

# MACHINE LEARNING SOLUTIONS FOR OBJECTIVE VISUAL QUALITY ASSESSMENT

*Paolo Gastaldo<sup>a</sup> and Judith A. Redi<sup>b</sup>*

<sup>a</sup>Dept. Biophysical and Electronic Engineering, University of Genoa, Via Opera Pia 11a, 16145, Genova (Italy)

<sup>b</sup>Dept. Mediamatics, Delft University of Technology, Mekelweg 4, 2628CD Delft (The Netherlands)

## ABSTRACT

Objective metrics for visual quality assessment usually improve their reliability by explicitly modeling the highly non-linear behavior of human perception; as a result, they often are complex, and computationally expensive. Conversely, Machine Learning (ML) paradigms allow to tackle the quality assessment task from a different perspective, as the eventual goal is to mimic quality perception instead of designing an explicit model the Human Visual System (HVS). Several studies already proved the ability of ML-based approach to address visual quality assessment. Indeed, a prerequisite for successfully using ML in modeling perceptual mechanisms is a profound understanding of the advantages and limitations that characterize learning machines. This paper illustrates and exemplifies the good practices to be followed.

## 1. INTRODUCTION

To fulfill users expectations in terms of quality of the visual experience, visual quality needs to be controlled in every stage of the digital media delivery chain, and restored if necessary. To do so in accordance with human preferences, control mechanisms should be based on automated quality assessment systems that replicate as accurately as possible the human visual system (HVS). At the same time, the eventual computational complexity of these systems should be kept low, to allow real-time and hardware implementations.

A variety of methods for objective quality prediction of images and video has been proposed in the literature [1]. These approaches usually decouple the quality assessment task into two steps: first defining an objective metric to estimate the intensity of the distortion(s) affecting the media, and then mapping this measure into quality scores by means of regression [2]. Most objective metrics attempt an explicit modeling of the highly non-linear behavior of the HVS; thus, usually they are complex, and often computationally expensive.

Machine Learning (ML) paradigms allow to tackle

the quality assessment task from a different perspective. They mimic the HVS reactions to quality losses, rather than explicitly modeling it. Quality assessment based on ML paradigms conforms to the 2-step approach of traditional methods, though modifying the balance between the computational effort involved in each step. In the first step, a meaningful feature-based representation of the distortion affecting the media is defined. In the second step, the learning machine handles the actual mapping of the feature vector into quality scores, attempting at reproducing perceptual mechanisms (Fig. 1.a). Such an approach relies on the ability of ML tools to learn from examples the complex, non-linear mapping function between feature vectors and quality scores. Consequently, a) relatively simple, computationally inexpensive metrics can be designed, and b) most of the computational power is spent in the off-line training phase of the learning machine. A trained ML-based system can therefore support real-time quality assessment on an electronic device with a minimal overhead to the metric computational cost.

As a matter of fact, the ML-based approach to perceived quality assessment has been proved so promising that it has been adopted in the ITU standard for audio quality assessment, the PEAQ [3]. Studies also proved the effectiveness of methodologies that exploit ML tools to address video [4-7] and image [8-15] quality assessment. In this paper we provide an overview of the benefits that the use of ML can bring to the Visual Quality Assessment problem in terms of both accuracy and computational complexity. To do so, we will comment and exemplify step by step the setup of a ML-based objective quality metric.

## 2. MACHINE LEARNING FOR VISUAL QUALITY ASSESSMENT

The typical setup of a ML-based objective metric includes two steps:

- 1) The definition of a feature-based description  $\mathbf{f} \in \mathcal{F}$ , where  $\mathcal{F}$  is a feature space, of the visual input (image or video)  $S$ :

$$\mathbf{f} = \phi(S, S') \quad (1)$$

where  $S'$  is the reference signal (if needed). The feature space  $\mathcal{F}$  represents a description of the perceived distortion in the image. In this sense, the feature extraction corresponds to the computation of an objective metric.

- 2) The learning of the non-linear mapping function,  $\gamma$ , between the feature space  $\mathcal{F}$  and the subjective quality score,  $q$ :

$$\gamma: \mathcal{F} \rightarrow q \in \mathcal{C} \subset \mathfrak{R} \quad (2)$$

In this step the complex relationships that link physical image characteristics measured in the feature space to subjective quality preferences are learned from examples by using a training algorithm. As a result, the challenging issue of designing an explicit model of perceptual mechanisms is partially by-passed.

Figure 1, path (a) schematizes the two-step approach just described. This basic approach can be augmented by an extra step (path b), that seeks to identify the distortion affecting the image (e.g., JPEG compression, noise, blur) prior to assessing its quality. This allows to use for the quality prediction either features or mapping functions specifically designed to quantify the quality loss produced by the identified distortion. ML has been recently shown to be excellent at supporting this task as well [12, 14].

A prerequisite for successfully using ML in modeling complex input-output relationships such as perceptual mechanisms is a profound understanding of the advantages and limitations that characterize learning machines and of the application-specific context. In general, three tasks need to be completed to develop a ML-based objective quality assessment system:

1. The definition of the feature space  $\mathcal{F}$  that describes the input signals.
2. The selection of the ML tool to be used to implement the mapping function  $\gamma$ .
3. The training of the system and the robust test of its generalization performance (i.e., its prediction accuracy on inputs not included in the training phase).

Good practices exist to effectively implement each of these steps. The following sections illustrate these good practices by discussing related literature and ad-hoc examples.

### 3. DEFINING THE FEATURE SPACE

Most of the effectiveness of a ML-based VQA system depends on a good design of feature space  $\mathcal{F}$  in which the media is mapped before entering the ML predictor. A mathematical function  $\gamma$  mapping the feature space  $\mathcal{F}$  and the perceived visual quality space should exist: *no ML paradigm can repair a defective feature space design by restoring missing information*. As a major consequence,

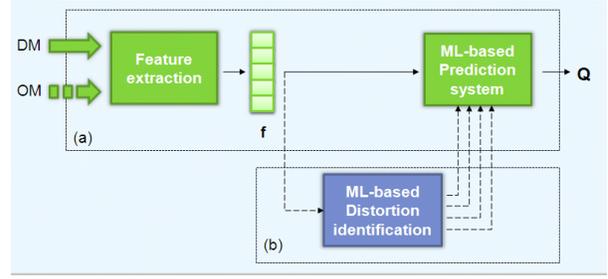


Figure 1 – CI-based image/video quality assessment system. The input (distorted) media DM and, if available, the original media (OM) are first represented in a low-dimensional feature space, and then mapped into the quality space by means of a CI tool (a). Prior to the mapping, further information on the distortion history of the media can be obtained through a CI-based Distortion Identification module (b).

the features need to encode relevant information on how the distortion in the image is perceived by the HVS. Indeed, most studies on ML-based methods for visual quality assessment rely on feature sets derived from perceptual models previously proposed in the literature [6, 9, 11, 14]. In [6] for example, a FR video quality assessment system exploits a 20-dimensional feature space, involving both HVS-based measures and distortion-oriented measures. In their recently proposed NR system, Moorthy and Bovik [15], structure an 88-dimensional feature space. Those features are obtained by exploiting the peculiarities of wavelet transforms, which can perform mirror models of spatial decompositions occurring in the V1 area of the primary visual cortex.

A second, critical issue is the eventual dimensionality of the feature space. If the function  $\gamma$  is defined over a high-dimensional domain, the convergence of a ML prediction system to the true value is known to be very slow, due to the so-called “curse of dimensionality”. This issue becomes more critical when a limited amount of data is available for the training, since the number of training samples required to sustain a given spatial density increases exponentially with the dimensionality of the input space  $\mathcal{F}$  [16]. In this case, even the overall performance of the system might be affected. As a consequence, the feature space  $\mathcal{F}$  should encode *all and only* the information that is relevant to quality assessment. To do so, models of the HVS have to be implemented to extract relevant features, and then simplified, possibly through an appropriate feature selection procedure. Feature selection can remove noise from data, i.e. it can eliminate non-informative features that encode either redundant data (correlation with other features) or inappropriate information (no correlation with the output space). This in turn avoids the computational complexity of the eventual prediction system to increase without control. On the other hand, one may rely on ML paradigms that are less prone to curse of dimensionality, e.g. Support Vector Machines (SVM) [17].

Only a few studies on ML-based methods give details on their feature selection strategy. A number of studies exploit second-order histograms to characterize distortion perception (e.g., [9, 12]). Several features can be deduced from those histograms, each characterizing specific statistical (and perceptual) characteristics of the image. To eliminate redundant information, in [9] PCA is adopted to select the most significant features. In [12] the Kolmogorov-Smirnov test is exploited instead. In [10] the feature selection is inherently embedded in the ML-based approach for FR image quality assessment. The system uses singular value decomposition (SVD) to define feature vectors that quantify major structural information in the images. Thus, the eventual dimensionality of the feature space depends on the number of singular vectors used to represent the images.

#### 4. SELECTION OF THE ML PARADIGM

The task of the learning machine in a ML-based quality assessment system is to find the function  $\gamma$  that best maps a feature vector  $\mathbf{f} \in \mathcal{F}$ , into its associate ‘target’ subjective quality score,  $q$ . As  $q$  typically assumes continuous scalar values, e.g.,  $q \in \mathcal{C} = [-1, 1]$ ,  $\gamma$  is a regression function. In the practical design of any ML-based system, the regression strategy implements the decision function,  $\hat{q} = \gamma(\mathbf{f})$  as a weighted series, whose basic terms,  $\phi(\mathbf{f})$ , typically embed nonlinear functions:

$$\hat{q} = \gamma(\mathbf{f}) = \sum_i \beta_i \phi_i(\mathbf{f}) + \beta_0 \quad (3)$$

Equation (3) is a general expression for a variety of paradigms, ranging from feed-forward neural networks (e.g., Multi-Layer Perceptron [16]) to kernel machines (e.g., Support Vector Machines [17]). In the training phase, the task of the ML paradigm is to learn through examples the mapping function  $\gamma$ . The training set  $TG = \{(\mathbf{f}_i, q_i); i=1, \dots, n_i\}$  gives an example-based formulation of the ideal input-output mapping. The training procedure consists in adjusting the degrees of freedom of  $\gamma$  (e.g., the coefficients  $\beta_i, \beta_0$ ) in such a way that (3) is able to reproduce that mapping, and predict quality  $\hat{q}$  on unseen data with a known accuracy.

Table 1 proposes a taxonomy of the ML paradigms adopted for image/video quality assessment in literature. Feed-forward neural networks such as The conventional MultiLayer Perceptron (MLP), Circular BackPropagation (CBP) network, and Extreme Learning Machine (ELM) proved to be very powerful and flexible tools to address highly non-linear regression problems [16]. Thus, they are often conveniently used in visual quality prediction.

Besides accuracy, the selection of the ML tool is also driven by the domain structure of the underlying problem. In [13], a general regression neural network (GRNN)

TABLE 1. TAXONOMY OF CI PARADIGMS FOR MULTIMEDIA QUALITY ASSESSMENT

	Image	Video
MultiLayer Perceptron		[4]
Circular BackPropagation	[8] [11] [12]	[5]
Extreme Learning Machine	[9] [15]	
Support Vector Machine	[10] [14] [12]	
Radial Basis Function		[6]
Time Delay Neural Network		[7]
General Regr. Neural Network	[13]	

supports NR image quality prediction. A GRNN is a probabilistic neural network, which in principle can guarantee fast learning but also lead to trained machines that are computationally less efficient than feed-forward neural networks. In [7], a Time Delay Neural Network (TDNN) is adopted for video quality assessment as it has the ability to represent a relationship between events in time. Also SVMs can serve as efficient regression machines. In [10] and [14], a SVM is used to tackle a regression problem because of the high-dimensional feature space, which does not favor the use of a feed-forward neural network.

In case the problem to be tackled is distortion identification rather than quality prediction (path (b) in figure 1) the prediction task becomes a classification one. In this setting, the output can assume as many values as the number of distortions considered. SVMs possibly represent the most powerful tool to implement classification systems [17]. Recently, two studies on image quality assessment [12, 14] exploited SVMs to implement the distortion identification module of Fig. 1.

#### 5. EFFECTIVE LEARNING

A critical issue with empirical training is the risk of *overfitting*, i.e., the excessive specialization of the mapping function  $\gamma$  on the training set. Overfitting typically results in a low generalization ability, i.e., poor performance when processing unseen data. Two factors concur to effective learning.

The first factor is a good composition of the training set. The eventual model will learn how to generalize quality prediction for the data distribution represented in the training space; i.e., its generalization ability will not cover samples diverging from this distribution. Thus, the training patterns (images or video) should give a sufficiently large and representative sample of the data population that one wants to address. Existing image and video quality databases (such as [18]) are excellent starting points, though presenting a rather modest number

of patterns. In this sense, larger subjective datasets should be produced in the future.

Second, the generalization ability of a ML algorithm is affected by its configuration settings, e.g. the number of neurons  $N_h$  in a feed-forward neural network, or the kernel function and specific parameterization in the case of SVMs. In literature, an instance of a ML algorithm with a specific setting configuration is referred to as a model. Overly complex models exhibit higher risks of overfitting [19]. Thus, in principle, the final model should be chosen so that this risk is minimized. No well-established theoretical guidelines to address model selection exist that also exhibit practical applicability. Conversely, data-driven criteria such as cross validation proved to be effective in practice [16]. Such technique is also one of the most practical tools to estimate the eventual generalization ability, compared to theoretical means for generalization error bounding [17].

### 5.1. Cross-validation for robust learning

Let  $DS = \{(x_i, y_i); i=1, \dots, n_d\}$  be the a subjective quality dataset.  $DS$  is randomly split into three, non-overlapping subsets: a training set  $TG$ , a validation set  $VS$  and a test set  $TS$ . The training set is, obviously, the collection of data to be used for the learning procedure. The role of the validation set is to support model selection. Given a set of possible models for the adopted ML paradigm, the set  $TG$  is used to complete the training for each of the settings of the model, thus leading to various trained systems. Then, the set  $VS$  is used to evaluate the generalization performance of each model. Eventually, the model with the best performance is selected. Finally, the model generalization ability is measured on  $TS$ , which includes only patterns that have not been involved in the training or validation phase.

As discussed above, in the domain of visual quality assessment usually the size  $n_d$  of the dataset  $DS$  (and, as a consequence, that of  $TG$ ) is rather small (i.e., hundreds of patterns). Hence, published studies often rely on a “multi-run version” of the conventional cross-validation strategy. In this case, multiple runs of the model selection procedure are completed by varying the distribution of data over the training, validation and test set. Such set up eventually leads to a more robust estimate in terms of generalization performance. The k-fold strategy represents a common way to implement such set up [16].

When dealing with visual quality assessment, it is good practice to split the dataset such that  $TG$ ,  $VS$ , and  $TS$  do not share any image content; this allows to evaluate the generalization of the trained system ability also with respect to completely new image content. The great majority of the studies published in the literature apply the multi-run strategy to attain a robust estimation of the generalization error. In a few cases, however, details about

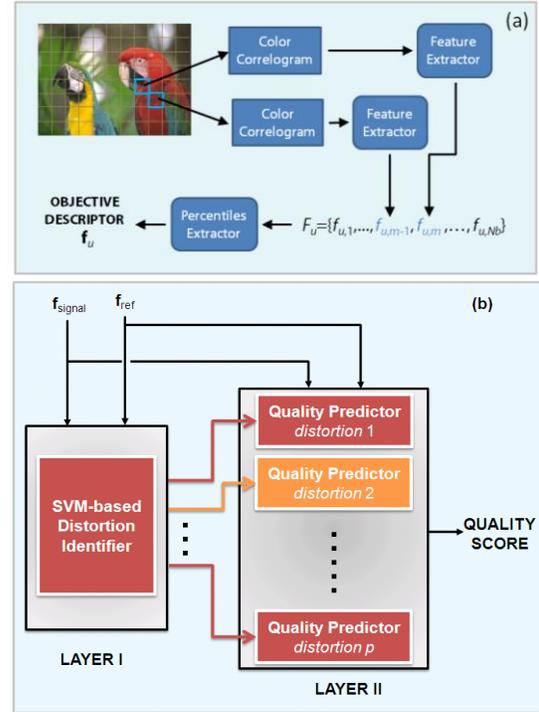


Figure 2 – Feature extraction process (a) and CI-based quality assessment system architecture (b) for the case study presented in [12].

model selection are missing [7, 10, 14].

## 6. A PRACTICAL EXAMPLE

Section 3, 4 and 5 listed a number of good practices to set-up a ML-based quality assessment system. This section exemplifies and motivates those practices with a case study, The Reduced Reference (RR) image quality assessment system proposed by Redi et al. [12] is used here as a reference.

### 6.1. The reference system

The system presented in [12] exploits color information to address reduced reference image quality assessment. A two-layer scheme (Fig. 2(b)) supports the quality assessment framework, according to the approach sketched in Fig. 1. Layer I is designed to identify which out of  $p$  distortions affects the image through a  $p$  one-versus-rest SVM classifiers. The system exploits then the output of layer I to select the distortion-specialized quality predictor to be used at Layer II. This layer hosts  $p$  different non-linear prediction systems, based on CBP neural networks.

The system was tested according to a 5 fold cross-validation. The overall performance of the system on

TABLE 2. COMPARISON OF PREDICTION ACCURACY BETWEEN IMAGE QUALITY METHODS ON THE LIVE DATABASE

		JP2K1	JP2K2	Noise	Blur	JPEG1	JPEG2
CI-based	FR Narwaria and Lin	0.95		0.98	0.95		0.94
	RR Redi et al.	0.77	0.88	0.98	0.85	0.93	0.99
	NR Moorthy and Bovik	0.92		0.98	0.92		0.92
Conventional	FR SSIM	0.94		0.98	0.90		0.94
	RR Li and Wang	0.94	0.96	0.96	0.95	0.82	0.95
	NR Saad et al.	0.80		0.91	0.87		0.59

LIVE database is reported in table II, along with the performance scored by two state-of-the-art ML-based methodologies: the full-reference (FR) approach proposed by Narwaria and Lin [10] and the no-reference (NR) approach recently proposed by Moorthy and Bovik [14]. For the sake of comparison, the table also reports the performance of the following conventional approaches: the FR metric SSIM [20], the RR metric proposed by Li and Wang in [21], and the NR metric proposed in [22]. Numerical results show that: 1) the performance of the ML-based methods is comparable to that of the conventional metrics, and 2) the system presented in [12] compares favorably with the alternative metrics.

## 6.2. Feature Selection

Color correlograms capture the color distribution and textural properties of the image, which have been shown to carry relevant information for image quality assessment. The reference system [12] exploits correlograms to capture such information on a local basis, still eventually providing the assessment system with a holistic representation of quality loss across the image. To do so, first color correlograms are computed for a set of  $N_b$  non-overlapping blocks of the image. From each correlogram 5 statistical features that summarize the changes in the structure of the block are extracted. This results in five collections of as many values as the number of blocks. To capture this structural information at a holistic level, the 20<sup>th</sup>, 40<sup>th</sup>, 60<sup>th</sup>, and 80<sup>th</sup> percentiles of each collection, together with their minimum and maximum values, are extracted and aggregated into a single image descriptor (see Fig. 2(a)). This results into as many global descriptors as the number of features (five). A Kolmogorov-Smirnoff test is then applied to select among the candidate features those two that better characterize the perceptual phenomenon. This entire procedure is repeated for both hue and luminance information. As a result, two hue and two luminance

features are extracted. The prediction systems in layer 2 which eventually receive thus as input two 24-dimensional vectors, one characterizing the input image and the other representing the reference image. Without feature selection, those modules would have received 60-dimensional input vectors.

To prove that feature selection can effectively eliminate non-informative features that may encode redundant data, we evaluated the performance of system [12] by implementing the “no features selection” set up. The experimental setup was kept identical as that used in [12], adopting a five-fold cross-validation strategy and using the LIVE dataset [18] as a benchmark. Table 3 reports the results obtained for an ideal system, i.e. when one assumes that Layer I does not bring about any error in identifying the distortion affecting the images. The same information is reported also for the reference system [12], which adopts feature selection. The comparison is based on the Pearson’s Correlation Coefficient between system predictions and actual subjective quality judgments. The results confirm that the “no features selection” set up cannot improve over the performance originally obtained in [12], rather, the opposite.

## 6.3. Selecting the ML paradigm

In the RR system proposed in [12], the quality predictors of layer II are implemented with Circular BackPropagation (CBP) networks. CBP extends the conventional MultiLayer Perceptron (MLP) with one additional input, the ‘circular’ input, being the sum of the squared values of all the network inputs. Such a structural enhancement improves the ability of the resulting ML model to tackle complex, non-linear problems, such as modelling perceptual mechanisms, still exploiting the powerful training algorithms designed for MLP [16].

Recently, another feed forward neural network, the Extreme Learning Machine (ELM) [23], has been successfully applied in challenging domains such as bioinformatics, computer vision, data mining and robotics. The ELM model has been introduced to overcome some issues in MLP network training, that is, possibly slow convergence rates and the presence of local minima that call for multistart and re-training strategies. In fact, also the ELM model can be enhanced by applying the circular input; the resulting model is called C-ELM [15]. Below we evaluate C-ELM algorithm when embedded in the RR quality assessment framework [12].

Table 3 compares the generalization performance attained by C-ELM on LIVE with that of CBP, still assuming an ideal layer I. C-ELM clearly improves over CBP, as its predictions are more correlated to human quality judgments for each of the six datasets involved in the experimental session. However, one should also take into account that in the case of feed forward neural

TABLE 3. COMPARISON OF PREDICTION ACCURACY BETWEEN THREE SET UP: “CBP – NO FEATURE SELECTION,” “CBP WITH FEATURE SELECTION,” AND “C-ELM WITH FEATURE SELECTION”

	JP2K1	JP2K2	Noise	Blur	JPEG1	JPEG2
CBP - No Feature Selection	0.820	0.809	0.910	0.864	0.894	0.902
CBP with Feature Selection [#15]	0.880	0.874	0.980	0.924	0.938	0.979
C-ELM with Feature Selection	0.928	0.911	0.975	0.953	0.961	0.990

networks the computational complexity of a decision function scales with the number of neurons in the hidden layer  $N_h$ ; that parameter is set by model selection. In this regard, Table 4 reports the configurations adopted for the different C-ELM predictors involved in the proposed framework; for comparison purposes, the Table also gives the same information for the CBP predictors. Those results show that by adopting C-ELM one would eventually obtain a more complex framework than that obtained with CBP, as a larger number of neurons is required in the former case to implement the quality predictors. It should be noticed that in both cases, the risk of overfitting is minimized thanks to cross-validation and model selection. Thus, the systems are equally reliable. The choice of accuracy over low complexity should be dictated in this case only by application-specific requirements. Such analysis confirms that details about model selection should be always included in works presenting ML-based approaches to visual quality assessment, as these parameters clearly characterize the overall system.

## 5. REFERENCES

- [1] W. Lin, C.-C. Jay Kuo, “Perceptual Visual Quality Metrics: A Survey,” *J. of Visual Comm. and Image Repr.*, vol. 22, no. 4, pp. 297-312, 2011.
- [2] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment”, <http://www.vqeg.org/>, 2000.
- [3] T. Thiede, W. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, B. Feiten, “PEAQ — the ITU standard for objective measurement of perceived audio quality,” *J. Audio Eng. Soc.*, vol. 48, no. 1/2, pp. 3–29, 2000.
- [4] F.-H. Lin, R. M. Mersereau, “Rate-quality tradeoff MPEG video encoder,” *Sig. Pro.: Image Comm.*, vol. 14, 1999, pp. 297-309
- [5] P. Gastaldo, S. Rovetta, R. Zunino, “Objective quality assessment of MPEG-2 video streams by using CBP neural networks,” *IEEE Trans. on Neural Networks*, vol.13, no.4, pp. 939-947, 2002
- [6] S. Yao, W. Lin, Z. Lu, E. Ong, and X. Yang, “Video Quality Assessment Using Neural Network Based on Multi-Feature Extraction,” in T. Ebrahimi and T. Sikora (eds.), “Visual Communications and Image Processing,” Proc. of the SPIE, Vol. 5150, , pp. 604-612, 2003
- [7] P. Le Callet, C. Viard-Gaudin, and D. Barba, “A Convolutional Neural Network Approach for Objective Video Quality Assessment,” *IEEE TNN*, vol. 17, no. 5, pp. 1316-1327, 2006

TABLE 4. NUMBER OF NEURONS IN THE HIDDEN LAYER FOR THE PREDICTORS OF LAYER II: COMPARISON BETWEEN C-ELM AND CBP

	C-ELM		CBP	
	Pred 1	Pred 2	Pred 1	Pred 2
JPEG1	150	10	7	7
JPEG 2	90	190	7	7
JP2K1	90	70	3	3
JP2K 2	70	50	3	3
Blur	150	80	5	5
Noise	120	170	3	3

- [8] P. Gastaldo, R. Zunino, “Neural networks for the no-reference assessment of perceived quality,” *SPIE Journal on Electronic Imaging*, vol.14, No.3, pp. 033004/1-11, Jul-Sep 2005
- [9] S. Suresh, R. V. Babu, and H.J. Kim, “No-reference image quality assessment using modified extreme learning machine classifier,” *Applied Soft Computing*, vol. 9, pp. 541–552, 2009
- [10] M. Narwaria, and W. Lin, “Objective Image Quality Assessment Based on Support Vector Regression,” *IEEE Trans. Neural Networks*, vol. 21, no. 3, 2010
- [11] Liu H, Redi J, Alers, H., Zunino R, Heynderickx I, “An Efficient Neural-Network based No-Reference Approach to an Overall Quality Metric for JPEG and JPEG2000 Compressed Images”, *SPIE J. of Elect. Imaging*, vol. 20, no. 4, pp. 1-15, 2011
- [12] J. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino, “Color Distribution Information for the Reduced-Reference Assessment of Perceived Image Quality,” *IEEE Trans. CSVT*, vol. 20, no. 12, pp. 1757-1769, 2010
- [13] Chaofeng Li, A.C.,Bovik, and Xiaojun Wu, “Blind Image Quality Assessment Using a General Regression Neural Network,” *IEEE TNN*, vol.22, no. 5, pp. 793-799, 2011
- [14] A. K. Moorthy and A. C. Bovik, “Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality,”*IEEE Trans. Image Proc.*, 2011, accepted, to be published.
- [15] S. Decherchi, P. Gastaldo, J. Redi, R. Zunino, E. Cambria, “Circular-ELM for the Reduced-Reference Assessment of Perceived Image Quality,” *Neurocomputing*, 2012, in press
- [16] C. M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2006
- [17] V. Vapnik, *Statistical Learning Theory*, New York: John Wiley, 1998.
- [18] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik: LIVE Image Quality Assessment Database at <http://live.ece.utexas.edu/research/quality>
- [19] E. Moody , “The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems”, in Moody, J.E., Hanson, S.J., and Lippmann, R.P., *Advances in Neural Information Processing Systems*, vol. 4, pp. 847-854, 1992,
- [20] Z. Wang and A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. on Image Processing*, vol. 13, no. 4, 2004
- [21] Q. Li and Z. Wang, “General-purpose reduced-reference image quality assessment based on perceptually and statistically motivated image representation,” in *Proc. ICIP 2008*, San Diego, CA, 2008, pp. 1192-1195.
- [22] M. A. Saad, A. C. Bovik, and C. Charrier, “A perceptual DCT Statistics Based Blind Image Quality Metric,” *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, 2010.
- [23] G.-B. Huang, D. Hui Wang, Y. Lan, “Extreme learning machines: a survey,” *Int. J. Mach. Learn. & Cyber.*, vol. 2, no. 2, pp. 107–122, 2011