Deep Neural Network Based Full Reference and No Reference Image Quality Assessment

S.Sathya moorthy
Assistant Professor,
Department of Computer Science,
Bharathiar University Arts and Science College
,Gudalur, The Nilgiris ,
Tamilnadu , India.
ssmoorthy4003@gmail.com

Dr.M.Dharmalingam
Assistant Professor,
Department of Computer Science,
Bharathiar University Arts and Science College ,
Modakkurichi, Erode ,
Tamilnadu , India.
emdharama@gmail.com.

***Abstract* -A deep Neural networks is an Artificial Neural Network (ANN) with multiple hidden layers between the input and output layers similar to shallow ANN. We present a deep neural network based approach to image quality assessment (IQA).We evaluate the proposed approach LIVE ,CISQ and TID 2013 databases as well as the LIVE in the wild image quality challenge state-of-the-art NR(No Reference) and FR (Full Reference) IQA(Image Quality Assessment).With the rise of machine learning ,recently a third category of IQA emerged comprising approaches that are purely data driven, do not rely on any explicit model and allow for end-to-end optimization of feature extraction and regression. Our approach presented in this paper belongs to this neural network for FR and NR IQA.**

Index Terms –Artificial Neural Network, No Reference, Full Reference, Image Quality Assessment.

I. INTRODUCTION

With the launch of networks handheld devices which can captured,store,compress,send and display a variety of audiovisual stimulai;High Definition Television(HDTV);Streaming Internet Protocol TV(IPTV) and websites such as youtube,facebook and Flicker etc.., an enormous amount of visual data is making its way to consumers. In deep-learning networks, each layer of nodes trains on a distinct set of features based on the previous layer's output. The further you advance into the neural net, the more complex the features your nodes can recognize, since they aggregate and recombine features from the previous layer.

The LIVE database comprises 779 quality annotated images based on 29 source reference images that are subject to 5 different types of distortions at different distortion levels. Distortion types are JP2K compression, JPEG compression, additive white Gaussian noise, Gaussian blur and a simulated fast fading Rayleigh channel. Quality ratings were collected using a single-stimulus methodology; scores from different test sessions were aligned. Resulting DMOS values span the range [0, 1], where a lower value indicates better visual quality.The LIVE In the Wild Image Quality Challenge Database (CLIVE) comprises 1162 images taken under real life conditions with a large variety of objects and scenes captured under varying luminance conditions using different cameras.Beyond capturing a large variety of image contents including pictures of objects,people and both outdoor and indoor scences ,such a database should also account for the various and diverse frequently used commercial image capture devices. The corpus of pictures should also have been obtained under varied illumination conditions and given the propensity of users to acquire their pictures in the improper ways

The TID2013 image quality database is an extension of the earlier published TID2008 image quality database containing 3000 quality annotated images based on 25 source reference images distorted by 24 different distortion types at 5 distortion levels each. The distortion types cover a wide range from simple Gaussian noise or blur over compression distortions such as JPEG to more exotic distortion types such as non-eccentricity pattern noise. This makes the TID2013 a more challenging database for the evaluation of IQMs. The rating procedure differs from the one used for the construction of LIVE, as it employed a competition-like double stimulus procedure. The obtained mean opinion score (MOS) values lie in the range [0, 9], where larger MOS indicate better visual quality.The CISQ image quality database contains 866 quality annotated images. 30 reference images are distorted by JPEG compression, JP2K compression, Gaussian blur, Gaussian white noise, Gaussian pink noise or contrast change. For quality assessment, subjects were asked to position distorted images horizontally on a monitor according to its visual quality. After alignment and normalization resulting they should exhibit a broad medium of authentic quality types,mixers and distortion severities.The result of new resources is called LIVE in the Wild Image Quality Challenge database.

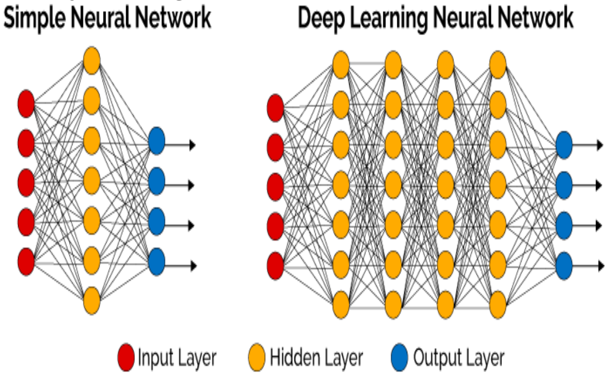

Fig.1.Deep learning neural network

II. RELATED WORK

A. *No Reference Image Quality Assessment*

With the launch of networked handheld devices which can capture, store, compress, send and display a variety of audiovisual stimuli; high definition television (HDTV); streaming Internet protocol TV (IPTV) and websites such as Youtube, Facebook and Flickr etc., an enormous amount of visual data of visual data is making its way to consumers. NR IQA refers to automatic quality assessment of an image using an algorithm such, that the only information that the algorithm receives before it makes a prediction on quality is the distorted image whose quality is being accessed. On the other end of the spectrum lie full reference [FR] algorithms that require as input not only the distorted image but also a clean, pristine reference image with respect to which the quality of the distorted image assessed, Somewhere between, these two extremes lie reduce reference (RR) approaches that possess some information regarding the reference image (e.g., a watermark), but not a actual image itself, a part from the distorted image.

The goal of an objective no reference image quality assessment (NR IQA) model is as follows: given an image (possibly distorted) and no other additional information, automatically and accurately predict its perceptual Quality. Given that the ultimate receivers of these images are humans the only reliable way to understand and predict the effect of distortion on a typical person’s viewing experience is to capture options. From a large sample of human subjects, which is termed subjective image quality assessment? The learning frame work adopted in this work is illustrated in figure key components in this framework include.1) Local feature extraction 2) Codebook construction 3) Local feature encoding 4) Feature pooling.

B. *Full Reference Image Quality Assessment*

Machine evaluation of image and video quality is important for many image processing systems, such as those for acquisition, compression, restoration, enhancement, reproduction, etc. The goal of quality assessment research is to design algorithms for *objective* evaluation of quality in a way that is consistent with subjective human evaluation. By “consistent” we mean that the algorithm’s assessments of quality should be in close agreement with human judgments, regardless of the type of distortion corrupting the image, the content of the image, or strength of the distortion.

III.COMPARISON WITH FULL REFERENCE AND NO REFERENCE IQA ALGORITHMS

A. *Two FR-IQA measures*:

Peak-signal-to-noise-ratio [PSNR] and the structural similarity index [SSIM] are tested for comparison. These results of FR-IQA measures are obtained as described above for obtaining results of our method. We also report the performance of two recent state-of-the-art NR-IQA measures. DIVINE and BLINDS Parameters: In the following experiments for the live the TID2008 database, we used patch size of 7-by-7 and for the TID2008 database, the number of patches extracted from each image is 10,000 and codebook size is fixed as 10,000.A statistical Evaluation of Recent full reference image quality Assessment algorithm.

B. *The image database*:

1) Source image content: Twenty-nine high resolution and high quality color images were collected from the Internet and photographic CD-ROMs. These images include pictures of faces, people, animals, close-up shots, wide-angle shots, nature scenes, man-made objects, images with distinct foreground/background Configurations and images without any specific object of interest.

2) Image Distortion Types: We chose to distort the source images using five different image distortion types that could occur in real-world applications. The distortion types are as follows.

a)JPEG2000 compression: The distorted images were generated by compressing the reference images (full color) using 1.JPEG2000 at bit rates ranging from

0.028 bits per pixel (bpp) to 3.15 bpp. Kakadu version 2.2 was used to generate the JPEG2000 compressed images.

b)JPEG compression: The distorted images were generated by compressing the reference images (full color) using JPEG at bit rates ranging from 0.15 bpp to 3.34 bpp. The implementation used was MATLAB's imwrite function.

c) White Noise: White Gaussian noise of standard deviation was added to the RGB components of the images after scaling the three components between 0 and 1. The same was used for R, G, and B components. The values of used were between 0.012 and 2.0. The distorted components were clipped between 0 and 1, and rescaled to the range 0 to 255.

d) Gaussian Blur: The R, G, and B components were filtered using a circular-symmetric 2-D Gaussian kernel of standard deviation pixels. The three color components of the image were blurred using the same kernel. The values of ranged from 0.42 to 15 pixels.

e)Simulated Fast Fading Rayleigh (wireless) Channel: Images were distorted by bit errors during transmission of compressed JPEG2000 bit stream over a simulated wireless channel. Receiver SNR was varied to generate bit streams corrupted with different proportion of bit errors. The source JPEG2000 bit stream was generated using the same codec as above, but with error resilience features enabled, and with 64 64 precincts. The source rate was fixed to 2.5 bits per pixel for all images, and no error concealment algorithm was employed. The receiver SNR used to vary the distortion strengths ranged from 15.5 to 26.1 dB.

In fact, choosing new types of distortions to be exemplified in TID2013, we had many options. On one hand, creation of a new database and MOS obtaining for it is not an easy task and new databases for metric verification are designed not often. Thus, having decided to create TID2013, we had to sufficiently modify and to add types of distortions which are really important from theoretical and practical viewpoints. On the other hand, we had to take into account some specific aspects to carry out experiments. First, experiment for each test image should not take too much time to prevent observer's tiredness. Second, the number of distorted images for each reference image has to be even to make each distorted image participating in equal number of visual quality comparisons. Taking all these into consideration, we have decided to introduce just seven new types of distortions to get the total number of distortion types equal to 24. With five levels of distortions, there are 120 distorted versions of each reference color image now (instead of 68 in TID2008).

The following distortion types have been introduced: change of color saturation (#18), Multiplicative Gaussian Noise (#19), Comfort noise (#20), Lossy compression of Noisy images (#21), Image color quantization with dither (#22), Chromatic aberrations (#23), sparse sampling and reconstruction (#24).

As it is seen, three types of distortions in one or another way relate to color (## 18, 22, 23). Including of them has been motivated by the facts that peculiarities of color distortions are not well represented in already existing databases whilst they are very important in modern practice [7, 11, 22], in particular, for color image printing. Note that many existing HVS-metrics are calculated for each component of a color image separately and then aggregated. Thus, they are unable to account for peculiarities of color distortion perception.

More in detail, changes in color saturation can result from larger quantization of color components in JPEG compression and JPEG-based algorithms in compression of video. Distortion modeling was carried out after image transformation into YCbCr color space. Distortions "Image color quantization with dither" have been modeled using the Matlab function rgb2ind that converts RGB image to indexed image using dither. The number of quantization levels was adjusted individually to provide a desired PSNR. Chromatic aberrations were modeled by slight mutual shifting of R, G, and B components with respect to each other with further blurring of shifted components. Multiplicative Gaussian spatially uncorrelated noise has been chosen to represent a wide class of distortions caused by signal-dependent noise which takes place in many modern applications of CCD sensors, in medical, ultrasound and radar imaging. Recent experiments have clearly demonstrated that existing HVS-metrics do not allow characterizing visual quality of images corrupted by different types of signal-dependent noise adequately.

Comfort noise has been added into consideration to take into account a specific feature of human vision that it practically does not matter for it what realization of the noise takes place for a given image. Similarly, human vision is less sensible to changes in regions with texture. Lossy compression of noisy images is one more type of distortions important for practice. Really, original color images acquired by digital cameras in bad illumination conditions as well as video frames are noisy. After lossy compression, additional distortions are introduced. Thus, one needs to assess visual quality in conditions of two types of distortions with their aggregate effect (similar tasks arise

In non-reference quality metric design where, e.g., blur and blocking artifacts are present simultaneously). More in detail, additive white Gaussian noise with variance σ2 has been added and then lossy compression by the DCT-based coder ADCT with the quantization step 1.73σ has been performed. Finally, sparse sampling and reconstruction (compressive sensing) of images has become a hot topic of research in recent years. meanwhile, HVS-metrics are practically not exploited for characterizing visual quality of reconstructed images although urgent need in this is obvious. 108 More in detail, the method [28] of compressive sensing image reconstruction has been used by us in generating distorted images.

In this study the use of a deep CNN with an architecture largely inspired by the organization of the primates' visual cortex, comprising 10 convolutional layers and 5 pooling layers for feature extraction, and 2 fully connected layers for regression, in a general IQA setting and shows that network depth has a significant impact on performance. We start with addressing the problem of FR IQA in an end-to-end optimization framework. For that, we adapt the concept of siamese networks known from classification tasks by introducing a feature fusion stage in order to allow for a joint regression of the features extracted from the reference and the distorted image.

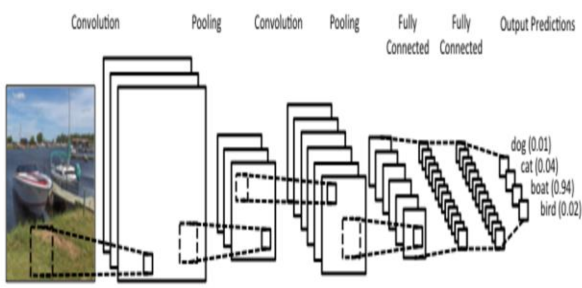

Fig.2.Architecture of Convolutional neural network

IV.NEURAL NETWORK-BASED FR IQA

Siamese networks have been used to learn similarity relations between two inputs. For this, the inputs are processed in parallel by two networks sharing their synaptic connection weights. This approach has been used for signature and face verification tasks, where the inputs are binarily classified as being of the same category or not. For FR IQA we employ a Siamese network for feature extraction. In order to use the extracted features for the regression problem of IQA,
feature extraction is followed by a feature fusion step. The fused features are input to the regression part of the network.

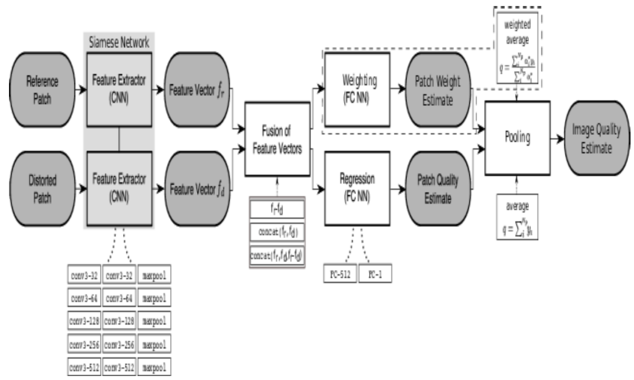

Fig. 3. Deep neural network model for FR IQA. Features are extracted from the distorted patch and the reference patch by a CNN and fused as difference, concatenation or concatenation supplementary with the difference vector. The fused feature vector is regressed to a patchwise quality estimate. The dashed-boxed branch of the network indicates an optional regression of the feature vector to a patchwise weight estimate that allows for pooling by weighted average patch aggregation.

For our IQA approach, images are subdivided into 32 × 32 sized patches that are input to the neural network. Local patch wise qualities are pooled into a global image wise quality estimate by simple or weighted average patch aggregation.

V.NETWORK ADAPTATIONS FOR NR IQA

Abolishing the branch that extracts features from the reference patch from the Siamese network is a straight forward approach to use the proposed deep network in a NR IQA context. As no features from the reference patch are available anymore, no feature pooling is necessary. However, both spatial pooling methods detailed in Section III-C are applicable for NR IQA as well.

The resulting approaches are referred to as Deep Image QuAlity Measure for NR IQA and Weighted Average Deep Image QuAlity Measure for NR IQA. This amounts to the same loss functions as for the FR IQA case.

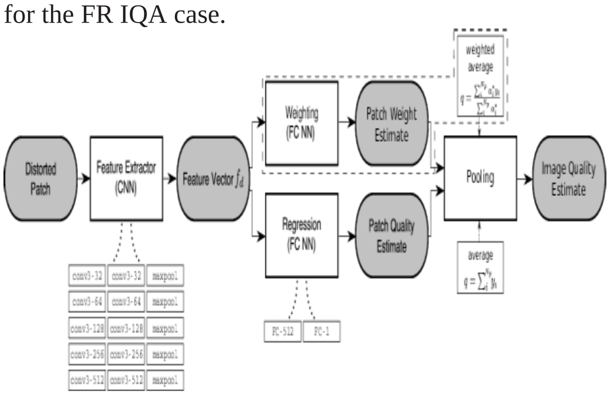

Fig. 4. Deep neural network for NR IQA. Features are extracted from the distorted patch by a CNN. The feature vector *fd* is regressed to a patchwise quality estimate. Patchwise estimates are aggregated to a global image quality estimate. The dashed-boxed branch of the network indicates an optional regression of the feature vector to a patchwise weight estimate that allows for pooling by weighted average patch aggregation.

VI RESULTS AND DISCUSSION

The proposed method is examined on the LIVE database consisting of quality annotated images that are subject to distortions of different kinds and differrent levels. The LIVE database is based on 29 source reference images, subject to 5 different types of distortions at three to five different distortion levels. MOS values were obtained under fairly controlled conditions. The TID2013 consists 25 colored reference images and 3000 differently distorted images, subject to 24 different distortion types. Subjective ratings were gathered by variations. The results from several viewing conditions of experiments in three different labs and on the internet were averaged.

Table 2 summarizes the performance the different proposed feature fusion schemes for LIVE and TID2013 databases. The table shows that the relationship between the two feature vectors can be learned by the model, but providing the difference f_r f_d explicitly leads to better results on both datasets.

Table 1.Benchmark IQA image datasets

Sl. No	Dataset	No.of reference images	No.of distorted images	No.of distortion types
1	LIVE	29	779	56
2	CSIQ	30	866	6
3	TID 2013	25	3000	24

Simple difference fusion f_r f_d has the advantage that it becomes zero if the feature vectors coincide (i.e., distorted image equals reference image), but it lacks flexibility, e.g. when only one of the feature vectors is informative. Due to the limited size of the training data set we did not evaluate more complex fusion techniques. However, although the differences in performance are rather marginal for both databases and aggregation methods, for further analysis the proposed method is evaluated based on concat(f_r; f_d; f_d f_r).

Table 2. Comparison of performance of the two suggested patch aggregation methods.

Dataset	Aggregation	f_d f_r	Concat (f_r; f_d)	Concat (f_r;f_d;f_d f_r)
LIVE	Average	0.976	0.974	0.976
	Weighted avg	0.982	0.977	0.982
TID 2013	Average	0.908	0.893	0.908
	Weighted avg	0.968	0.958	0.965

The LCC was computed on the validation set of one random split for each dataset and with N_p = 1024 random patches per image

Therefore the background is relative to the foreground less important for the globally perceived quality of the image. Also in the last example the weightings have a region separation effect, namely they separate the scores assigned to the sky from the scores assigned to the rest. The former ones overestimate the MOS value, thus are down weighted by the proposed method

VII CONCLUSION

We have studied the problem of the RR image QA by measuring the changes in suitably weighted entropies between the reference and distorted images in the wavelet domain. A distinguishing feature of the RRED indexes is that these algorithms do not depend on any parameters that need to be trained on databases. The algorithms differ in the nature of the distortion measurement (by computing quality in different orientated sub bands at different scales) and the quantity of the information required from the reference to compute quality (by filtering and sub sampling in every sub band).We plan to regularly update the versions of this database. In particular, updated versions will provide more reliable data (in statistical sense) due to taking into account the results of future experiments.

Moreover, new versions will include new types of distortion that take place in different applications of image processing and/or those distortions that might correspond to new peculiarities of HVS found in future experiments. The overall performance of the single-number algorithms may be further improved by better aligning the scores obtained for different distortion categories. This is a subject of future research. Even though a relative generic neural network is able to achieve high prediction performance, incorporating IQA specific adaptations to the architecture may lead to further improvements. Our results show that there is room for optimization in terms of feature dimensionality and balancing the ratio between network parameters. Here, prediction performance and generalization ability is important to be studied.

REFERENCES

[1] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik," No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, December 2012.

[2] Deepti Ghadiyaram and Alan C. Bovik," Massive Online Crowdsourced Study of Subjective and Objective . Picture Quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, January 2016.

[3] Peng Ye, Jayant Kumar, Le Kang, David Doermann,"Unsupervised Feature Learning Framework for No-reference Image Quality Assessment," Institute for Advanced Computer Studies University of Maryland,College Park, MD, USA.

[4] Jonathan Long, Evan Shelhamer, Trevor Darrell," Fully Convolutional Networks for Semantic Segmentation," UC Berkeley.

[5] Wei Zhang, Ali Borji, Zhou Wang, Patrick Le Callet, and Hantao Liu, "The Application of Visual Saliency Models in Objective Image Quality Assessment: A Statistical Evaluation," IEEE Trans. On Neural Networks And Learning Systems, vol. 27, no. 6, June 2016

[6] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik," A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, November 2006.

[7] Nikolay Ponomarenko (1), Oleg Ieremeiev (1) Vladimir Lukin(1), Karen Egiazarian (2),Lina Jin (2), Jaakko Astola (2), Benoit Vozel(3), Kacem Chehdi(3), Marco Carli (4), Federica Battisti (4), C.-C. Jay (5), "Color Image Database Tid2013: Peculiarities And Preliminary Results," (1) National Aerospace University,Kharkov, Ukraine (2) Tampere University of Technology, Tampere, Finland.(3) University of Rennes . 1, Lannion, France.(4) University of Roma III, Rome, Italy.(5) Media Communications Lab, USC Viterbi School of Engineering, USA.

[8] Lin Zhang, *Member*, Ying Shen and Hongyu Li." . VSI: A Visual Saliency-Induced Index for Perceptual . Image Quality Assessment," *IEEE Trans.Image . Process.*, vol. 23, no. 10, October 2012

[9] Sebastian Bosse , Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek," Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," *IEEE Trans. Image Process.*, . vol. 27, no. 1, January 2018.