M, Oittinen P, Hakkinen J. CID2013: A database for evaluating no-reference image quality assessment algorithms. *IEEE Transactions on Image Processing*. 2015 Jan; 24(1): 390–402. <https://doi.org/10.1109/TIP.2014.2378061> PMID: 25494511
67. Li L, Xia W, Lin W, Fang Y, Wang S. No-reference and robust image sharpness evaluation based on multi-scale spatial and spectral features. *IEEE Transactions on Multimedia*. 2016 Dec.
68. Chow LS, Rajagopal H, Paramesran R, Alzheimer's Disease Neuroimaging Initiative. Correlation between subjective and objective assessment of magnetic resonance (MR) images. *Magnetic Resonance Imaging*. 2016 Jul; 34(6): 820–831. <https://doi.org/10.1016/j.mri.2016.03.006> PMID: 26969762