Improved hybrid blind IQA using alternative NSS characterization in the spatial domain

Antonis Mairgiotis
Dept. of Computer Science and Tel.
University of Thessaly
Lamia, Greece
mairgiotis@gmail.com

Dimitra Tsampra
Dept. of Computer Science and Engin.
Univeristy of Ioannina
Ioannina, Greece
demitsampra@gmail.com

Lisimachos P. Kondi
Dept. of Computer Science and Engin.
Univeristy of Ioannina
Ioannina, Greece
lkon@cse.uoi.gr

***Abstract*— The adoption of a Natural Scene Statistics (NSS) model has been an important research direction in the selection of perceptual features capable of giving satisfactory results in the problem of image quality assessment (IQA). In this work, trying to improve the performance of a blind IQA methodology, we simultaneously consider quality aware features from both the spatial and the transform domains. Moreover, for the first time, a statistical description of the spatial domain is investigated through the Student's t distribution, trying to predict the subjective evaluation of humans and to reduce the total number of features. In essence, a large number of features are used, which are optimized by the consequent characterization with the distribution's parameters. The proposed model is then fed to a tool to learn a simple regression model. In this way the extracted trained model is used to predict the graded image quality score, based on known publicly available datasets. The results are interesting and show high levels of agreement with the subjective human perception while maintaining a low total number of features.**

Keywords—blind image quality assessment, natural scene statistics, image distortion, Student's t, hybrid model

I. Introduction

The objective evaluation of the perceived quality of an image is a growing necessity in modern image processing systems. This has a catalytic effect on a wide range of applications, starting with the capture of visual information and ending with applications like image restoration, image super-resolution, image denoising and compression to name a few [1]-[4]. Although the goal is the human viewer of information, it is particularly difficult to have one person at a time as the ultimate judge of the information you hold on a device. In addition, large scale applications with real-time constraints and the necessity for real-time quality score make the evaluation a cumbersome procedure [2]. Therefore, the problem shifts from its subjective nature and over the years it has become necessary to define algorithmic procedures that will give us an objective result (Objective IQA) [1]-[4].

Depending on the availability or the amount of information of the original reference image, we can distinguish the full-reference (FR) [7]-[9], reduced reference (RR) [10]-[12] and no-reference (NR) methodologies [13]-[15]. In the case of this work, we deal with the category of No Reference (NR)/blind IQA where the original image is not available. There are several cases where this category becomes the only choice since no information about the pristine image is provided. In addition, it is important to emphasize that no knowledge is available about any distortion that has taken place. Although the prior knowledge of the distortion may have worked well enough for a specific type of degradation [16]-[18], in this paper we work within the framework of 'distortion-unaware' methods. Thus, we seek to predict the quality of distorted images without any knowledge of pristine reference images and in accordance with human perception of quality [4], [5]. Thus, this general-purpose NR IQA category does not depend on prior knowledge of any kind of distortion.

NSS-based blind IQA methods assume that statistical regularities of natural scene images are retained in a large set of images and change, even significantly, depending on the distortion they undergo [4]. These regularities are predictably modified by the presence of distortions. Thus, most existing methods construct NSS models by exploiting some perceptual-specific features either from the spatial domain or from the transform domain [13]-[15]. We usually have an effective feature extraction procedure, followed by training a regression module using those features. In case we don't have any human scored images, then there is an approach of estimation development of the deviations from statistical regularities inherent in natural images [6]. Furthermore, it is obvious that the predictive capabilities of image quality evaluation algorithms are based on the appropriate selection of quality aware features through their characterizing parameters like mean, variance or skewness [1]-[4].

Depending on the features used, various methodologies with high performance have been proposed. In their work, Moorthy et al., proposed the BIQI metric [5], trying to describe wavelet coefficients. The subsequent modification, entitled DIIVINE [13], had a richer set of features derived from a NSS wavelet coefficient model. In the work of BLIINDS-II [14] the proposed image quality measures were based in the DCT domain. Then, in the case of BRISQUE [15], an exclusive spatial domain NSS based NR-IQA model is proposed. This has achieved better performance than DIIVINE and BLIINDS-II with low computational complexity. BRISQUE used a parametric model to describe the regularity of natural scene statistics, introducing fitting errors due to variations of image contents. Usually, parametric models of Generalized Gaussian Density (GGD) and the Asymmetric form (AGGD) are used to describe natural scene statistics [13]-[15]. Consequently, support vector machine regression (SVR) can map it to more accurate image quality scores [32].

It is noteworthy that the use of deep learning has shown competitive performance, but more research is needed for specific models to be considered as a basis for comparison. For example, the size of existing datasets is intolerably small for training a deep model in a direct way, thus highlighting issues related to the type of common performance appraisal platform and efficiency issues. In addition, there are still issues related to perceptual features or other issues of normalization within networks [26]-[28]. In this work, we do not adopt this direction but based on NSS we extract quality aware features, with a significantly smaller amount of data [4], [13]-[15]. More specifically, the work presented here is focused on the deployment of two types of features in the spatial domain and the transform domain.

2021 Picture Coding Symposium (PCS) | 978-1-6654-2545-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/PCS50896.2021.9477453

978-1-6654-2545-2/21/$31.00 ©2021 IEEE

In the spatial domain, we propose a flexible enough distribution to represent the histogram of our image that is affected by various types of distortion. Thus, we propose the Student's t distribution and we realize that the shape of our data in the spatial field is inherent in its properties [22]-[23]. This means that it can describe a range of shapes of a distribution, from Gaussian to sharper shapes at the center of the distribution. In addition, the combination of the proposed alternative statistical description together with the fastest parameter estimation algorithm, helps us to maintain high levels of evaluation along with computational cost reduction. While the present study is related to known methodologies in IQA, it capitalizes on the spatial domain's feature space, which is the first time that is considered.

The framework of the proposed approach is summarized in Fig. 1. Feature extraction is based on various existing methods [4], [13]-[15]. The refinement of our methodology is based on the determination of the number of features. This is achieved mainly by choosing to configure the proposed statistical description of our data. Thus, the composition of an N-dimensional feature vector (large enough, but much smaller than other similar methodologies), helps to improve performance and correlate with human subjective evaluation. In what follows, we assume that the feature vector to be used is defined as: $\mathbf{f} = \left[f_1, f_2, ..., f_N\right]$ with N being the total number of features. In the following, we formulate each feature of the NSS based methodology in detail.

II. Proposed Spatial Based Features

A. Locally Normalized Luminance Coefficients

Natural images are known to be highly correlated [20]. The correlation can be affected by distortions in different forms. In addition, strong statistical regularities are present across images with different contexts [21]. However, the presence of distortions causes a change in the values that describe them, so there is the possibility to measure the deviation from the normality and therefore to calculate their degradation. Usually, in the spatial domain, NSS features are considered from the luminance map through mean subtracted contrast normalized (MSCN) coefficients [4], [15], [21]. The statistical properties of the MSCN coefficients have been widely studied in the context of IQA methodologies. These can be adequately described through Generalized Gaussian distribution [15]. In this work, for the first time, we make use of quality-related features by adopting an alternative one-dimensional distribution - other than the known GGD distribution, meaning the Student's t distribution [22] - [24].

Thus, to model the statistical regularities observed in natural images, the MSCN coefficients of an intensity image $I(i, j)$ are computed by [15]:

$$L(i,j) = \frac{I(i,j) - \mu_L(i,j)}{\sigma_L(i,j) + C} \quad (1)$$

where i, j are spatial domain indices. L is a luminance image and C is a constant we use to avoid zero denominator, thus avoiding instability issues. In this context, with, $\mu_L(i, j)$ and $\sigma_L(i, j)$ we denote the estimated local mean and local deviation [15]. These normalized luminance values tend to be described by a unit normal Gaussian distribution [21]. In our proposed framework, this statistical behavior of the coefficients is described through the alternative parametric

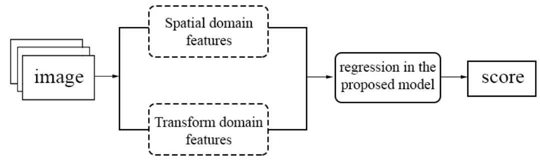

Fig. 1. High – level overview of the proposed methodology

distribution we propose. Moreover, the combination between these two distributions, GGD and our proposal, is also investigated.

B. Commonly used probability density function

In order to quantify the quality of an image in the presence of some kind of distortion, we usually resort to a parametric statistical distribution that describes the data shape (e.g. the histogram) in a compact representation through its parameters. For example, in the case of the known BRISQUE model [15], the univariate Generalized Gaussian distribution (GGD) has been used to fit the data of interest, thus providing us with the appropriate features related to the perceived quality of an image. This distribution is characterized by a large concentration of values around zero and heavy tails. Of particular importance is also the investigation of the statistical relationship in the spatial domain between the adjacent normalized coefficients [15]. In this framework, more features are typically produced by the shape, mean, left variance and right variance from an Asymmetric Generalized Gaussian Distribution (AGGD) of the pairwise products in the horizontal (H), vertical (V) and diagonal (main -D1 and secondary – D2) directions [15]. These form the first set of features that will be used to capture image distortion, as shown in Table I.

Previous work in IQA has shown that extracting features and performing analysis at multiple scales can improve the quality assessment method [14]. Thus, these features refer to two scales (the original and a reduced one – low pass filtered and downsampled by a factor of 2). That means that, at the end, we have in total 36 features – 18 at each scale, in order to perform the goal of quality assessment. The feature vector can be written as: $\mathbf{f} = \left[f_1, f_2, ..., f_{18}, f_{19}, f_{20}, ..., f_{36}\right]$ and in Table I, we can summarize the feature description for the first scale using GGD and AGGD.

TABLE I. Summary Of Features Of GGD and AGGD

Feature ID	***Feature Description (for a unique scale)***
$f_1 - f_2$	Shape and variance of GGD for MSCN coefficients
$f_3 - f_6$, $f_7 - f_{10}$, $f_{11} - f_{14}$ $f_{15} - f_{18}$	Shape, mean, left variance, right variance of AGGD for H, V, D1, D2 directions of MSCN pairwise products respectively

C. Proposed Statistical Distribution

In what follows, we propose the application of Student's t distribution regarding the characterization of MSCN coefficients and the pairwise products of neighboring MSCN coefficients. The rationale behind this choice is threefold.

The first one is related to the investigation of whether the use of different perceptual features, through alternative statistical characterization, can give us a more accurate characterization of the data. Then, in this base, we see if we can contribute to improving the predictive power or maintain high performance levels and finally if we can improve the consequent computational costs. The Student's -t distribution is defined as [24]:

$$St(x;\mu,\sigma^2,\nu)=\frac{\Gamma\left((\nu+1)/2\right)}{\Gamma\left(\nu/2\right)\sqrt{\pi\nu\sigma^2}}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}} \tag{2}$$

where μ is the mean and σ^2 the known scale parameter and $\Gamma(x)$ is the known gamma function. Parameter ν defines the degrees of freedom (dof) where $\nu \in (0, \infty)$ and is responsible for the control of the tail length. This means that for small values of ν we have considerable weight in the tails. For larger values of ν $(\nu \to \infty)$, the distribution increasingly resembles the normal distribution giving less weight in the tails. The variance of the distribution, is $\sigma\nu/(\nu-2)$ for $\nu > 2$ and infinite if $0 < \nu < 2$. The t-distribution can be thought of as a weighted sum of normal distributions with the same mean, but variance that depends inversely on the gamma distribution.

In order to estimate the distribution's parameters, we usually resort to the known EM (Expectation Maximization) algorithm [24], [25]. After various experiments, in this work, in order to improve the efficiency of our proposal, we follow an alternative approach for model parameter's estimation. It is important to emphasize that the use of the proposed parameter estimation, led us to further reduce the time complexity. Indicatively, the execution time of the experiments (1000 iterations) was reduced from a total of 8 hours to a few minutes. The scale parameter σ^2 and degrees of freedom ν, are computed from the histogram of our interest, either MSCN coefficients, or their adjacent neighbors product, using the following formulas:

$$\sigma^2=\frac{\nu-2}{\nu}\sigma_x^2\ ,\ \ k_x=\frac{6}{\nu-4} \tag{3}$$

where σ_x^2 is the variance and k_x is the kurtosis, and they can be estimated from the second and fourth moment of the data [26]. Notice that, the Student's-t distribution is symmetric around its mean and using various values of degrees of freedom we can observe various forms of t-prior. In addition, the t-prior can express the fat tail and excess kurtosis (Fig. 2). Thus, describing MSCN coefficients and the pairwise products, we have in total 10 features –at each scale. The feature vector can be written as: $\mathbf{f}=\left[f_1,f_2,...,f_{10},f_{11},f_{12},...,f_{20}\right]$ and in the Table II, we can summarize the feature description for the first scale. From the modeling results it was observed that the Student's t distribution, with this parameterization could adequately describe the coefficients of our interest.

D. Entropy feature

As an additional feature we derive the image entropy, based on the work of Fang et al. [30]. This feature is based in the spatial domain and is not related with the statistical description of the MSCN coefficients of our interest.

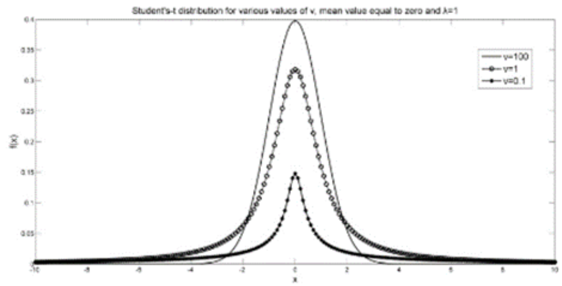

Fig. 2 Modelling of the coefficients of our interest with Student's-t distribution. Note that Student's t pdf provides good fit to both tail and mode of histograms of actual data.

TABLE II. SUMMARY OF FEATURES OF STUDENT'S T DISTRIBUTION

Feature ID	*Feature Description (for a unique scale)*
f_1-f_2	Scale parameter and dof of Student's-t (MSCN coefficients)
f_3-f_4, f_5-f_6, f_7-f_8, f_9-f_{10}	Scale parameter and dof of Student's-t (pairwise product in Horizontal direction, Vertical, Diagonal and SubDiagonal direction)

Thus, by taking blocks of size 5x5, we can divide every block into three energy subbands. As we can see in Fig. 3 (left), there are three subbands denoted with $n=1,2,3$, represented by different levels of shading. Then, by resorting to the work of [14], we can define the energy subband ratio of DCT coefficients of each block as:

$$R_n=\frac{|\,E_n-\frac{1}{n-1}\sum_{j<n}E_j\,|}{E_n+\frac{1}{n-1}\sum_{j<n}E_j} \tag{4}$$

where $n=2,3$ and E_n denotes the energy of every subband (n=1 corresponds to the low energy subband, n=2, corresponds to the medium energy subband and n=3 corresponds to the high energy subband). By taking the mean value of R_2 and R_3, the average value for all blocks in the image and the highest 10th percentile average of the local scores across all the image, provide us with two additional features. By filtering our image twice with a low-pass Gaussian filter and repeating the previous procedure, we can also derive four features in addition. Thus, in total we have six features related with the energy of the image in a local base in transform domain.

E. Features based on coefficient of frequency variation

In this group of features, we exploit three orientations based on block DCT. These orientations, are in essence three local subareas, which are defined sequentially along the subdiagonal area. Thus, for a block of size equal to 5x5, we define fequency variation ζ_j by:

$$\zeta_j=\frac{\sigma_j}{\mu_j}\ ,\ j=1,2,3 \tag{5}$$

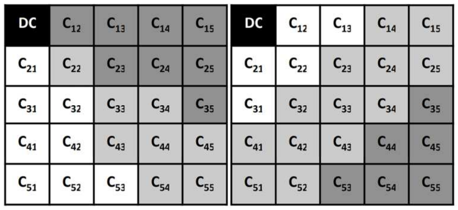

Fig. 3 DCT coefficients along three frequency bands (left) and three orientations (right).

where σ_j and μ_j are the mean and standard deviation of the DCT coefficient magnitudes respectively. The frequency of variation ζ_j in three orientations is computed for all blocks in the image. Then, the variance of ζ_j for all blocks, is computed. By taking the mean and the highest 10th percentile of the local block scores across the image, we define another six features. The rationale behind this choice is related to rich form that is inherent to these kind of differences. The number of entropy-related features of the previous section and the features of this section are listed in the Table III

III. Experimental Results

The performance of our methodology can be evaluated by using subjective image datasets, where each image has been scored by human observers. The proposed method is examined on two known image databases, but for the favor of space, the image data set we use in this work is LIVE-C [31]. In order to obtain the results presented in this section, we repeat 1000 times the division into 80% of the images for training purposes and 20% of the images for testing purposes (without overlapping of each case). The final performance is the average value using the results of all these trials. Thus, we report the Spearman rank order correlation coefficient (SROCC) and Pearson linear correlation coefficient (PLCC) among the predicted scores and the objective scores. For the purpose of managing high-dimensional data along with the assumption that it is a commonly accepted option of various IQA methodologies, we resort to SVR. In Table IV, for comparison, we indicate the performance of known methodologies.

TABLE III. Summary of the Additional Features That We Make Use, except the Features that Are Based on the Statistical Description of the MSCN Coefficients

Feature ID	***Feature Description (for a unique scale)***
Spatial domain	
f_1	Entropy of the image
Transform domain	
$f_2 - f_7$	Energy subband ratio measure (10% and mean)
$f_8 - f_{13}$	Coefficients of variation ζ

A. Performance on known datasets and computational complexity

In Table V, we present various combinations and the number of used features that provide us with competitive results. Thus, we can verify the effectiveness of our proposal based on four known IQA methodologies as these depicted in Table IV. As one observes the performance based on our proposed methodology, it clearly shows competitive performance in combination with a reduced number of features.

TABLE IV. Mean SROCC and PLCC Of Known NR-IQA methods Over 1000 trials On Live-C Database

Method	***SROCC***	***PLCC***
DIIVINE	0.58	0.60
BLIINDS-II	0.45	0.46
BRISQUE	0.59	0.61
NIQE	0.43	0.45

TABLE V. Mean SROCC Values Over 1000 TRials On Live-C Database

MSCN distr.	**Prod. of neighb. MSCN distr.**	**Add/nal features**	**Total # of feat.**	**SROCC**	**PLCC**
Student-t	AGGD	14 DCT 1 entropy	51	0.5576	0.5408
Student-t +GGD	AGGD	14 DCT 1 entropy	55	0.5556	0.5428
Student-t	Student -t	12 DCT 1 entropy Kurtosis + asymmetry for MSCN prod	49	0.5657	0.5653
Student-t	Student -t	Kurtosis + asymmetry for MSCN prod	36	0.5319	0.5312

From Table V, we verify the importance of using an alternative statistical characterization for the known MSCN coefficients. Since the proposed distribution allows extra flexibility in the kurtosis, we can account for the uncertainty in the nature of the tail of the empirical distribution we study. In the results we also added kurtosis and asymmetry as an additional feature. Notice that in case of DIIVINE [13], 88 features based on NSS model of wavelet coefficients were used. In summary, we observe that the modeling of either the distribution of MSCN coefficients or the products of neighboring MSCN coefficients with Student's t distribution, provide high levels of prediction while maintaining a small number of features.

IV. Conclusions

In this work, we provided a hybrid methodology, while we studied the importance of an alternative statistical description of our data. The confirmation of the validity of using the proposed statistical distribution in spatial domain is investigated through its competitive results along with the reduced number of features we make use. In addition, the choice of methodology for calculating the parameters is fast, with the result that the computational complexity of our proposal is further reduced.

REFERENCES

[1] Z. Wan and A. C. Bovik, Modern Image Quality Assessment (Syntheses Lectures on Image, Video and Multimedia Processing), San Mateo, CA, USA: Morgan and Claypool, 2006

[2] A. K. Moorthy and A. C. Bovik, "Visual Importance Pooling for Image Quality Assessment," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193-201, April 2009

[3] D.M. Chandler, Seven Challenges in Image Quality Assessment: Past, Present, and Future Research, ISRN Signal Processing

[4] Z. Wang and A. C. Bovik, "Reduced- and No-Reference Image Quality Assessment," in *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29-40, Nov. 2011,

[5] A. K. Moorthy, A.C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," in IEEE Signal Processing Letters vol. 17, no. 5, pp.513 – 516, June 2010

[6] A. Mittal, R. Soundararajan, A.C. Bovik, Making a 'completely blind' image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212

[7] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612

[8] Sheikh, H.R., Bovik, A.C.: 'Image information and visual quality', IEEE Trans. Image Process., 2006, 15, (2), pp. 430–444

[9] Zhang L, Zhang L, Mou X, Zhang D. FSIM: a feature similarity index for image quality assessment. IEEE Trans Image Process. 2011 Aug;20(8):2378-86

[10] X. Gao, W. Lu, D. Tao, X. Li, Image quality assessment based on multiscale geometric analysis

[11] Rehman, A., Wang, Z.: 'Reduced-reference image quality assessment by structural similarity estimation', IEEE Trans. Image Process., 2012, 21, (8), pp. 3378–3389

[12] Tao, D., Li, X., Lu, W., et al.: 'Reduced-reference iqa in contourlet domain', IEEE Trans. Syst., Man Cybern. B Cybern., 2009, 39, (6), pp. 1623–1627

[13] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," IEEE Trans. Image Process., 20 (12), 3350-3364, 2011

[14] M. Saad, A.C. Bovik, C. Charrier, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," IEEE Trans. Image Process. 21 (8) (2012) 3339–3352

[15] A. Mittal, A. K. Moorthy and A. C. Bovik, "No reference image quality assessment in the spatial domain," IEEE Trans. Image Process., 21 (12) (2012) 4695-470

[16] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in Proc. IEEE Int. Conf. Image Process., 2002, pp. 57–60.

[17] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 4, pp. 529–539, Apr. 2010

[18] H. Liu and I. Heynderickx, "A perceptually relevant no-reference blockiness metric based on local image characteristics," EURASIP J. Adv. Signal Process., vol. 2009, Art. no. 263540, 2009

[19] Ke Gu, G. Zhai, X. Yang, W. Zhang, "Using free energy principle for blind image quality assessment," IEEE Trans. On Multimedia, vol. 17, no.1, Jan. 2015

[20] A. Srivastava, A. B. Lee, E. P. Simonceli and S. C. Zhu, "On advances in statistical modeling of natural images", J. Math. Imag. And Vision, 18, pp. 17-33, 2003

[21] D.L. Ruderman, The statistics of natural images, Network: Comput. Neural Syst. 5 (4) (1994) 517–548

[22] P. Gendron, B. Nandran, "Modeling heavy tailed correlated noise with packet basis functions," J. Stat. Planning Infer., 11, pp. 99-114

[23] A O' Hagan, "Modeling with heavy-tails. Bayesain Statistics", Oxford University Press, Oxford 2003, pp. 345-359

[24] A. Mairgiotis, L. P. Kondi and Y. Yang, "Weak Signal Watermark Detection Through Rao-T Hypothesis and Lightweight Detection," *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 2018

[25] A. Mairgiotis, G. Chantas, N. Galatsanos, K. Blekas and Y. Yang, "New Detectors for Watermarks with Unknown Power Based on Student-t Image Priors," *2007 IEEE 9th Workshop on Multimedia Signal Processing*, Chania, Greece, 2007

[26] W. H. Press, S. A. Teukolsky, W. T.. Flannery, "Numerical recipes in C", 2nd edn, Cambridge University Press, 1992

[27] S. Bosse, D. Maniry, K. -R. Muller, T. Wiegand, W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment", *in IEEE Trans. Image Process*. 27 (1) (2018) 206–219

[28] X.Yang, F. Li, H. Liu, "Deep feature importance awareness based no-reference image quality prediction", Neurocomputing, 401 (2020), 209-223

[29] K.-Y. Lin and G. Wang, "Hallucinated – IQA: No-reference image quality assessment via adversarial learning," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., Apr. 2018, pp. 723-731

[30] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang and G. Zhai, "No-Reference quality assessment of contrast distorted image based on natural scene statistics," *IEEE Signal Process*. Lett., vol. 22, no.7, pp.838-842, Jul. 2015

[31] Ghadiyaram and A.C. Bovik, "LIVE In the Wild Image Quality Challenge Database, "Online: http://live.ece.utexas.edu/research/ChallengeDB/index.html, 2015.

[32] V. N. Vapnik, The Nature of Statistical Learning Theory, 1995