# Image Quality Assessment using a Neural Network Approach

A. Bouzerdoum[1], A. Havstad[2], and A. Beghdadi*

School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, NSW 2522, AUSTRALIA
Email: a.bouzerdoum@ieee.org, ahavstad@hades.fste.ac.cowan.edu.au

*L2TI, Institut Galilée, Université Paris 13, 93430 Villetaneuse, FRANCE
Email: Beghdadi@galilee.univ-paris13.fr

***Abstract*—*In this paper, we propose a neural network approach to image quality assessment. In particular, the neural network measures the quality of an image by predicting the mean opinion score (MOS) of human observers, using a set of key features extracted from the original and test images. Experimental results, using 352 JPEG/JPEG2000 comp-ressed images, show that the neural network outputs correlate highly with the MOS scores, and therefore, the neural network can easily serve as a correlate to subjective image quality assessment. Using 10-fold cross-validation, the predicted MOS values have a linear correlation coefficient of 0.9744, a Spearman ranked correlation of 0.9690, a mean absolute error of 3.75%, and an rms error of 4.77%. These results compare very favorably with the results obtained with other methods, such as the structural similarity index of Wang et al. [17].*

*Keywords—Image Quality Assessment, Neural Networks, Mean Opinion Score, Multilayer Perceptron.*

## I. INTRODUCTION

Image quality assessment (IQA) plays a very crucial role in image and video processing. The aim is to replace human judgment of perceived image quality with a machine evaluation. As a consequence, over the past three decades a large effort has been devoted to developing IQA measures that try to mimic human perception [1]-[10]. While many methods and models still rely on simple measures, such as the *peak-signal-to-noise-ratio* (PSNR) and the *mean-squared error* (MSE), many others use sophisticated signal processing techniques, such as multi-channel filtering [4]-[5], discrete cosine transform [7]-[8], multi-scale Wavelet decompositions [9]-[10], and Wigner-Ville distribution [11]. To date, however, it has been very difficult to find a reliable objective measure that correlates very highly with human perception [12].

Since invariably the end user of visual information is the human observer, it is generally recognized that subjective IQA methods are the ultimate solution. However, subjective measures are difficult to design and time consuming to compute; furthermore, they cannot be readily incorporated into the design and optimization of image and video processing algorithms, such as compression and image enhancement. For this reason, there has been an increasing interest in objective IQA techniques that can automatically predict or approximate the perceived image quality. Watson and Malo proposed a class of distortion metrics for video quality measurement, based on the standard observer vision model [13]. Gastaldo et al. used *continuous back-propagation* (CBP) *neural* networks to assess the quality of MPEG2 video streams [15]; the neural networks were trained to predict human ratings of video streams. The same type of neural networks was used to assess the quality of images that are processed by an enhancement algorithm [14]; here, the networks were trained to predict whether the quality of the processed image is better or worse than that of the original one. Wang et al. used second order statistics of the original and distorted images to compute a measure of image quality, which they named the *structural similarity* (SSIM) *index* [17]. They tested this measure on 344 (JPEG and JPEG2000) compressed images and compared the results with the *mean opinion scores* (MOS) of human observers; they found that the mean SSIM (MSSIM) scores correlate very well with the MOS, after applying logistic regression. Furthermore, the MSSIM was compared with other IQA measures and found to perform better than them.

In theory, artificial neural networks can approximate a continuous mapping to any arbitrary accuracy; therefore, they may be well suited to learning the salient characteristics of human perception. In this paper, we propose a method for image quality assessment based on neural networks. More specifically, a feedforward neural network, namely the *multilayer perceptron* (MLP), is trained to predict directly the MOS of JPEG and JPEG2000 compressed images. The proposed method is tested on 352 images and its performance is compared to that of the MSSIM of Wang et al. [17].

The paper is organized as follows. In the next section, image quality assessment methods are described briefly; in particular the MSSIM is introduced and discussed. Section III introduces the MLP neural network and the new IQA based on neural networks. Section IV presents the experimental results and comparisons between the MSSIM and the new Index. Finally, Section V concludes the paper.

[1] A. Bouzerdoum was a visiting professor at L2TI, Institut Galilée, Université Paris 13, for the period May 22 to June 30, 2004.

[2] A. Havstad was with Edith Cowan university, Perth, Australia

0-7803-8689-2/04/$20.00 ©2004 IEEE



Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 11:57:04 UTC from IEEE Xplore. Restrictions apply.

## II. IMAGE QUALITY ASSESSMENT

Image quality assessment methods can be categorized into three approaches: *full-reference* IQA, "blind" or *no-reference* IQA, *reduced-reference* IQA. In the full-reference IQA, a copy of the original image is available, with which the distorted image is compared. In this class of methods, the image quality metric measures image fidelity. By contrast, in the no-reference approach image quality is assessed based solely on the information content of the test image; that is, there is no reference image with which the test image can be compared. In the reduce-reference approach, only partial information about the original image is available. The neural network approach we propose here is a full-reference approach, where the fidelity of a test image is computed based on features extracted from the reference and test images.

### *II.1 Subjective Versus Objective Measures*

There are two main classes of IQA metrics: objective and subjective methods. While objective methods attempt to quantify the amount of degradation present in the image using a well-defined mathematical model, subjective measures are based on evaluation by human observers.

The *mean opinion score* (MOS) is the most common approach for subjective image quality assessment. Here a group of people is asked to visually compare an original image with a degraded image and estimate the image quality of the degraded image, and the mean score is taken as the image quality index. While this process reflects more faithfully human perception, it is time consuming and impractical to use in conjunction with other image processing algorithms. For this reason, there is strong interest in developing objective methods that correlate very well with the subjective assessment.

There are six classes of objective quality or distortion assessment methods:

- Pixel difference-based measurement: peak signal-to-noise ratio (PSNR) and the mean-squared error.
- Correlation-based measures: correlation of pixels, or of the vector angular directions.
- Edge-based measures: displacement of edge positions or their consistency across resolution levels.
- Spectral distance-based measures: measuring the magnitude and/or phase spectral discrepancies.
- Context-based measures: penalties based on various functions of the multidimensional context probability.
- Human Visual System (HVS) based measures: measure image quality by incorporating aspects of the human visual system characteristics. The quality of an image, as perceived by a human, depends on many factors, such as contrast, color, spatial frequency and masking effects.

By far the most common objective IQA methods are the pixel difference-based metrics because they have low computational complexity, and can easily be incorporated into other image processing algorithms. They are also independent of the viewing conditions and the individual observers. However, such simple measures, which do not take into account the HVS characteristics, are not adequate for describing perceptual image quality. Other more sophisticated measures do exist, such as the Universal Image Quality index (UIQI) [16] and the Structural Similarity (SSIM) Index [17], which are better correlated with subjective image quality.

### *II.2 Structural Similarity Index*

In 2000, Wang and Bovik proposed a measure the *universal image quality index* (UIQI) [16], where the comparison between the reference and test images is broken down into three different comparisons: luminance, contrast, and structural comparisons. The luminance comparison $l(x,y)$ between a reference image $X$ and a test image $Y$ is describe by

$$l(x,y)=\frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2},$$

where $\mu_x$ and $\mu_y$ denote the mean values of the images $X$ and $Y$, respectively. The contrast comparison is defined as

$$c(x,y)=\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2},$$

where $\sigma_x$ and $\sigma_y$ are the standard deviations of $X$ and $Y$, respectively. The structural comparison is given by

$$s(x,y)=\frac{\sigma_{xy}}{\sigma_x\sigma_y},$$

where $\sigma_{xy}$ is the covariance of $X$ and $Y$.

Based on these three comparison measures, the UIQI was defined as

$$\mathrm{UIQI}(x,y)=l(x,y)c(x,y)s(x,y)=\frac{4\mu_x\mu_y\sigma_{xy}}{\left(\mu_x^2+\mu_y^2\right)\left(\sigma_x^2+\sigma_y^2\right)}$$

The UIQI is a simple measure, which depends solely on first and second order statistics of the reference and test images. However, it is somewhat unstable, especially at uniform areas, where the denominator term is very small. Furthermore, rigorous tests showed that the UIQI doesn't correlate well with subjective assessment.

In order to alleviate the problem of stability and improve the correlation between the objective and subjective measures, Wang et al. [17] proposed the *structural similarity*



Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 11:57:04 UTC from IEEE Xplore. Restrictions apply.

*index* (SSIM) as an improvement to the UIQI. The SSIM has been defined as follows [17]:

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

$$C_1=(K_1L)^2,\quad C_2=(K_2L)^2$$

where $L$ is the dynamic range of the pixel values (255 for 8-bit images), and $C_1$ and $C_2$ are small positive constants.

At every pixel $(i, j)$, a local SSIM index, SSIM$(i, j)$, is defined by evaluating the mean, standard deviation and covariance on a local neighborhood $N_{ij}$, around that pixel. The overall image quality is measured by the *mean SSIM* (MSSIM) index given by

$$\mathrm{MSSIM}=\frac{1}{M}\sum_i\sum_j\mathrm{SSIM}(i,j)$$

where $M$ is the total number of local SSIM indexes.

Wang et al. compared the MSSIM and the MOS of human assessors, using a database of JPEG and JPEG2000 compressed images at various bit rates. They found that although the MSSIM does not exhibit a linear relationship with the MOS, it is well correlated with it when the MOS is estimated from the MSSIM using nonlinear regression. Furthermore, a comparison with other IQA methods, using different metrics, showed that the MSSIM predicts the MOS better than existing IQA methods [17].

## III. NEURAL NETWORKS

Neural networks have the ability to learn complex data structures and approximate any continuous mapping. They have the advantage of working fast (after a training phase) even with large amounts of data. The results presented in this paper are based on a multilayer feedforward network architecture, known as the *multilayer perception* (MLP). The MLP is a powerful tool that has been used extensively for classification, nonlinear regression, speech recognition, hand-written character recognition and many other applications. The elementary processing unit in a MLP is called a *neuron* or *perceptron*. It consists of a set of input synapses, through which the input signals are received, a summing unit and a nonlinear *activation* transfer function. Each neuron performs a nonlinear transformation of its input vector; the input-output relationship is given by

$$\varphi(\boldsymbol{x})=f\left(\boldsymbol{w}^T\boldsymbol{x}+\theta\right),$$

where $\boldsymbol{w}$ is the synaptic weight vector, $\boldsymbol{x}$ is the input vector, $\theta$ is a constant called the bias, $\varphi(\boldsymbol{x})$ is the output signal, and $^T$ is the transpose operator.

An MLP architecture consists of a layer of input units, followed by one or more layers of processing units, called hidden layers, and one output layer. Information propagates, in a feedforward manner, from the input to the output layer [18]; the output signals represent the desired information. The input layer serves only as a relay of information and no information processing occurs at this layer. Before a network can operate to perform the desired task, it must be trained. The training process changes the training parameters of the network in such a way that the error between the network outputs and the target values (desired outputs) is minimized [18].

In this paper, we propose a method to predict the MOS of human observers using an MLP. Here the MLP is designed to predict the image fidelity using a set of key features extracted from the reference and test images. The features are extracted from small blocks (say 8x8 or 16x16), and then they are fed as inputs to the network, which estimates the image quality of the corresponding block. The overall image quality is estimated by averaging the estimated quality measures of the individual blocks. Using features extracted from small regions has the advantage that the network becomes independent of image size. The key features are based on the features of Wang and Bovik with some modifications. Six features, extracted from the original and test images, were used as inputs to the network: the two means, the two standard derivations, the covariance, and the mean-squared error between the test and reference blocks.

## IV. EXPERIMENTAL RESULTS

The experimental results are based on a database of distorted images and their corresponding mean-opinion scores. This database, which can be found at Zhou Wang's Homepage [19], consists of images that have been compressed by JPEG and JPEG2000 at different bit rates. We used 354 pairs of reference and test images to train and test the neural network: 343 pairs were taken from the database and 9 pairs were added. The 9 added pairs have identical reference and test images, and hence their MOS values are set to 100%. These images are added so as to test the network on images with maximum MOS values.

The results presented here are obtained from using an MLP architecture with 6 inputs, 6 neurons in the first hidden layer, 6 neurons in second hidden layer, and 1 output neuron. We used the logistic sigmoid activation function in the hidden layers and the linear activation function in the output layer. Ten networks, with the same architecture, were trained and tested using the method of 10-fold cross-validation. Each network was trained on 90% of the images from the available set, and the other 10% were used to test the performance of the network; the test set is shifted for each network. In this way, all the images in the database are used to test the network. The desired output of the neural network is the MOS value of the test image.

To test the ability of the neural network to predict the MOS, its performance is assessed using different metrics, as recommended by VQEG (Video Quality Expert Group) in [20]. For a metric relating to performance accuracy we use



Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 11:57:04 UTC from IEEE Xplore. Restrictions apply.

Pearson's linear correlation coefficient $\rho$. Mono-tonicity of the model is assessed using Spearman's rank-order correlation coefficient $\rho_r$. We also used the root mean square error (RMSE), the mean absolute error, (MAE) and the standard error ($\sigma_E$). The performance of the neural network is compared to that of the MSSIM. First, logistic regression is applied to find a nonlinear mapping between the MSSIM scores and the MOS. The 10-fold cross-validation method is also applied to assess the fit of the nonlinear regression, in the same way as with the neural network.

Table 1 presents the different assessment metrics for the neural network and MSSIM predictions. Clearly the neural network outperforms the MSSIM, even after nonlinear regression, for every metric. Figure 1 (a) and (b) show the fit between the objective and subjective measures. It is clear that the fit between the neural network output and the MOS is linear, whereas, as expected, the fit between the MSSIM and the MOS is nonlinear. Fig. 1 (c) and (d) show the error histograms of the two fits.

Table 1. Comparison of image quality assessment using neural networks and MSSIM with and without logistic regression: $\rho$ is Pearson's linear correlation coefficient, $\rho_r$ is Spearman's rank-order correlation coefficient, RMSE is the root mean square error, MAE is the mean absolute error, $\sigma_E$ is the standard error.

| Metric | $\rho$ | $\rho_r$ | RMSE | MAE | $\sigma_E$ |
|---|---|---|---|---|---|
| MSSIM | 0.9114 | 0.9499 | 27.951 | 26.320 | 9.422 |
| MSSIM-fit | 0.9517 | 0.9492 | 6.512 | 5.396 | 6.521 |
| NNet | 0.9744 | 0.9690 | 4.775 | 3.750 | 4.774 |

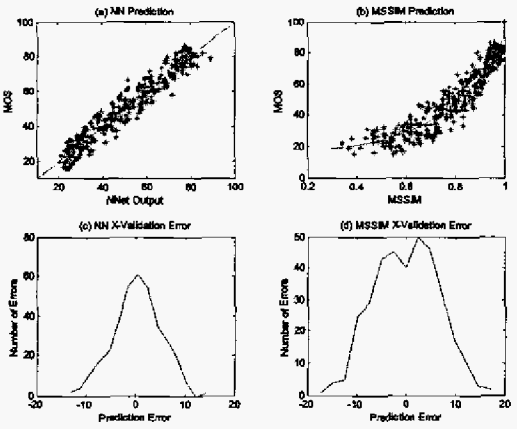


Fig. 1. MOS vs objective assessment: (a) MOS vs NN output, (b) MOS vs MSSIM, (c) and (d) error histograms for (a) and (b).

## V. CONCLUSION

A new approach for image quality assessment using neural networks has been presented in this paper. Experimental results show that a neural network can be trained to accurately predict the MOS values using 6 features from the reference and test images. When compared with the MSSIM of Wang et al. [17], the neural network was found to correlate better with the subjective assessment than the MSSIM does.

## VI. REFERENCES


[1]. J. L. Mannos and D. J. Sakrison,"The effects of a visual fidelity criterion on the encoding of images," IEEE *Trans. on Information Theory*, Vol. 10, pp. 525–536, 1974.

[2]. G. C. Higgins, "Image quality criteria," *J. Applied Photgr. Eng.*, Vol. 3, No. 2, pp. 53–60, 1977.

[3]. H. L. Snyder, "Image quality: measures and visual perfor-mance," *Flat Panel Displays and CRTs*, L. E. Tannas Jr. (ed.), pp 70–90, Van Nostrand Reinhold, New York, 1985.

[4]. S. Daly, "The visual difference predictor: an algorithm for the assessment of image fidelity," in *Human Vision, Visual Processing, and Digital Display*, Proc. SPIE, Vol. 1666, pp. 2–15, San Jose, CA, 1992.

[5]. D. J. Heeger and P. C. Teo, "A model of perceptual image fidelity," *Proc. IEEE International Conference on Image Processing*, Vol. 2, pp. 343–345, 23–26 Oct. 1995.

[6]. C. J. van den Branden Lambrecht (ed.), "Special Issue on image and video quality metrics," *Signal Processing*, Vol. 70, Nov. 1998.

[7]. J. Malo, A. M. Pons, and J. M. Artigas, "Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain," *Image and Vision Computing*, Vol. 15, pp. 535–548, 1997.

[8]. A. B. Watson, J. Hu, and J. F. McGowan, "Digital video quality metric based on human vision," *J. of Electronic Imaging*, Vol. 10, No. 1, pp. 20–29, 2001.

[9]. Y. K. Lai and C.-C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," *J. Visual Com-munication and Image Repres.*, Vol. 11, pp. 17–40, 2000.

[10]. A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," *Proc. Seventh Intern. Symp. Signal Proces. its Applications* (ISSPA-2003), Vol. 1, pp. 485–488, Paris, 1–4 July 2003.

[11]. A. Beghdadi, "Design of an image distortion measure using spatial/spatial frequency analysis," *Proc. First Inter. Symposium on Control, Communications and Signal Processing*, pp. 29–32, 21–24 March 2004.

[12]. Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" *Proc. IEEE Inter. Conference Acoustics, Speech, and Signal Processing* (ICASSP-2002), Vol. 4, pp. 3313–3316, Orlando, FL, 13-17 May 2002.

[13]. A. B. Watson and J. Malo, "Video quality measures based on the standard observer," *Proc. IEEE Int. Conf. Image Proc.*, Vol. III, pp. 41–44, Rochester, 22–25 Sep. 2002.

[14]. P. Carrai, I. Heynderickx, P. Gastaldo, and R. Zunino, "Image quality assessment by using neural networks," *Proc. IEEE Inter. Symposium on Circuits and Systems* (ISCAS-2001), pp. V-253–256, 6–9 May 2001, Sydney.

[15]. P. Gastaldo, S. Rovetta, and R. Zunino, "Objective Quality Assessment of MPEG-2 Video Streams by Using CBP Neural Networks," IEEE *Trans. Neural networks*, Vol. 13, pp. 939–947, 2002.

[16]. Z. Wang and A. C. Bovik, "A universal image quality index," IEEE *Signal Proc. Letters*, vol. 9, pp. 81–84, 2002.

[17]. Z. Wang, 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, pp. 600–612, Feb. 2004.

[18]. J. M. Zurada, *Introduction to artificial neural systems*: PWS publishing company, 1992.

[19]. Z. Wang's Homepage, http://www.cns.nyu.edu/~zwang/

[20]. VQEG. (2000, Mar.) *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. http://www.vqeg.org/




Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 11:57:04 UTC from IEEE Xplore. Restrictions apply.