International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-7, Issue-6S4, April 2019

# No-reference Fundus Image Quality Assessment using Convolutional Neural Network

**Bhargav J Bhatkalkar, Dheeraj Rajaram Reddy, Vishnu Asutosh Dasu, Advait Raykar, Srikanth Prabhu, Sulatha Bhandary, Govardhan Hegde**

***ABSTRACT--- Computerized fundus image analysis is a well-established research area in the field of medical imaging. The cause of noise in fundus images is due to many factors like the low lighting conditions, adverse illumination effects, camera malfunctioning, etc. The presence of noise may lead to data loss and sometimes to the wrong data interpretation. Classifying the fundus images into either good quality or bad quality is very important as the good quality fundus images can be directly sent for processing without any preprocessing, hence reducing the computational time and the bad quality images can be forwarded for the required preprocessing stages. In this paper, we are using a convolutional neural network (CNN) to assess the quality of fundus images automatically. We use No-reference image quality assessment technique (IQA) classify the fundus images into good quality or bad quality based on their quality. A Mean Opinion Square (MOS) of 12 image quality assessment participants is taken for labeling the 300 fundus images based on their quality. The participants have rated the fundus images on the scale of 0-10, where the 0-rating is given for very bad quality fundus images, and 10-rating is given for the very good quality fundus images. The experimental study has proven that the classification result of the proposed CNN outperforms the best-known blind image quality assessment algorithms, namely, DIVINE, BLIINDS-II, and BRISQUE when trained on the public databases LIVE, TID2013 and on our fundus image dataset.***

***Keywords—Fundus image analysis, image quality assessment, convolutional neural network***

**Revised Manuscript Received on April 05, 2019.**

**Bhargav J Bhatkalkar,** Department of Computer science & Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: bhargav.jb@manipal.edu)

**Dheeraj Rajaram Reddy*,** Department of Computer science & Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: dheeraj.r@learner.manipal.edu)

**Vishnu Asutosh Dasu,** Department of Computer science & Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: vishnu.asutosh@learner.manipal.edu)

**Advait Raykar,** Department of Computer science & Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: advait.narendra@learner.manipal.edu)

**Srikanth Prabhu,** Department of Computer science & Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: srikanth.prabhu@manipal.edu)

**Sulatha Bhandary,** Department of Ophthalmology, Kasturba Medical College, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: sulatha_bhandary@yahoo.co.in)

**Govardhan Hegde,** Department of Computer science & Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, Udupi, India, 576104 (E-mail: Govardha.hegde@manipal.edu)

## I. INTRODUCTION

Medical imaging is used widely nowadays for diagnosing various diseases. Noise is a factor that is unavoidable in digital images due to many reasons. There can be many different types of noises introduced in digital images which makes it difficult to process them further. Noises are introduced due to improper lighting conditions, malfunctioning of digital sensors, slow shutter speeds, high sensitivity modes, etc. As the quality of medical image plays a very crucial role in further decision making, assessing its quality becomes the first step in image processing. By preprocessing only the poor quality images to enhance their quality reduces the great amount of computational cost and time.

Automated image quality assessment (IQA) techniques have the advantages of speed and accuracy in comparison to the manual image quality assessment. There are three categories of IQA methods, namely, 1) Full-reference Methods 2) Reduced-reference Methods 3) No-reference Methods. These three methods consider different image quality assessment metrics for their assessment.

The Full-reference IQA methods use the metrics like Feature Similarity Index Metrics (FSIM) [1], Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metrics (SSIM). The Reduced-reference IQA methods use the metrics like Peak Signal to Noise Ratio (PSNR), Modified Structural Similarity Index (MSSIM) and Lubin's Sarnoff model [2] [3], etc. The No-reference methods use most of the Full-reference metrics and Natural Scene Statistics (NSS) based features [4].

Machine learning techniques are extensively used for feature extraction and classification of the data. The Convolutional Neural Network (CNN) is a category of the algorithm in machine learning which is widely used for image segmentation and image classification since past recent years. These categories of algorithms are also known as Deep Learning algorithms.

CNN has several sets of convolutional layers. Between every set of the convolutional layer, there is a pooling layer as shown in Fig. 1. In between each set of the convolutional layer and in the pooling layer we have an activation function. The convolutional layer runs a filter on an input image and alters the individual pixel intensities. Activation function determines the state of neurons and fires the signal to the next connected neuron. Pooling layer is used to reduce the convolution responses to a lower dimension. This layer can be implemented with different types of pooling functions like Max pooling, Min pooling, Average Pooling, etc.

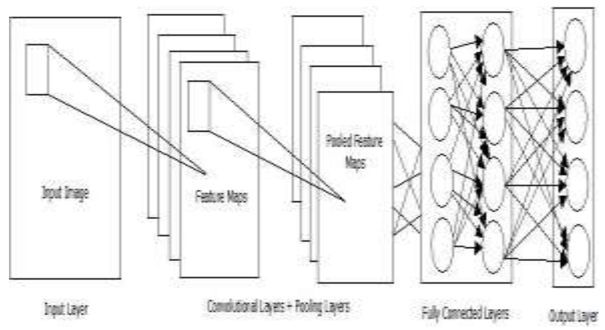


**Figure 1: A common architecture of CNN**

We have implemented a class of CNN known as VGGNet-16 [5] to classify the fundus images based on their quality. The aim is to classify the good quality fundus images from the bad quality fundus images. Only, the classified bad quality fundus images are taken up for the further preprocessing stages to enhance their quality. Great amount of computation time is saved as good quality fundus images are not considered for further preprocessing and can be directly used for feature extraction. The accuracy of classifying the good quality fundus images from the bad quality fundus images using VGGNet-16 is found to be exceptionally well when compared to DIVINE [6], BLIINDS-II [7] and BRISQUE [8] algorithms.

## II. RELATED WORK

Generally, different approaches to IQA are classified by the amount of information known about the original, undistorted reference image input to the algorithm. The full reference (FR) approaches to IQA require an undistorted reference image and have full access to the whole reference image. The reduced reference (RR) approaches require only the partial information, typically a set of extracted features of the distortion-free reference image. In no-reference IQA techniques, the specific information available about the reference image is not used, and it is assumed that a distortion-free reference image is not available for image assessment [9].

No-reference IQA algorithms fall under two categories: The distortion-specific algorithms and general-purpose algorithms. The former categories of algorithms know that there is a specific type of distortion in the images and the later does not know about the type of distortion available, and in this case, there may be more than one type of distortion available [9].

IQA algorithms may also consider the human visual system (HVS) evaluation methods for better accuracy [10] [11] [12] [13]. But since the lack of undistorted reference image, no-reference IQA methods may not precisely use HVS. The commonly used quality metrics for no-reference IQA are PSNR, SSIM, FSIM and Natural Scene Statistics (NSS) based features [14].

PSNR gives the ratio between the maximum intensity of the original image and the noise introduced. The higher PSNR value means the image quality is high and vice versa [15] [16]. The SSIM measure is more accurate compared to PSNR. It highly correlates with the quality human visual system. SSIM models any distortion in the image as by combining luminance distortion, loss of correlation, contrast distortion [11] [15]. The FISM algorithm considers the fact that HVS interprets the image by considering the low-level features. In this method, the low-level features of an image are extracted compared with the low-level features of the reference image. More difference between comparison indicates that the quality of the image is not good [17] [18].

NSS deliberates the fact that all the natural scenes hold some statistical properties. These properties are transformed by the distortion making the scenes un-natural. By using the scene statistics, NSS algorithms attempt to recreate the original scene without any reference [19]. NSS uses both wavelet transform and DCT transform [7] to extract the features.

Image quality assessment experiments require training data as well as testing data. Training data should have distortion specific images as well as images having more than one type of distortion. As the human visual system is considered to be the best IQA method, human subject opinion plays an important role in building the training data-set.

The LIVE database [20] comprises 779 labeled images. These images are formed using the original 29 source reference images, and every image is subjected to 5 different types of known distortions at different distortion levels. The best known full-reference image quality assessment database till date is TID2013 [22]. This database has 3000 test images derived from 25 reference color images.

There are various algorithms available for No-reference IQA like DIVINE, BLIINDS-II, and BRISQUE which were considered to be the best in the recent days. These algorithms were tested on the LIVE database for their performance. The latest introduction of Convolutional Neural Network for IQA has been proven to outperform all the previous No-reference IQA algorithms [23]

Overfitting is a problem in machine learning where the neural network starts to adapt to the data set given during the training. It mainly occurs due to the less amount of training data. This feature is called overfitting. Overfitting causes the network to classify the new test data wrongly. This problem of overfitting is easily eliminated using methods like “Dropout” [24].

There are various activation functions available for the neurons to fire. Most significant among all are Sigmoid, Tanh, Linear function, a Step function. Activation functions fall under two categories: Linear Activation Function and Non-Linear Activation Function. Rectified Linear Unit (ReLU) is one activation function which is most used in the world right now. Almost all CNN and Deep learning algorithms use ReLU for activating the neuron. It is a Non-Linear activation function, and it also reduces the computational cost compared to other activation function [25].


Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

## III. METHODOLOGY

We have built a data set of 300 fundus images for testing the efficiency of our CNN for classifying between good quality fundus images and bad quality fundus images. This data-set has various quality fundus images ranging from extremely good quality fundus images where we can see all the retinal structures to extremely bad quality fundus images where not even a single retinal structure is visible. There were 12 human participants to rate these fundus images in the range of 10 to 100 where 10 is the rating given to the extremely bad quality fundus images and 100 being the rating given to the extremely good quality fundus images. A Mean Opinion Square (MOS) of the rating is taken for every image, and a MOS fundus image dataset is created to train our CNN.

The Table 1 shows the different levels of implemented CNN architecture which is a 224 × 224 × 64, 112 × 112 × 128, 56 × 56 × 56 × 256, 28 × 28 × 28 × 512, 14 × 14 × 14 × 512, 1 × 1 × 4096, 1 × 1 × 4096, 1 × 1 × 4096 structure. The input is fixed-size 224 × 224 locally normalized RGB image. The convolution in all the layers is performed using 3 × 3 filters, and the convolution stride is fixed to 1. In some of the convolution layers, we also use a 1 × 1 convolution filter to carry the effect of a linear transformation of the input channels. Max-pooling in between the convolution layers is performed over a 2 × 2 pixel window, with a stride of 2.

Every convolutional layer in the architecture has a different depth. These convolutional layers are followed by the three Fully-Connected (FC) layers at the end as shown in Fig 2. The first two fully-connected layers have 4096 channels each and the third layer is an output layer which performs a 2-way classification and thus contains 2 channels (one for each class, i.e., good-quality fundus image or bad-quality fundus image). The final layer is the soft-max layer [5].

The CNN is trained in an iterative process over a number of epochs. In deep learning, one epoch is the time duration in which every sample from the training set has been used once. We have used 100 epochs with early stopping. The optimizer used for CNN is Adam, and no dropout layers are used in the architecture. The network learning rate is set to 1e-4 with exponential decay.

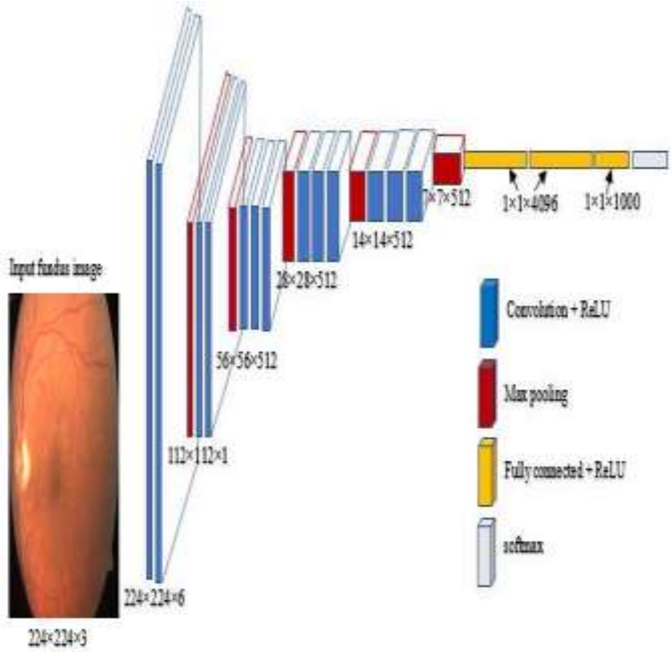


**Figure 2: Different layers of the implemented CNN**

**Table 1: Implemented CNN architecture [5]**

| Input (224 × 224 RGB image) | | | | | |
|---|---|---|---|---|---|
| **conv3-64** | conv3-64<br>LRN | conv3-64<br>conv3-64 | conv3-64<br>conv3-64 | conv3-64<br>conv3-64 | conv3-64<br>conv3-64 |
| **max pool** | | | | | |
| **conv3-128** | conv3-128 | conv3-128<br>conv3-128 | conv3-128<br>conv3-128 | conv3-128<br>conv3-128 | conv3-128<br>conv3-128 |
| **max pool** | | | | | |
| **conv3-256**<br>**conv3-256** | conv3-256<br>conv3-256 | conv3-256<br>conv3-256 | conv3-256<br>conv3-256<br>conv1-256 | conv3-256<br>conv3-256<br>conv3-256 | conv3-256<br>conv3-256<br>conv3-256<br>conv3-256 |
| **max pool** | | | | | |
| **conv3-512**<br>**conv3-512** | conv3-512<br>conv3-512 | conv3-512<br>conv3-512 | conv3-512<br>conv3-512<br>conv1-512 | conv3-512<br>conv3-512<br>conv3-512 | conv3-512<br>conv3-512<br>conv3-512<br>conv3-512 |
| **max pool** | | | | | |
| **conv3-512**<br>**conv3-512** | conv3-512<br>conv3-512 | conv3-512<br>conv3-512 | conv3-512<br>conv3-512<br>conv1-512 | conv3-512<br>conv3-512<br>conv3-512 | conv3-512<br>conv3-512<br>conv3-512<br>conv3-512 |
| **FC-4096** | | | | | |
| **FC-4096** | | | | | |
| **FC-1000** | | | | | |
| **Soft-max** | | | | | |


Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

## IV. RESULTS AND CONCLUSION

The performance of the proposed algorithm is evaluated based on the LIVE and TID2008 datasets. The Spearman's rank correlation coefficient (SRCC) and Pearson correlation coefficient (PCC) accuracy matrices are used for the performance evaluation of the implemented CNN with the DIVINE, BLIINDS-II and BRISQUE algorithms. The following Table 2 lists the accuracy of CNN over the other algorithms on different databases. It evident that our CNN outperforms during the classification of fundus images based on their quality compared to the other three image quality assessment algorithms.

**Table 2: Accuracy of CNN in fundus image classification**

| | LIVE | | TID2013 | | Fundus Dataset | |
|---|---|---|---|---|---|---|
| | PLCC | SRCC | PLCC | SRCC | PLCC | SRCC |
| CNN (Ours) | 0.97 | 0.956 | 0.989 | 0.933 | 0.81 | 0.821 |
| DIVINE | 0.885 | 0.916 | 0.851 | 0.855 | 0.783 | 0.712 |
| BLIINDS-II | 0.93 | 0.931 | 0.841 | 0.877 | 0.733 | 0.74 |
| BRISQUE | 0.94 | 0.94 | 0.917 | 0.922 | 0.785 | 0.798 |

The following Fig 3 shows that our CNN performs extremely well during convergence in adjusting the weights of the network.

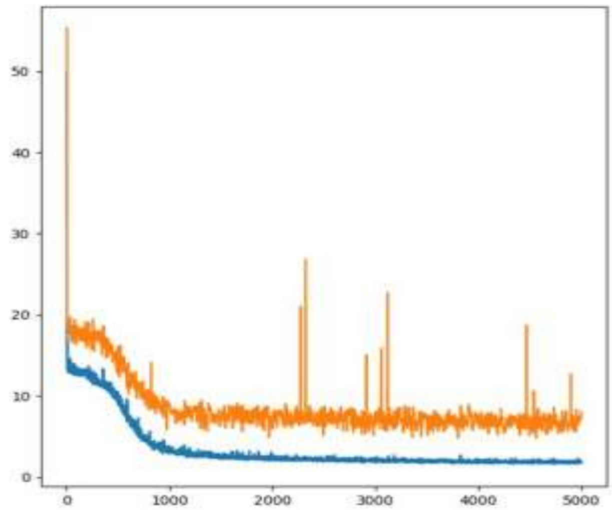


**Figure 3: The CNN Convergence Graph**
**(Blue line - training loss Orange line - Validation loss)**

Following Fig 4 and Fig 5 shows the fundus images being classified into good quality images and bad quality images by our CNN.

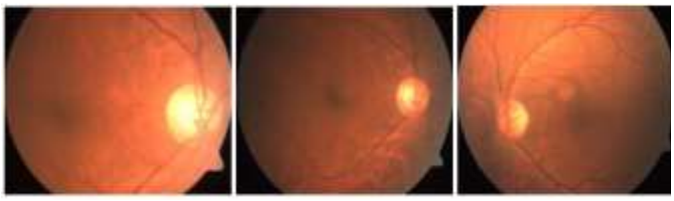

**Figure 4: Good Quality Fundus Images**

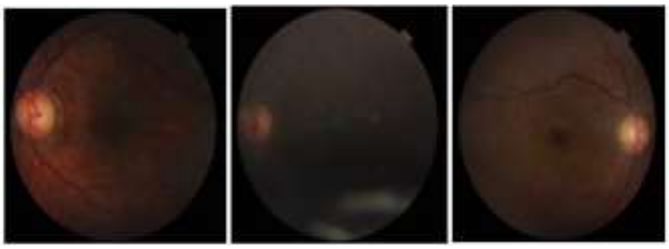

**Figure 5: Bad Quality Fundus Images**

## V. ACKNOWLEDGMENT

We sincerely thank Kasturba Medical College (KMC), Manipal, India for providing the required fundus images for our experimentation. We also thank Manipal Academy of Higher Education (MAHE) for the support and the resources provided during our research work.

## REFERENCES

[1] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "FSIM: a feature similarity index for image quality assessment", IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, 2011.
[2] J. Lubin, “The use of psychophysical data and models in the analysis of display system performance,” in Digital images and human vision, A. B. Watson, ed., pp. 163–178, The MIT Press, Cambridge, Massachusetts, 1993.
[3] J. Lubin, “A visual discrimination mode for image system design and evaluation,” in Visual Models for Target Detection and Recognition, E. Peli, ed., pp. 207–220, World Scientific Publishers, Singapore, 1995.
[4] Yan Fu and Shengchun Wang, A No Reference Image Quality Assessment Metric Based on Visual Perception.
[5] K.Simonyan and A. Zimmerman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556v6 [cs], Apr. 2015.
[6] Moorthy, A. K., and Bovik, A. C. 2011. Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE transactions on Image Processing 20(12):3350–3364.
[7] Saad, M. A.; Bovik, A. C.; and Charrier, C. 2012. Blind image quality assessment: A natural scene statistics approach in the DCT domain. IEEE Transactions on Image Processing 21(8):3339–3352.
[8] Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21(12):4695–4708.
[9] A No Reference Image Quality Assessment Metric Based on Visual Perception, Yan Fu and Shengchun Wang.
[10] Liu, H.; Heynderickx, I. Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data. IEEE Trans. Circuits Syst. Video Technol. 2011, 21, 971–982.
[11] Wang, B.; Wang, Z.; Liao, Y.; Lin, X. HVS-based structural similarity for image quality assessment. In Proceedings of the International Conference on Signal Processing, Beijing, China, 26–29 October 2008.
[12] Babu, R.V.; Perkis, A. An HVS-based no-reference perceptual quality assessment of JPEG coded images using neural networks. IEEE Int. Conf. Image Process. 2005, 1, I-433-6.
[13] Zhang, Y.D.;Wu, L.N.; Wang, S.H.; Wei, G. Color image enhancement based on HVS and PCNN. Sci. Chin. Inf. Sci. 2010, 53, 1963–1976.
[14] No-reference/Blind Image Quality Assessment: A Survey ,Shaoping Xu,Shunliang Jiang &Weidong Min Pages 223-245 | Published online: 08 Apr 2016.

[15] Image Quality Metrics: PSNR vs. SSIM, Alain Hore, Djemel Ziou, 2010 20th International Conference on Pattern Recognition,2010

[16] Scope of validity of PSNR in image/video quality assessment, Q. Huynh-Thu 1 and M. Ghanbari 2

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[18] FSIM: A Feature Similarity Index for Image Quality Assessment Lin Zhanga, Student Member, IEEE, Lei Zhanga,1, Member, IEEE Xuanqin Moub, Member, IEEE, and David Zhanga, Fellow, IEEE

[19] Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality, Anush Krishna Moorthy, Alan Conrad Bovik IEEE Transactions on Image Processing ( Volume: 20, Issue: 12, Dec. 2011 )

[20] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. LIVE image quality assessment database release 2. Online, ttp://live.ece.utexas.edu/research/quality

[21] Image database TID2013: Peculiarities, results and perspectives, Pages 57-77, Volume 30, Signal Processing: Image Communication, 1 Jan 2015.

[22] Anush Krishna Moorthy ; Alan Conrad Bovik, Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality, IEEE Transactions on Image Processing ( Volume: 20 , Issue: 12 , Dec. 2011

[23] Convolutional Neural Networks for No-Reference Image Quality Assessment Le Kang, Peng Ye, Yi Li, David Doermann; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1733-1740

[24] Droup out: G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.

[25] https://towardsdatascience.com/activation-functions-and-its-types-which-is-better-a9a5310cc8f