Subjective Image Quality Prediction based on Neural Network

Sertan Kaya[a], Mariofanna Milanova[a], John Talburt[a], Brian Tsou[b], Marina Altynova[c]
[a]University of Arkansas at Little Rock, 2801 S. University Av, Little Rock, AR, USA 72204;
[b]U.S. Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA 45431;
[c]Qbase, 2619 Commons Boulevard, Dayton, OH, USA 45431

Abstract: This paper investigates the applicability of the Neural Network approach for image quality assessment. The aim is to predict the subjective quality score, namely the difference mean opinion score (DMOS) obtained from human observers, by incorporating a neural network algorithmic approach utilizing extracted statistical features from test and original images. To ease this approach, a MATLAB user interface is developed and presented here. To validate the proposed approach, an image database is selected consisting of various distortion types as test bed in which a DMOS value is provided for each distorted image. Experimental results show that the obtained output of Neural Network correlates well with DMOS values and Neural Network can mimic human observers.

Keywords: subjective image quality assessment, neural network, difference mean opinion score

1. INTRODUCTION

The growth of advance technology allows images to be captured, stored and transmitted from one device to another with ease. During these processes, imaging systems may add some distortion, degradation or artifacts to images. Recent advances in multimedia technology dramatically reveal an increasing demand for quality controlled images. The problem of Image Quality Assessment has been given great attention in literature. Traditionally, Image Quality Assessment methods can be classified into two major categories, namely objective and subjective quality assessments [1]. On the one hand, the objective quality assessment methods measure the quality between original and distorted images based on mathematical analysis. The objective quality methods are also divided into 3 subgroups based on availability of original signal as follows; Full Reference Methods (FR), Reduced Reference Methods (RR), and No-Reference Methods (NR). Some well-known metrics are such as MSE (Mean Squared Error), PSNR (Peak Signal to Noise Ratio) [2], SSIM (Structural Similarity) [2], and VIF (Visual Information Fidelity) [3]. On the other hand, the subjective quality assessment method [1] essentially measures the perceived quality on images based on human observers' opinion. In literature, most commonly used metric are MOS (Mean Opinion Score) and DMOS (Difference Mean Opinion Score) [4].

Gastaldo et all. [5] use Circular Back-Propagation (CBP) neural network, including one additional input to the neural network architecture, extracting first-order histogram, features drives from co-occurrence matrix and discrete cosine transform features from an enhanced image to obtain objective image quality assessment. Bouzerdoum et all. [6] uses Multi-Layer Perceptron (MLP) using modified Wang and Bovik features from JPEG/JPEG2000 images in order to predict Mean Opinion Score (MOS) for images. Paolo Gastaldo and Rodolfo Zunino [7] improves his previous work without using reference image and also incorporates Principal Component Analysis (PCM) to extract best characterize features for JPEG/JPEG2000 images. Here, our work is applicable for various distortion techniques while predicting Difference Mean Opinion Score with ease of using an interface.

In subjective quality assessment, even though MOS or DMOS establish the fundamental aspect for signal processing analyzing purposes, they also bring several drawbacks including: 1-) subjective quality assessment test is costly and time consuming; 2-) it is not automated or cannot be automated; 3-) it is not a real time evaluation. In objective quality assessment, several limitations of this method can be addressed when: 1-) the correlation of objective quality assessment with human visual perception has been questioned; 2-) all quality effecting parameters can not be considered at the same time; 3-) high computing power may be required. In this study, the ultimate goal is to design and develop a hybrid system based on a neural network algorithmic approach while taking into account subjective parameters. Besides, the unique framework is devised as automated and real-time. The MATLAB interface exhibits user-friendliness which is one of the major intended characteristic of the framework.

The organization of the paper is addressed as follows: Section 2 provides a brief overview of Image Quality Assessment. The neural network is summarized and the approach is explained with details in Section 3. The experimental results of our approach are presented in Section 4. In Section 5, the conclusions of this paper are summarized.

2. IMAGE QUALITY ASSESMENT

The basic approach for developing image quality assessment models includes: 1) The selection of a set of features which can be objectively measured and 2) the establishment of a model which can simulate the HVS for the processing and analyzing of selected features. By weighing individual features, perceptual models can predict not only specific attributes but also a global image quality. The main challenges are: first the HVS is extremely complex and not fully understood and second, complex models are time consuming for real-time applications. Different objective metrics simulate different features of Human Visual System (HVS).

Natural images signals are highly structured. It is assumed that the measure of structural information changes provides a good estimation of the perceptual image distortion [8].The fundamental principle of the structural approach is that the human visual system is highly adapted to extract structural information from the visual scene and therefore measurement of structural similarity (or distortion) could provide good approximation of subjective perceptual image quality.

Consider two images $x = \{x_i \mid i = 1,2,...,N\}$ and $y = \{y_i \mid i = 1,2,...,N\}$ where N is the number of pixels and x_i and y_i are the ith pixels of the images of x and y, respectively. SSIM- $\mathrm{SSIM}(x, y)$ combines three comparison components, namely luminance- $l(x, y)$, contrast- $c(x, y)$ and structure- $s(x, y)$ [8]:

$$\mathrm{SSIM}(x, y) = f(l(x, y), c(x, y), s(x, y)) \tag{1}$$

Luminance, contrast and structure comparisons are defined as follows:

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad C_1 = (K_1 L)^2 \\ c(x, y) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad C_2 = (K_2 L)^2 \\ s(x, y) &= \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad C_3 = \frac{C_2}{2} \end{aligned} \tag{2}$$

where μ_x, μ_y, σ_x, σ_y and σ_{xy} are means of x and y, variances of x and y and correlation coefficient between x and y. K_1 and K_2 are scalar constants that $K_1, K_2 << 1$ and L is the dynamic range of the pixel values. Finally, SSIM index yields to:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{3}$$

In Bovik [8] C1=6.50, C2= 58.52

The main drawback of SSIM algorithm in spatial domain is that it is highly sensitive to translation, rotation and scaling of images. Complex-wavelet SSIM algorithm was developed to work in transform domain and can capture non-structured image distortions that is topically caused by movements of image acquisition devices. CW-SSIM methods works only when the amount of translation, scaling and rotation is small compared to the wavelet filter size. The problem can be solved using multi scale SSIM.

3. NEURAL NETWORK

The Neural Network model we adopt in this study is multilayer neural network. The fundamental of multilayer network [9] is based on the backpropagation algorithm. The backpropagation is also known as gradient descent in error. The backpropagation is essentially generalization of the LMS (Least-Mean-Squares) algorithm. The idea of the backpropagation is to increase training speed, improve performance and obtain desired output values by adjusting weights and scaling inputs. The single perceptron can only be used for the classes that are linearly separable whereas a typical multilayer network [10], which is learned by the backpropagation algorithm, has aptitude to classify the classes which are not necessarily linearly separable.

3.1. Feedforward Operation

A simple three-layer neural network [9], illustrated in Figure 1, has input, hidden, and output layers. There is also a single bias unit that has connection with all units except input units. Since these connected components treat like biological neurons, they are named as neurons. All input vectors are presented to the input layer, and the outputs of input neurons corresponds to component vectors. Net activation is computed in hidden neurons which take the weighted sum of inputs into account. Briefly, net activation is the sum of the inner products of weight and input vectors added by initial weight. The net activation is shown below.

$$net_j=\sum_{i=1}^{d}x_i w_{ji}+w_{j0}=\sum_{i=0}^{d}x_i w_{ji}\equiv w^t x, \tag{4}$$

where *i* denotes *ith* input, and *j* denotes a node (neuron) in the hidden layers. The activation function $f(.)$ produces output of the hidden neuron:

$$y_j = f(net) \tag{5}$$

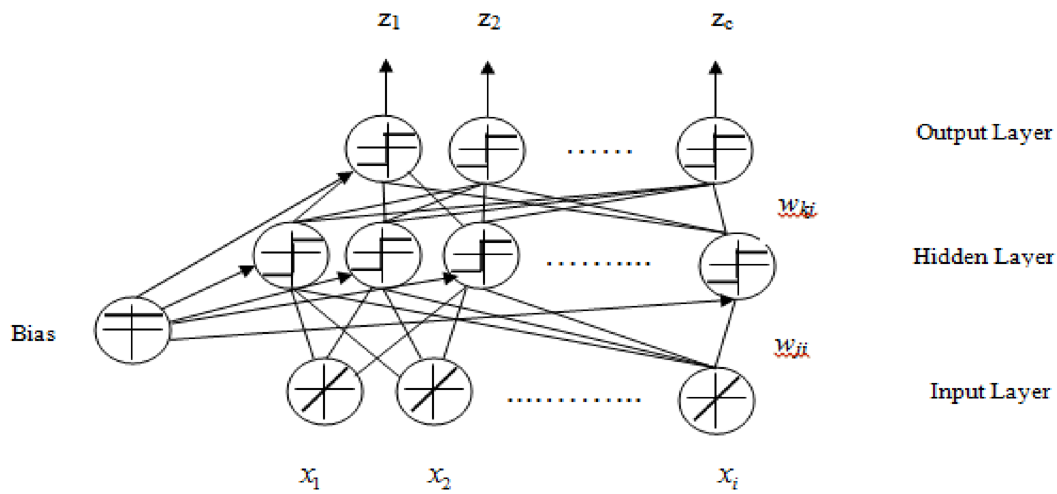

Figure 1: A simple three-layer neural network scheme

3.2. The Backpropagation Algorithm

The Backpropagation algorithm [9] is one of the most common methods in multilayer neural networks for supervised learning. The major issue here is to obtain the desired outputs by adjusting weight vectors that are based on inputs provided by the training sets. The idea behind the backpropagation algorithm is to obtain the sum of the smallest squared difference between actual outputs and desired outputs. This learning rule is valid for not only two-layer architecture, but also for three-layer architecture. The underlying idea behind network learning is to present a training pattern to an untrained network's input layer and to let the signal pass through hidden layer to output layer. Output layer then generates output. The difference between generated output and desired output (target value) is called an error. The aim of the learning rule, which is presented here, is to minimize the error adjusting the weights and presenting learning rule is also pattern basis method. The training error on a pattern basis can be defined as the sum of the squared difference between desired output (target value) and the actual output. This can be expressed with the following equation. A description of the backpropagation algorithm can be found in [9].

$$Y(w) \equiv \frac{1}{2}\sum_{k=1}^{c}(t_k - z_k)^2 \tag{6}$$

In this work, we use neural network to model and evaluate in real-time how human subjects estimate image quality when distorted by changes in quality effecting parameters such as JPEG, JPEG2000, Gaussian blur, fast fading, and white noise. Here, the neural network architecture we adopt is multilayer feedforward neural network. Besides, the backpropagation is selected as learning algorithm for proposed framework. In classification view point, the backpropagation is capable of mapping nonlinear relations between the quality effecting parameters and DMOS.

4. EXPERIMENTAL RESULTS

The LIVE database [11] is selected as test bed to perform experiments and validate our approach. The LIVE image database contains 29 high-resolution 24 bits/pixel color reference images and theirs distorted images under five distortion types: JPEG2000, JPEG, White noise, Gaussian blur, and bit errors. Here, we chose 50 distorted images from each distortion type. Each distorted image has a computed Difference Mean Opinion Score (DMOS) ranging from 1 to 100. JPEG2000 images were generated using various bit rates. White noise images were obtained using White Gaussian noise. Gaussian kernel was used to create Gaussian blurred images. Fast-fading Rayleigh channel model was utilized to generate transmission errors in JPEG2000 bit stream.

The experiment consists of two major steps, namely, training and testing. The training section encompasses three aspects as follows; creating feature vectors, obtaining target vectors and designing neural network architecture. To create feature vector, we divide an image into grids such that 8 by 8 sliding window can scan through entire image. In this 8x8 window, statistical features such as mean, standard deviation and covariance are computed for each original and distorted image pair for a batch process. This process is repeated for each 4 pairs of original and distorted images among 5 pairs. The reason is that 4 pairs of images (original and distorted) are used for training and 1 pair is considered for testing purposes. By doing so, the feature vectors are generated for a type of images. This is for only one batch training process. This process is repeated for 50 image pairs for each distortion type. Obtaining target vectors are essentially based on subjective score, namely DMOS, carried out by human observers. The DMOS value is already provided for each distorted image given in the LIVE database set. To be able obtain the DMOS value corresponding each 8x8 window, we use mean weighed technique as follows. We calculate mean of each window and entire image. The DMOS score corresponds to mean of the entire image, and computed DMOS is assigned each window based on theirs weighed mean. With this fashion, target vectors are generated corresponding feature vectors. As we mentioned in the previous section selected neural network architecture is multilayer feedforward neural network. The backpropagation is chosen as learning algorithm for proposed framework. The number of hidden layers is composed of 3 and, each hidden layer consists of 6 neurons. The logistic sigmoid activation function is used in the hidden layers and the linear activation function is employed in the output layer. Using feature vectors and target vectors under determined neural network architecture, training process is achieved which gives a net which is saved for testing section. The MATLAB training interface screen shot is depicted in Figure 2 for training section.

In the testing section, the first pair image on the left, among 5 pairs for the same type of images is used to yield input vectors with the same method explained above, as the MATLAB testing interface screen shot illustrated in Figure 3. After the input vectors are fed into the neural network system, the ultimate goal is obtain DMOS score. The output of neural network would be predicted DMOS score. This process is repeated for all 250 images under various distortion types of the original images. With this procedure, obtained data is depicted in the scatter plot in Figure 4. Non-linear regression analysis was performed to fit data. Each sample point in the scatter plots has corresponding DMOS and neural network output values. DMOS values are represented in the y-axis, whereas predicted DMOS values are located in the x-axis of scatter plot. As shown in Figure 4, as DMOS values increase, predicted DMOS values increase meaning that they are in proportion.

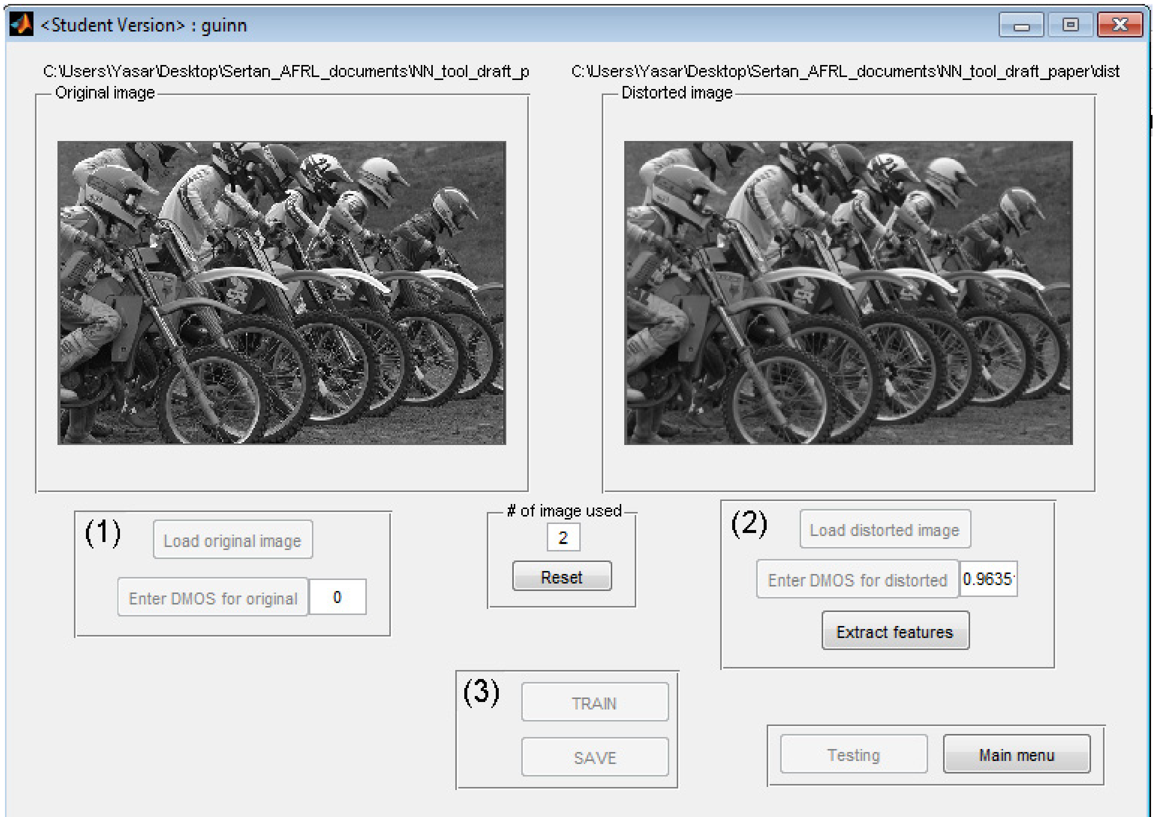

Figure 2: The MATLAB training interface

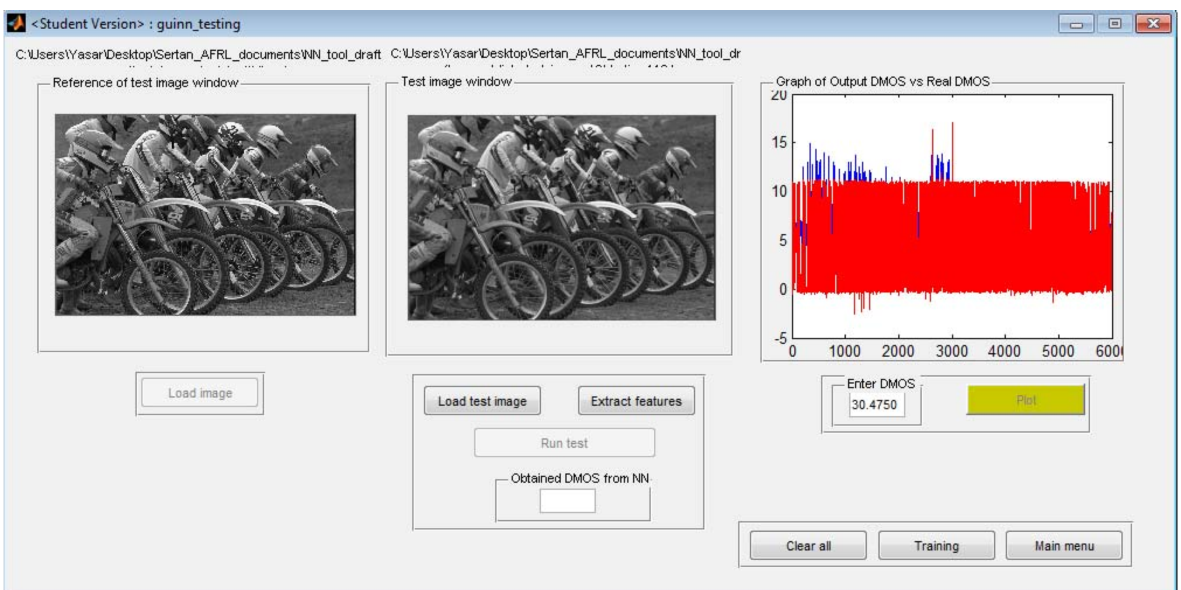

Figure 3: The MATLAB testing interface

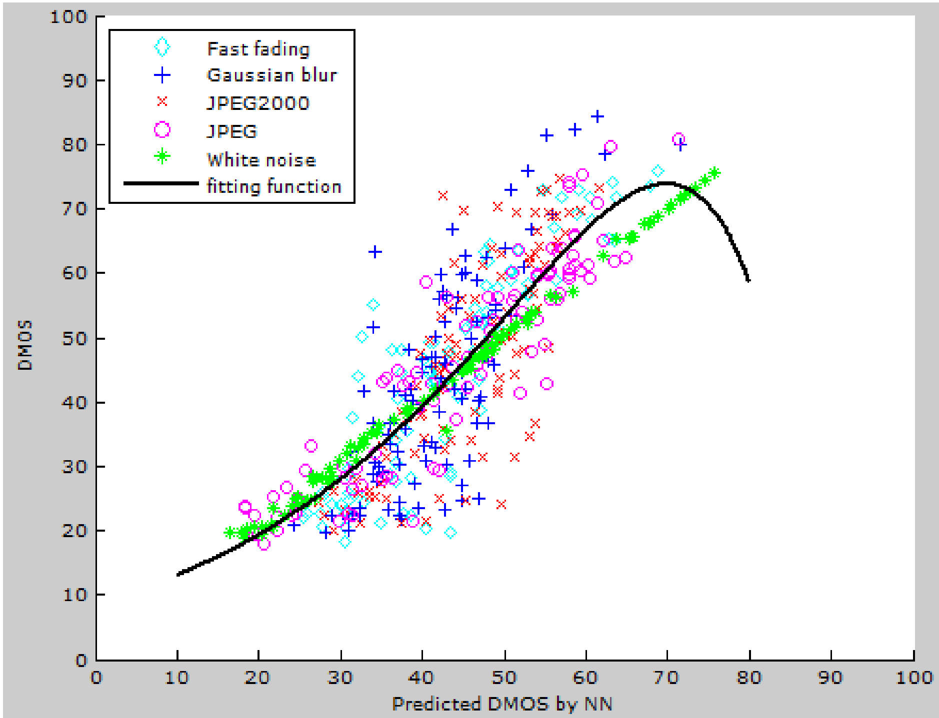

Figure 4: Scatter plot of DMOS vs. Predicted DMOS by NN

5. CONCLUSIONS

In this research, the primary goal is to design a model based on multilayer neural network for predicting the subjective image quality score, known as the difference mean opinion score carried out by participants, in image quality assessment. Along with the application of this technique, statistical features are extracted from both test and original images using a handy tool. We validate our proposed method using the LIVE Image Dataset Release 2. Based on our proposed experiment, the obtained output of Neural Network correlates well with the DMOS values. The trained Neural Network can obtain the ability to imitate human observers..

REFERENCES

[1] Wang Z., Bovik A. C., *Modern Image Quality Assessment*, Morgan & Claypool, 2006

[2] Wang Z., Bovik A. C., “Mean Squared Error: Love It or Leave It?”, IEEE Signal processing magazine, vol 1, no 1, January , 98-118, (2009).

[3] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” IEEE Trans. Image Process., vol. 14, no. 12, pp. 2117–2128, Dec. 2005

[4] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” IEEE Trans. Image Process., vol. 15, no. 11, pp. 3440–3451, Nov. 2006

[5] Gastaldo P., Zunino R., Vicario E., Heynderickx I., “CPB neural network for objective assessment of image quality”, Proc. IJCNN, Portland USA, pp. 194-9, 2003.

[6] Bouzerdoum A., Havstad A., and Beghdadi A., “Image quality assessment using a neural network approach,” in Proc. 4th IEEE Int. Symp.Signal Process. Inf. Technol., pp. 330–333,2004.

[7] Gastaldo P., Zunino R., "Neural networks for the no-reference assessment of perceived quality", J. Electron. Imaging 14, 033004, Aug 17, 2005.

[8] Wang Z., Bovik A C., “Image Quality Assessment: From error Visibility to structural Similarity” IEEE Transaction on Image Processing, vol 15, no 4, 600-612, April (2004).

[9] R. Duta, P. Hart, and D. Stork, *Pattern Classification*, New York, NY: John Wiley and Sons Inc., 2001.

[10] T. M. Mitchell, *Machine Learning*, McGraw-Hill International Editions, 1997.

[11] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2", http://live.ece.utexas.edu/research/quality