# No-reference Image Quality Assessment using Convolutional Neural Network


Ravishankar Gaikwad
*Electronics and Communication Engineering*
*Visvesvaraya National Institute of Technology*
Nagpur, India
ravishankargaikwad00@gmail.com

Ankit Rahangdale
*Electronics and Communication Engineering*
*Visvesvaraya National Institute of Technology*
Nagpur, India
ankitr9949@gmail.com

Kashyap Mishra
*Electronics and Communication Engineering*
*Visvesvaraya National Institute of Technology*
Nagpur, India
kashyapmishra666@gmail.com

Yash Chandekar
*Electronics and Communication Engineering*
*Visvesvaraya National Institute of Technology*
Nagpur, India
yashchandekar975@gmail.com

Mallineni Bhaskar
*Electronics and Communication Engineering*
*Visvesvaraya National Institute of Technology*
Nagpur, India
bhaskarmallineni2017@gmail.com



***Abstract*— We propose no reference image quality assessment with the help of convolution neural network.Distorted image is cut into several number of patches.After preprocessing, convolution neural network is used to automatically extract features, two convolution neural network models are designed with different structures .Finally this models are ensembled to improve the accuracy of image quality assessment. Experiments have been performed on LIVE and VCL@FER databases and corresponding results show that the proposed method can successfully predict image quality score.**



***Keywords*— convolution neural network, image quality assessment,ensemble learning**


## I. Introduction

In our digital world digital images have very wide usage such as information expression, communication and entertainment. However, they are affected by many interference factors in the process of image acquisition, transmission, storage because of which images are getting distorted and affecting quality of images. The quality of image greatly affects the viewers subjective feelings and information acquisition. Therefore, it is really important to take accurate measures for image quality assessment.

Usually Image Quality Assessment (IQA) is of two types: subjective image quality assessment and objective image quality assessment. Subjective assessment is more accurate as it involves a large number of observers' scores over the images of their subjective feelings, however it is of less use in real life as it is time consuming, inflexible and computationally expensive. Hence objective image quality assessment obtains greater importance.

Objective image quality assessment can further be divided into few categories based on reference images using full reference (FR) IQA, reduced-reference (RR) IQA, no-reference (NR) IQA.

## II. Technologies used

### A. Convolution layer

We have used Convolution neural network algorithms to develop our model. This algorithm takes image as input.The objective of the convolution is to extract high level features such as edges etc.algorithm adapt in better way if we add more layers

*Activation function*: With the help of activation function ,features obtained in previous layers are passed to next layers,Rectified Linear Unit is the most commonly used activation function. Activation Function can also speed up training. It is expressed as

$$f(x) = \max(0,x)$$

*Optimizer used*: Optimizers are algorithms used to change the attributes of neural networks such as weights and learning rate in order to reduce losses.We have used ADAM optimizer.It is computationally efficient and requires little memory.It is also well suited for algorithms with large dataset

### B. Pooling layers

Pooling layer is an important part of our CNN model. Pooling layer basically extracts useful features from the selected filter size matrix. This layer effectively decreases the size of the input feature matrix without much loss of features and thus reduces computational requirements of the model. This also helps in preventing overfitting. Pooling layer has two hyper parameters viz. kernel size and step size. We have used two types of pooling layer: max pooling and average pooling.

### C. Ensembling

Ensembling is a technique which combines several weak models to get one single strong model. This helps in avoiding overfitting and high variance and thus improving accuracy. Whole dataset is divided into several overlapping parts, each of which is used to train different models which were combined to get a single and accurate model. Base model with greater accuracy are given higher weightage and one with

## III. Ensembling Convolutional Neural Network

### A. CNN Structures Used

Two CNN models of different depth have been used in this paper. Table 1. given below shows the model used. In the table '3X3, 16' means kernel size of convolution is 3X3 and 16 filters have been used. The entries in the pool column indicate the kernel size of the pooling layer.

Table 1: Layers of CNN

| CNN1 | | CNN2 | |
|---|---|---|---|
| conv | 3X3 , 8 | conv | 3X3 , 8 |
| conv | 3X3 , 16 | conv | 3X3 , 16 |
| conv | 3X3 , 32 | conv | 3X3 , 32 |
| pool | 2X2 | pool | 2X2 |
| conv | 3X3 , 64 | conv | 3X3 , 64 |
| pool | 2X2 | pool | 2X2 |
| conv | 3X3 , 128 | conv | 3X3 , 128 |
| pool | 2X2 | conv | 3X3 ,256 |

After the convolutional and pooling layers there are fully connected layers in each CNN and the last layer of each CNN has only 1 node at the end which indicates the quality score of the image. Each CNN has been trained for 10 epochs.

The results of the two CNN networks have been combined using following

$$Y(x)=(2/3)\, y_1(x) \;+\; (1/3)\, y_2(x)$$

Here, (⅔) weight is assigned to predicted values of CNN having lower mean squared error.

### B. Image Preprocessing

CNN requires a large amount of data. To meet this demand,distorted images are cut into small patches of size 40 x40 for Live database and 50 x 50 in VCL@FER database. As the images are uniformly distorted we assign each patch a same quality score.

### C. Training

In the training stage,the quality score of each patch is predicted through each CNN model.Score of the whole image is the average score of the predicted image patch.Quality of whole image is calculated as

$$\hat{q} = \frac{1}{N_p}\Sigma_i^{N_p} y_i$$

where $y_i$ is the score of $i^{th}$ patch.

## IV. RESULT AND ANALYSIS

### A. Dataset used and Processing

The LIVE database consists of 29 types of coloured reference images. various types of distortions and different degrees into images have been appended by the researchers with various techniques resulting in a total of 779 different quality images .The obtained distorted images has a resolution range varying from 512 x 768 to 438 x 634.JPEG2000 compression distortion,white noise distortion ,JPEG compression distortion are the five various types of distortions which are obtained by adding Gaussian white noise to RGB components of the image Gaussian blur(GBLUR) distortion and Analog receiver signal to noise ratio fast fading channel distortion(FF).Under the same apparatus and monitoring environment the quality of every images judged by 29 different spectators.After rearrangement the different mean opinion score(DMOS) is obtained.The range of DMOS value of each image is 0-100,greater DMOS value represents the lower quality image.

VCL@FER database consists of 575 images, out of 575 images 23 images are originals, with 4 different distortion types per image and 6 degrees of distortion. JPEG2000, JPEG, Gaussian noise, White noise are types of distortions in the dataset. 118 subjects were used for testing who had to grade 11307 images. Every image was evaluated between 16 and 36 times. In each session every subject had to grade different types of distortion which will influence on objective measure performances.

For representing the relationship between objective and subjective values the spearman rank order correlation(SROCC) uses monotonicity and the range of values is [-1,1] which can be expressed as

$$\text{SROCC}=1-\frac{6\Sigma_{i=1}^{n} d_i^2}{N(N^2-1)}$$

In above formula d is the variation in the ranking of i'th image in the order of ground truth and predicted value.Another metric is Linear correlation coefficient(LCC)that takes the measurement of linear dependence between objective and subjective values and the range of values is[-1,1],which ca be expressed as

$$\text{LCC}= \frac{\Sigma_{i=1}^{N}(X_1-\bar{X})(Y_i-\bar{Y})}{\sqrt{\Sigma_{i=1}^{N}(X_i-\bar{X})}\sqrt{\Sigma_{i=1}^{N}(Y_i-\bar{Y})^2}}$$

In formula $X_1$ and $Y_1$ represent ground truth and predicted values of $i^{th}$ image , $\bar{X}$ and $\bar{Y}$ represent the average ground and average projected value of distorted image

*B. Experimental Results on LIVE Database*

The LIVE database includes 779 distorted images distributed across five types of distortions. For training the CNN models 70 percent images of each type of distortion have been selected randomly to minimize the error. After training the remaining images have been used for testing.

The obtained values of SROCC and LCC have been shown in the following tables. It can be observed that the proposed method provides better accuracy.

TABLE 2: COMPARISON OF SROCC FOR LIVE DATABASE

| Model | JP2K | JPEG | WN | BLUR | FF | ALL |
|---|---|---|---|---|---|---|
| BRISQUE | 0.914 | 0.965 | 0.979 | 0.951 | 0.877 | 0.940 |
| DIIVINE | 0.913 | 0.910 | 0.984 | 0.921 | 0.863 | 0.916 |
| CORNIA | 0.943 | 0.955 | 0.976 | 0.969 | 0.906 | 0.942 |
| BLIIDNS-II | 0.929 | 0.942 | 0.969 | 0.923 | 0.889 | 0.931 |
| SHALLOW CNN | 0.952 | 0.977 | 0.978 | 0.962 | 0.908 | 0.956 |
| Ensemble CNN (Proposed) | 0.985 | 0.988 | 0.998 | 0.992 | 0.984 | 0.989 |

TABLE 3: COMPARISON OF PLCC FOR LIVE DATABASE

| Model | JP2K | JPEG | WN | BLUR | FF | ALL |
|---|---|---|---|---|---|---|
| BRISQUE | 0.922 | 0.973 | 0.985 | 0.951 | 0.901 | 0.942 |
| DIIVINE | 0.922 | 0.921 | 0.988 | 0.923 | 0.888 | 0.917 |
| CORNIA | 0.951 | 0.965 | 0.987 | 0.968 | 0.917 | 0.935 |
| BLIIDNS-II | 0.935 | 0.968 | 0.980 | 0.938 | 0.896 | 0.930 |
| SHALLOW CNN | 0.953 | 0.981 | 0.984 | 0.953 | 0.933 | 0.953 |
| Ensemble CNN (Proposed) | 0.943 | 0.973 | 0.991 | 0.946 | 0.941 | 0.958 |

*C. Experimental Results on VCL@FER Database*

The *VCL@FER* database includes 575 images in total, out of these 23 images are originals and hence have been ignored while training and testing. From the 552 distorted images 480 images have been used for training the CNN models and the remaining 72 images have been used for testing.

The obtained values of SROCC have been shown in the following tables.

TABLE 4: COMPARISON OF SROCC FOR VCL@FER DATABASE

| Model | SROCC |
|---|---|
| BRISQUE | 0.812 |
| DIIVINE | 0.727 |
| BLIIDNS-II | 0.810 |
| Ensemble CNN (Proposed) | 0.988 |

## V. CONCLUSION

No reference image quality assessment using convolutional neural networks has been proposed in this paper. CNN avoids feature selection and dimensionality reduction because it automatically extracts features from the given data. Obtained experimental results show that CNN provides better accuracy than other existing models.

## REFERENCES

[1] A.Zaric, N.Tatalovic, N.Brajkovic, H.Hlevnjak, M.Loncaric, E.Dumic, S.Grgic, "VCL@FER Image Quality Assessment Database", AUTOMATIKA Vol. 53, No. 4, pp. 344–354, 2012

[2] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transactions on Image Processing, 2012, : pp. 4695-4708

[3] Moorthy A K , Bovik A C . Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality[J]. IEEE Transactions on Image Processing, 2012, pp. 3350-3364.

[4] Ye P, Kumar J, Kang L,. Unsupervised feature learning framework for no-reference image quality assessment[C]//In IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2012, pp.1098-1105

[5] Saad M A, Bovik A C, Charrier C.Blind image quality assessment:a natual scene statistics approach in the DCT domain[J]. IEEE Transactions on Image Processing,2012,vol 21,pp.3339-3352

[6] Kang L, Ye P, Li Y. Convolutional neural networks for no-reference image quality assessment[C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus : IEEE , 2014, pp. 1733-1740

[7] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek,"Deep neural networks for no-reference and full-reference image quality assessment," IEEE Trans. Image Process.2016, pp. 206–219