# No-Reference Image Quality Assessment Using Shearlet Transform and Stacked Autoencoders

Yuming Li, Lai-Man Po, Xuyuan Xu, Litong Feng, Fang Yuan, Chun-Ho Cheung*, Kwok-Wai Cheung**

Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China
*Department of Information Systems, City University of Hong Kong, Hong Kong, China
**Department of Computer Science, Chu Hai College of Higher Education, Hong Kong, China

*Abstract*—**In this work, we describe an efficient general-purpose no-reference (NR) image quality assessment (IQA) algorithm that is based on a new multiscale directional transform (shearlet transform) with a strong ability to localize distributed discontinuities. The algorithm relies on utilizing the sum of subband coefficient amplitudes (SSCA) as primary features to describe the behavior of natural images and distorted images. Then, stacked autoencoders are applied to exaggerate the discriminative parts of the primary features. Finally, by translating the NR-IQA problem into classification problem, the differences of evolved features are identified by softmax classifier. The resulting algorithm, which we name SESANIA (ShEarlet and Stacked Autoencoders based No-reference Image quality Assessment), is tested on several databases (LIVE, Multiply Distorted LIVE and TID2008) and shown to be suitable to many common distortions, consistent with subjective assessment and comparable to full-reference IQA methods and state-of-the-art general purpose NR-IQA algorithms.**

***Keywords - No-reference image quality assessment; Shearlet transform; Stacked autoencoders; Softmax classification***

## I. INTRODUCTION

The ability to quantify the visual quality of an image or video is a vital step for many systems that process digital media. Automatic image or video quality measurement methods aim to assign quality scores to images or videos which are in consistent with human perception. Over the last few decades, numerous image quality assessment (IQA) algorithms have been developed and can be divided into three types: full-reference (FR), reduced-reference (RR) and no-reference (NR). In FR-IQA and RR-IQA methods, the whole reference images or partial information of the reference images are assumed to be available. However, in many practical applications, it is hard to obtain the information about the original images. Therefore, NR-IQA method is required. Recently, most successful NR-IQA approaches can be generally classified as: 1) Distortion-specific approach [1]; 2) Natural scene statistics (NSS) based approach [2-5]; and 3) Training-based approach [6].

Based on these observations, we developed SESANIA (ShEarlet and Stacked Autoencoders based No-reference Image quality Assessment) that is a combination of second and third trends. It is also a general purpose NR-IQA method that can estimate a wide range of image distortions. However, different from our previous works [7], SESANIA does not directly use the property of NSS model in shearlet domain to construct a predictor, but utilize the sum of subband coefficient amplitudes (SSCA) as primary features to describe the behavior of natural images and distorted images. The main idea of SESANIA is based on the finding that the statistical property of most natural images in shearlet domain is relatively constant. Nevertheless, distorted images usually contain more or less spread discontinuities in all directions. Particularly, for natural images, the SSCA in different scales has relatively constant relationship in shearlet domain. However, this constant relationship will be disturbed if a natural image is distorted by some common distortions. Motivated by this idea, SSCA can act as a feature descriptor to describe an image. Thus, natural images and distorted images can be distinguished by these primary features. To exaggerate the discriminative parts of the primary features, we propose to apply stacked autoencoders as ‘evolution process’ to ‘amplify’ the primary features and make them more discriminative. Finally, by translating the NR-IQA problem into classification problem, the differences of evolved features can be easily identified by Softmax classifier. In the implementation process, SESANIA does not incorporate any prior knowledge about distortions, which makes it suitable to many distortions and easy to be extended.

The remainder of the paper is organized as follows. Section 2 introduces the detailed implementation and related techniques about SESANIA. In Section 3, experimental results and related analysis are given. Finally, conclusion and future works are presented in Section 4.

## II. METHIDOLOGY

The framework of the proposed approach is illustrated in Fig. 1. The major components in this framework include: 1) SSCA extraction in shearlet domain, 2) Feature evolution using stacked autoencoders, 3) Evolved feature identification using softmax classifier, 4) Quality score calculation. More details will be described in the following sections.

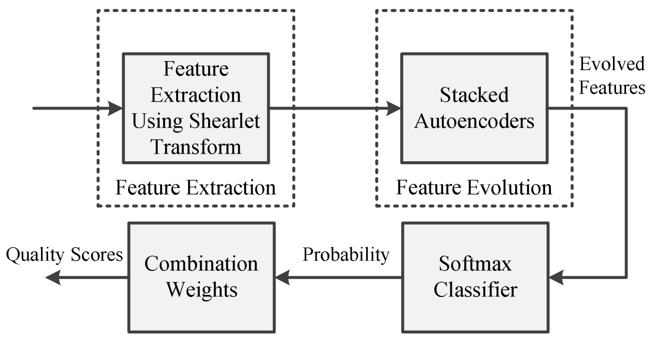


Figure 1. Overview of the SESANIA framework.

978-1-4799-8391-9/15/$31.00 ©2015 IEEE

### A. ***Shearlet Transform***

The new NR-IQA method proposed in this paper is based on the shearlet transform [8]. This multiscale transform is a multidimensional edition of the traditional wavelet transform, and is capable for addressing anisotropic and directional information at different scales. When the dimension $n=2$, the affine systems with composite dilations are the collections of the form:

$$SH_\phi f(a,s,t) = \langle f, \phi_{a,s,t} \rangle, a>0, s\in R, t\in R^2 \quad (1)$$

where the analyzing factor $\phi_{a,s,t}$ is called shearlet coefficient, which is defined as:

$$\phi_{a,s,t}(x) = |\det M_{a,s}|^{-\frac{1}{2}}\phi(M_{a,s}^{-1}x-t) \quad (2)$$

where $M_{a,s}=B_s A_a=\begin{pmatrix} a & \sqrt{as} \\ 0 & \sqrt{a}\end{pmatrix}$, and $A_a=\begin{pmatrix} a & 0 \\ 0 & \sqrt{a}\end{pmatrix}$, $B_s=\begin{pmatrix} 1 & s \\ 0 & 1\end{pmatrix}$. $A_a$ is the anisotropic dilation matrix and $B_s$ is the shear matrix. The analyzing functions associated to the shearlet transform are anisotropic and are defined at different scales, locations and orientations. Thus, shearlets have the ability to detect directional information and account for the geometry of multidimensional functions, which overcome the limitation of the wavelet transform.

### B. ***Feature Extraction***

The normalized sum of subband coefficient amplitudes (SSCA) is used to express this property in the shearlet domain, which is defined as

$$SSCA(a,s)=\frac{\sum_t |SH_\phi f(a,s,t)|}{\max\left(\sum_t |SH_\phi f(a,s,t)|\right)} \quad (3)$$

where, $SH_\phi f(a,s,t)$ is the shearlet transform of an image and $a$ is scale parameter, $s$ is direction parameter and $t$ is time shift. Fig. 2 plots the mean SSCA for grayscale images in logarithmic coordinates, which is generated by all the 29 original images and their associated distorted versions in laboratory for image and video engineering (LIVE) database [9]. Every original image and distorted image are randomly sampled several times by the size of 256 × 256. Totally 12,000 sampled blocks are obtained and 2,000 for each type. Shearlet transform with 4 scales and 6 directions for each scale is applied to each of the sampled blocks, and SSCA is calculated. It can be seen from Fig. 2 that common distortions disturb image statistics and make statistical property vary from that of natural images in shearlet domain. Besides, SSCA in fine scales increase or decrease monotonously with the reduction in image quality.

### C. ***Feature Evolution***

The SSCA extracted from images can serve as features to distinguish natural images and distorted images. However, an intuitive idea is before sending this primary feature into classifier, whether we can design a system to 'amplify' the difference between natural image features and distorted image features. Recent works about deep neural networks provide us some ideas to solve this problem. In this paper, we propose to use stacked autoencoders to serve as an amplifier to increase the distance between natural image features and distorted image features and make them more discriminative. A stacked autoencoder is a neural network consisting of multiple layers of sparse autoencoders in which the outputs of each layer are wired to the inputs of the successive layer. In our work, feature evolution process is shown in Fig. 3. In this process, the primary features are a vector which contains SSCA from RGB channel and is normalized. The primary features are evolved in stacked autoencoders and the final evolved features are sent to Softmax classifier.

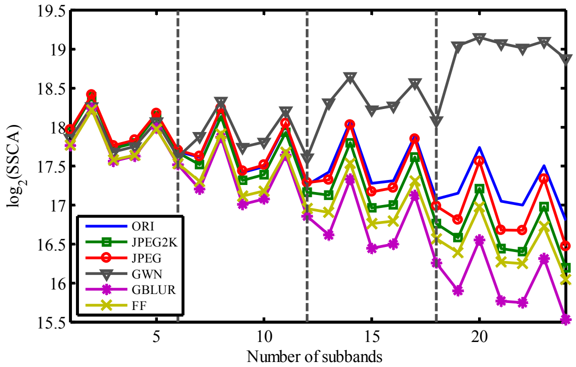


Figure 2. Mean SSCA versus subband enumeration index for natural images and different distorted images in LIVE IQA database.

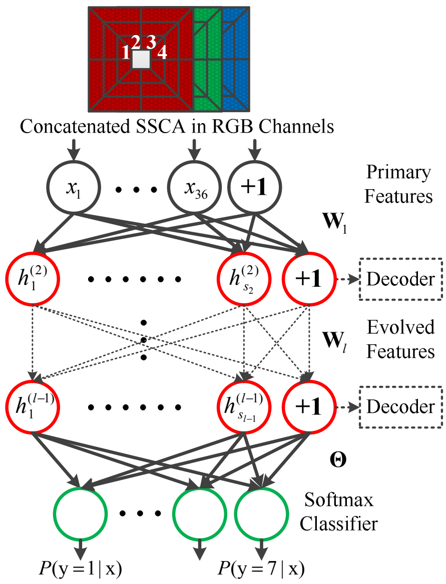


Figure 3. Feature evolution process. The final evolved features are sent to Softmax classifier (Green circle). The red circle indicates sigmoid function. Dashed decoder means decoders are discarded after pretraining.

### D. ***Classification and Quality Evaluation***

In section, the NR-IQA is translated into classification problem and Softmax classifier is applied to identify the image quality through final evolved features. Now, we have a training set $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ of $m$ labeled data, where the input features are $x^{(i)}$, which is the final evolved features from the last layer of stacked autoencoder. $y^{(i)} \in \{1, 2, \dots, 7\}$ is the label of a training image, which is obtained based on Table I. Given an image feature $x^{(i)}$, the output of a Softmax classifier is

$$P_\theta(x^{(i)}) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{K} e^{\theta_l^T x^{(i)}}} \tag{4}$$

where, $\theta$ are the parameters of Softmax classifier and $K$ is the number of classes. The final quality score of an image is calculated by:

$$Q_i = \omega \cdot P_\theta^T(x^{(i)}) \tag{5}$$

where the combining weights $\omega$ are learned by calculating the least square solution of this over determined equation upon the training set.

TABLE I. THE RELATIONSHIP BETWEEN IMAGE MOS AND ITS LABEL

| **MOS** | **< 25** | **25-34** | **35-44** | **45-54** | **55-64** | **65-74** | **> 75** |
|---|---|---|---|---|---|---|---|
| **Class** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

## III. EXPERIMENTS AND RELATED ANALYSIS

To train, test and compare the proposed NR-IQA algorithm, the following three IQA databases were used: 1) LIVE IQA database; 2) LIVE Multiply Distorted IQA database [10]; and 3) TID2008 database [11]. We tabulated the performance of three FR-IQA and four state-of-art NR-IQA algorithms. Three FR-IQA methods include: peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM) and multi-scale structural similarity index (MS-SSIM). Four general purpose NR-IQA methods include: BIQI [2], BLIINDS-II [3], BRISQUE [4] and DIIVINE [5]. When testing on these databases, we divide each database into two randomly chosen subsets and there is no overlap between train and test sets. The training subsets contain 80% original images and their distorted versions and the testing subsets contain 20%. This train-test procedure is repeated 1000 times, and the median of the performance across these 1000 iterations is reported. For SESANIA, Fourier based shearlet transform is applied and the RGB channel of an image is decomposed into 4 scales (exclude approximation component) and every scale has 6 directions. The number of primary features is 72. The layer number of the stacked autoencoder is 3 and each layer has 64 neurons. Table II, Table III and Table IV report the median LCC and SROCC results tested on three databases respectively. Fig. 4(a) and (b) show the box plot of LCC and SROCC distributions across 1000 trials of experiments on the LIVE IQA database. It can be seen that SESANIA achieves comparable testing results and approaches the performance of the reliable FR-IQA methods and state-of-art general purpose NR-IQA methods on LIVE database. For LIVE Multiply Distorted IQA database and TID2008, SESANIA still achieves acceptable testing results.

## IV. APPLICATION TO BLIND IMAGE DENOISING

In this section, we will use SESANIA to automatically determine the threshold of shearlet denoising algorithm and guides this algorithm to recover the original image from Gaussian noise. One simple shearlet denoising algorithm for additive Gaussian noise is hard threshold method, which is defined as

$$SH\,f[s,d,k] = \begin{cases} SH\,f[s,d,k], & \text{if } \left|SH\,f[s,d,k]\right| > T \times E[s,d] \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $T$ is the threshold and $E[s,d]$ is the $l^2$ norm of shearlets. In this application, we assume the standard variance $\sigma$ of Gaussian noise is unknown. To get the value of $T$, we can consider to optimize an objective function and obtain the optimized threshold $T^*$. If the original image has already obtained, the objective function can be any FR-IQA method, in this application, we use PSNR as the FR-IQA and the objective function can be expressed as

$$T^*_{PSNR} = \underset{T}{argmax}\ \mathrm{PSNR}(T, f_{ori}, f_{noi}) \tag{7}$$

where $f_{ori}$ and $f_{noi}$ indicate original image and noisy image, respectively. However, if the original image cannot be obtained, we use SESANIA as objective function and the objective function can be expressed as

$$T^*_{SESANIA} = \underset{T}{argmax}\ \mathrm{SESANIA}(T, f_{noi}) \tag{8}$$

Lena acts as the original image. Gaussian noise with randomly generated $\sigma$ (in the range of 10 to 40) is added to it, totally 100 noisy images are obtained. SSIM between denoised image and original image is calculated. Fig. 4(c) shows the box plot of SSIM of these 100 images before and after denoising. It can be seen that PSNR shows good performance, since it has original image as reference. However, without the reference image, the performance of SESANIA is still good. It also guides the denoising algorithm to find a reasonable threshold and helps to improve the final SSIM between denoised image and original image.

## V. CONCLUSION

In this paper, we proposed a general-purpose NR-IQA algorithm SESANIA, which is based on the statistical characterization in the shearlet domain. The general idea about SESANIA is to quantify the naturalness of an image and no prior information about distortions is incorporated in this algorithm. SESANIA is tested on LIVE database, LIVE Multiply distorted database and TID2008 database. It has shown good performance on these databases and is comparative to FR-IQA methods and state-of-art general-purpose NR-IQA algorithms.


## ACKNOWLEDGMENT

*The work described in this paper was substantially supported by a grant from the City University of Hong Kong, Kowloon, Hong Kong with Project number of 7004058.*

## REFERENCES

[1] Sheikh, Hamid R., Alan C. Bovik, and Lawrence Cormack. "No-reference quality assessment using natural scene statistics: JPEG2000." Image Processing, IEEE Transactions on 14, no. 11 (2005): 1918-1927.

[2] Moorthy, Anush Krishna, and Alan Conrad Bovik. "A two-step framework for constructing blind image quality indices." Signal Processing Letters, IEEE 17, no. 5 (2010): 513-516.

[3] Saad, Michele A., Alan C. Bovik, and Christophe Charrier. "Blind image quality assessment: A natural scene statistics approach in the DCT domain." Image Processing, IEEE Transactions on 21, no. 8 (2012): 3339-3352.

[4] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain." Image Processing, IEEE Transactions on 21, no. 12 (2012): 4695-4708.

[5] Moorthy, Anush Krishna, and Alan Conrad Bovik. "Blind image quality assessment: From natural scene statistics to perceptual quality." Image Processing, IEEE Transactions on 20, no. 12 (2011): 3350-3364.

[6] Ye, Peng, Jayant Kumar, Le Kang, and David Doermann. "Unsupervised feature learning framework for no-reference image quality assessment." In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 1098-1105. IEEE, 2012.

[7] Li, Yuming, Lai-Man Po, Xuyuan Xu, and Litong Feng. "No-reference image quality assessment using statistical characterization in the shearlet domain." Signal Processing: Image Communication 29, no. 7 (2014): 748-759.

[8] Easley, Glenn, Demetrio Labate, and Wang-Q. Lim. "Sparse directional image representations using the discrete shearlet transform." Applied and Computational Harmonic Analysis 25, no. 1 (2008): 25-46.

[9] Sheikh, Hamid R., Zhou Wang, Alan C. Bovik, and L. K. Cormack. "Image and video quality assessment research at LIVE." (2003). http://live.ece.utexas.edu/research/quality/.

[10] Ayaraman, Dinesh, Anish Mittal, Anush K. Moorthy, and Alan C. Bovik. "Objective quality assessment of multiply distorted images." In Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on, pp. 1693-1697. IEEE, 2012.

[11] Ponomarenko, Nikolay, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, M. Carli, and F. Battisti. "TID2008-A database for evaluation of full-reference visual quality assessment metrics." Advances of Modern Radioelectronics 10, no. 4 (2009): 30-45.


TABLE II. MEDIAN LCC AND SROCC ON LIVE IQA DATABASE. (ITALICIZED ALGORITHMS ARE NR-IQA ALGORITHMS.)

| | LCC | | | | | | | SROCC | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **JPEG2K** | **JPEG** | **GWN** | **GBLUR** | **FF** | **ALL** | | **JPEG2K** | **JPEG** | **GWN** | **GBLUR** | **FF** | **ALL** |
| PSNR | 0.8692 | 0.8402 | 0.9412 | 0.8318 | 0.8569 | 0.8588 | | 0.8483 | 0.8314 | 0.8964 | 0.8393 | 0.8354 | 0.8461 |
| SSIM | 0.9048 | 0.8303 | 0.9321 | 0.8823 | 0.8538 | 0.8823 | | 0.9363 | 0.9730 | 0.9638 | 0.9758 | 0.9832 | 0.9719 |
| MSSSIM | 0.8807 | 0.8062 | 0.8841 | 0.8298 | 0.8004 | 0.7939 | | 0.9460 | 0.9742 | 0.9660 | 0.9770 | 0.9795 | 0.9795 |
| *BIQI* | 0.8086 | 0.9011 | 0.9538 | 0.8293 | 0.7328 | 0.8205 | | 0.7995 | 0.8914 | 0.9510 | 0.8463 | 0.7067 | 0.8195 |
| *DIIVINE* | 0.9220 | 0.9210 | 0.9880 | 0.9230 | 0.8680 | 0.9170 | | 0.9319 | 0.9483 | 0.9821 | 0.9210 | 0.8714 | 0.9116 |
| *BLIINDS-II* | 0.9386 | 0.9426 | 0.9635 | 0.8994 | 0.8790 | 0.9164 | | 0.9323 | 0.9331 | 0.9463 | 0.8912 | 0.8519 | 0.9124 |
| *BRISQUE* | 0.9229 | 0.9734 | 0.9851 | 0.9506 | 0.9030 | 0.9424 | | 0.9139 | 0.9647 | 0.9786 | 0.9511 | 0.8768 | 0.9395 |
| *SESANIA* | 0.9350 | 0.9295 | 0.9861 | 0.9805 | 0.9480 | 0.9481 | | 0.8914 | 0.9105 | 0.9693 | 0.9761 | 0.9179 | 0.9189 |

TABLE III. MEDIAN LCC AND SROCC ON LIVE MULTIPLY DISTORTED IQA DATABASE. (ITALICIZED ALGORITHM IS NR-IQA ALGORITHMS.)

| | LCC | | | SROCC | |
|---|---|---|---|---|---|
| | **Part 1 (GBLUR + JPEG)** | **Part 2 (GBLUR + GWN)** | | **Part 1 (GBLUR + JPEG)** | **Part 2 (GBLUR + GWN)** |
| PSNR | 0.7679 | 0.8083 | | 0.7183 | 0.7551 |
| SSIM | 0.8376 | 0.8566 | | 0.8745 | 0.9028 |
| MSSSIM | 0.8361 | 0.8698 | | 0.8743 | 0.8887 |
| *SESANIA* | 0.7888 | 0.8197 | | 0.8470 | 0.8698 |

TABLE IV. MEDIAN LCC AND SROCC ON TID2008 DATABASE. (ITALICIZED ALGORITHM IS NR-IQA ALGORITHMS.)

| | LCC | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **WN** | **WNC** | **SCN** | **MN** | **HFN** | **IN** | **QN** | **GBLUR** | **IDN** | **JPEG** | **JPEG2K** | **JPEGTE** | **JP2KTE** | **ALL** |
| PSNR | 0.9416 | 0.9218 | 0.9575 | 0.8795 | 0.9698 | 0.8559 | 0.8876 | 0.9271 | 0.9446 | 0.8681 | 0.8814 | 0.6365 | 0.8484 | 0.7528 |
| SSIM | 0.7576 | 0.7783 | 0.7805 | 0.7525 | 0.8263 | 0.6206 | 0.7462 | 0.8906 | 0.9276 | 0.9469 | 0.9472 | 0.8390 | 0.8357 | 0.8160 |
| MSSSIM | 0.7546 | 0.7817 | 0.7714 | 0.7826 | 0.8247 | 0.6223 | 0.7807 | 0.8721 | 0.9211 | 0.9353 | 0.9361 | 0.8154 | 0.8073 | 0.8213 |
| *SESANIA* | 0.9400 | 0.8427 | 0.9251 | 0.8944 | 0.9417 | 0.8350 | 0.9229 | 0.9701 | 0.9568 | 0.9557 | 0.9412 | 0.9402 | 0.8450 | 0.8252 |
| | SROCC | | | | | | | | | | | | | |
| PSNR | 0.9177 | 0.8962 | 0.9245 | 0.8747 | 0.9318 | 0.8721 | 0.8843 | 0.9335 | 0.9418 | 0.8753 | 0.8248 | 0.7727 | 0.8349 | 0.7813 |
| SSIM | 0.8168 | 0.7997 | 0.8218 | 0.7956 | 0.8695 | 0.6712 | 0.8815 | 0.9598 | 0.9638 | 0.9354 | 0.9604 | 0.8761 | 0.8649 | 0.8760 |
| MSSSIM | 0.8184 | 0.8042 | 0.8281 | 0.8138 | 0.8677 | 0.6888 | 0.8827 | 0.8721 | 0.9690 | 0.9371 | 0.9361 | 0.8742 | 0.8668 | 0.8840 |
| *SESANIA* | 0.9296 | 0.8261 | 0.9270 | 0.7343 | 0.9264 | 0.8924 | 0.9279 | 0.9682 | 0.9229 | 0.9327 | 0.9212 | 0.9249 | 0.8669 | 0.8433 |

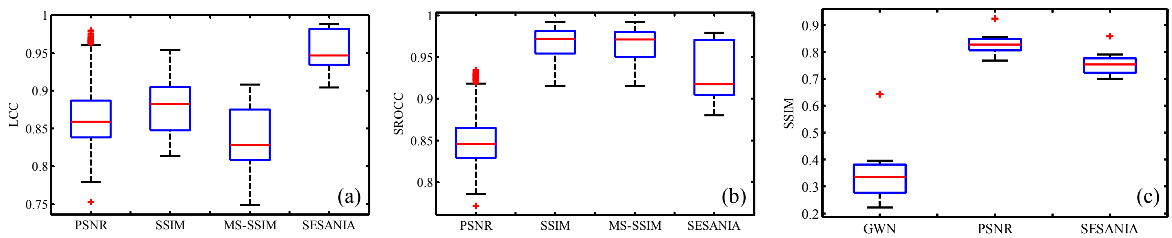


Figure 4. Box plot of LCC and SROCC distributions on LIVE IQA database and SSIM of 100 noisy images before and after denoising. (a) Box plot of LCC distribution. (b) Box plot of SROCC distribution. (b) Box plot of SROCC distribution. (c) Box plot of SSIM of 100 noisy images before and after denoising.