International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Blind Synthetic Image Quality Assessment Using EfficientNet-V2

Yogita Gabhane[1*], Tapan Kumar Jain[1†] and Vipin Kamble[2†]

Submitted: 18/10/2023 **Revised**: 10/12/2023 **Accepted**: 21/12/2023

Abstract: We propose a simple and efficient deep convolutional neural network (CNN) model using EfficientNet-V2 S for Blind Image Quality Assessment (BIQA) that works for the distorted images from the TID2013 dataset consisting of images subjected to different varieties and levels of distortion. The ability of the EfficientNet-V2 S over its base model is effectively used for faster training and lower computational and time complexity. For quality assessment, the image is resized and normalized in such a manner as to retain the information content. The pre-trained model is used to evaluate weights using transfer learning over 80% of the TID2013 datasets while the model is tested on the remaining 20% of the distorted images. The actual scores and the predicted scores are compared based on two correlation coefficients namely Spearman's Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC) with state-of-the-art techniques used for BIQA. We fine-tuned the model by modifying the structure of the network especially the last layers for the target dataset and achieved remarkable performance for sixteen cases from 24 varied distortion cases.

Keywords*: Deep convolutional neural network, EfficientNet-V2 S, Blind Image Quality Assessment, TID2013 dataset, transfer learning, SROCC, and PLCC.*

1. Introduction

Modern image acquisition systems are highly enhanced to acquire quality images whether they are stationary or mobile however problem persists in storing, compressing, and transmitting the images. Perceptual loss and unwanted distortions are introduced in any of the stages of the image acquisition system. Therefore, image quality assessment (IQA) becomes an increasingly important aspect to assure the reliability of the system. Perceptual assessment is the matter better concerned with the human visual system, subjective IQA is complex and expensive being the most reliable. The foremost task is to design an efficient and accurate IQA to evaluate the quality of a distorted image.

Depending upon the reference information, traditionally objective IQA are categorized as no-reference, reduced, and full-reference IQA. Different from the NR-IQA algorithms, which can only work with a distorted digital image, the FR-IQA [1-9] methods can use the full reference image in their analysis. To evaluate an image's quality, reduced-reference (RR)-IQA algorithms only have access to a subset of the whole reference image and therefore must rely on a pool of extracted features instead. IQA computation techniques that seem to be objective are evaluated using benchmark databases that contain images that have been purposefully altered, along with the correlating mean opinion scores (MOSs) that were acquired through subjective user testing. Such MOSs are employed to compare the accuracy of various objective IQA techniques.To reiterate, researchers can use freely accessible IQA databases to create and analyze IQA algorithmic structures. There exists an abundance of algorithms for No-Reference IQA (NR-IQA) that have been devised over the years, they can be broadly categorized as either distortion-specific or universal. Such algorithms can determine an image's quality by identifying and analyzing one or more distortions present, including but not limited to blur, blocking, ringing, and noise.

In most of the real-world cases, reference information does not exist or is unavailable, in such cases, BIQA has attracted significant researchers in this area. Earlier research carried out using the BIQA approach included traditional handcrafted features or the learned features which covered the low and high-level features from the image. Nowadays, semantic deep blind features are extracted using deep convolutional neural networks. A MEON model for quality prediction was suggested in [10]. It was based on multitask learning for distortion detection and regression. The features were extracted using the CNN while the distortion detection network was learned on image descriptors. Lastly, the output of network and quality features were collectively used to predict the visual quality of the image. As an alternative to handcrafted features of a single scale, local features were extracted from the distorted image [11]. CNN was

[1*]*Department of ECE, IIIT, Nagpur, 441108, Maharashtra, India.*
[1]*Department of ECE, IIIT, Nagpur, 441108, Maharashtra, India.*
[2]*Department of ECE, Visvesvaraya National Institute of Technology, Nagpur, 440010, Maharashtra, India.*
[*]*Corresponding author(s). E-mail(s): yogitagabhane.yg@gmail.com; Contributing authors: tapan.jain@iiitn.ac.in; vipinkamble@ece.vnit.ac.in;*

used with local weights and local quality jointly as a weight-based quality metric. This improved the overall performance which considered features on multiscale. HyperNet [12] was able to extract deep local and global features at different levels. Integrating features at different levels is an issue and requires extensive research and collaborative discussion. Although much research can be seen in deep learned IQA techniques and has made remarkable contributions in synthetic and authentic distortions (SAD), still there is a large scope for improvement.

The paper focused on the following aspects:

1. An efficient preprocessing stage to resize and normalize the images while retaining the details of the image and minimizing the computational complexity through experimental analysis.

2. The use of EfficientNet V2-S [13] pre-trained CNN model as a backbone for extracting deep semantic features from the image.

3. Fine tuning of model parameters to achieve improved quality prediction.

4. The correlation coefficient values SROCC and PLCC are preferable to the competing methodologies on the TID2013 [14] data set.

Further, this paper has been organized as follows: In the next section, we express a brief review of the recent state-of-the-art research relevant to the topic discussed in this paper. In the third section, the suggested methods for NR-IQA based on deep neural networks (DNN) are explained. In these methods, the source image is pre-processed before being sent to the DNN model for feature extraction and quality prediction. In the fourth section, experimental work is explored. In the fifth section, the outcomes are discussed, recent approaches are compared, and state-of-the-art quality predictions. Further, in the last part, conclusions are drawn.

2. Literature Review

The authors in [15] modelled SAD as independent variations and further pooled bilinearly the two pre-trained descriptors as a collective representation resulting in a deep bilinear network for quality predictions called DB-CNN. They evaluated their model on five IQA datasets and represented the performance of DB-CNN for both distortions. For synthetic distortions, they used 21,869 images from two well-known datasets including the Waterloo Exploration [16] and the PASCAL VOC 2012 [17] datasets, and introduced an additional five distortions along with the four common distortions excluding distortions caused due to under-and over-exposures. To pre-train the network, 852,891 distorted images were used. Whereas, VGG-16 [18] was used to extract the relevant features in the authentic distortion stage considering the hypothesis that the network adapts to authentic distortions as a natural consequence of the ImageNet [19] dataset to improve the classification accuracy. They limited their model to a few distortions and the VGG-16 network while other distortions and networks such as ResNet [20] could be used considering other bilinear pooling variants. Also, the proposed DB-CNN model can be extended for feature extraction for a more unified quality assessment.

Multilevel deep semantic features using deep learning were acquired from a strong vision-transformer structure [21]. The authors focused on extracting features at different levels instead of traditional handcrafted features using a fusion module for effective hierarchical features. The fusion module was followed by an attention module for eliminating the redundant features and successively different granularity distortions were represented using the low and high level features. The resultant image quality score (IQS) was represented as a map using the local and global features. The features were obtained using local average and max pooling for the local features whereas global features were gained using the global average and max pooling after the convolutional layer (CNL). They used CSIQ [22] and TID2013 [14] datasets for the synthetic distortions while BID [23], LIVEC [24], and the KONIQ-10k [25] were part of authentic images. Although their proposed framework achieved good performance over 18 other state-of-the-art works, they failed to detect the type and area of the distortion. Their model suffered from time complexity.

The problem of overfitting due to scarce dataset images was tackled in [26]. The images were subjected to different levels and types of distortion and various strategies were adopted for augmenting the images. A No-reference IQA technique was used for training and Full-reference IQA was used for measuring distorted image score. The augmented images were trained using the Resnet-50 Network to obtain more robust weights and the weights were used for predicting the final image score on the target IQA images from the fine-tuning stage. The performance of the NR-IQA framework based on quality-aware features was carried out on authentic and synthetic distortion images from LIVE, CSIQ, TID2013, LIVEC, and KonIQ-10K datasets. They limited their work to some distortions while augmenting the dataset images. Further, both the authentic and the synthetic datasets were handled independently and not in a unified manner which would improve the generalization capability of their framework. More deep semantic features can be used instead of just considering the output from the last CNL.

Resnet-50 was the backbone of semantically oriented and perceptual quality-oriented features in [27]. One of the parallel models is fixed and the other is learnable with the removed last CNL and the average pooling layer. A global average pooling was used to flatten the features, and 3-fully connected layers (FCL) with ReLu and sigmoid were used for quality regression. Cross-domain features were extracted using a 3-stacked CNL, a GAP layer, and 3-FCL. Experimental analyses were carried out on KonIQ-10k, LIVEC datasets, and the TID2013 to validate the performance of the cross-domain feature similarity-guided network. The proposed framework indicated higher generalization ability concerning the other four competing IQA techniques.

The work proposed in [28] focused on monitoring the objects of interest during BQIA (Blind Quality Image Assessment). They used an end-to-end detection transformer cum IQA model (DETR-IQA) by adding simple multilayer perceptrons at the decoder. Using the KonIQ-10k dataset, they considered five objects to detect in the images and used a fusion of distortion degree comprising of the region of interest (ROI belonging to object region) and other remaining part of the image. They found that the predicted IQS of ROI was higher than the score of the remaining region. The input image was fed to a feature extractor module like ResNet-50 to acquire multi-scale features and the flattened sequence was then provided to the transformer encoder. The decoder transformer containing cross and self-attention modules processes 100 object queries along with the encoder output. The feed-forward network governs the bounding boxes while the classification is obtained through the linear projections. Their proposed framework can assess image quality with and without objects with a small amount of computational complexity.

The authors in [29] concentrated on distortion diversity and image variations. They suggested a model for improving the BIQA adaptability schemes for distinguished image contents and a variety of distortions. They introduced two content-aware networks (ResNet-50 and ResNet-18) for capturing the quality characteristics of the image from different perspectives for QA. The HyperIQA network was evaluated on authentic and synthetic image datasets which possessed remarkable generalization ability still leaving room for effectively handling diverse image contents and distortions. The research by [30] suggested distortion awareness, distortion fusion, and quality prediction modules. It captures synthetic and real distortions in distorted images using synthetic and real distortion-aware networks, but for TID2013 distortion-wise correlation is not considered. Although each patch's quality differs according to its content and actual spatial differences, allocating the MOS to the patches that are sampled from the respective original image remains a common practice.

For BIQA, work suggested in [31] the dual-branch vision transformer (ViT). It concurrently takes into account both global semantic information and multiscale local distortions. Pre-trained CNN ResNet-50 is used to obtain two scale features which are input to the ViT with two branches, and then to attain global image semantics, content-aware-IQA is employed. Finally, to accurately anticipate the IQS, numerous FFBs integrate the results of the vision transformer which is a dual branch. The researchers conducted a Neural Architecture Search (NAS) to develop a novel network. This network was then scaled up to create a range of models known as EfficientNet. The results of this study demonstrate that the EfficientNet achieves greater precision and efficiency than any prior neural network design. Unlike, prior works this model uses EfficientNetV2-S for feature extraction to give a better IQA.

3. Methods and Materials

The experiments are performed on the synthetic IQA datasets provided by TID2013. The TID2013 dataset consists of a collection of 3000 images that have been proned to various forms of distortion. These distortions are sourced from twenty-five references and consist of a total of twenty-four different distortions with five degradation levels. The various types of distortion involve a broad spectrum, ranging from additive Gaussian noise, and compression distortions like JPEG, and JPEG2000 to more unconventional forms of distortion such as local block-wise noise, and non-eccentricity pattern noise. The TID2013 database can be considered a more demanding IQA dataset. The TID2013 database is collection of images, each of which is assigned a mean opinion score (MOS) within the numerical range of [0, 9]. In contrast to the DMOS, higher MOS indicates higher quality. Within this database, we select 80% of the images that have been distorted at random for training purposes, and 20% to test our system. These train and test sets are completely distinct in terms of the contents of the images they contain.

The proposed work uses a pre-trained EfficientnetV2-S network for feature extraction of the input image which is faster as compared to its base model EfficientNet when training is considered. To mitigate the computational overhead and time complexity, the input image from the TID2013 dataset is resized to 384x384 from 512x312. The resize operation reduces the input image size in only one dimension which ensures minimum information loss. The pixel values are then normalized in the range [0 1] for fast processing and eliminate overfitting using the preprocessing block shown in Figure 1.

The final layers of EfficientnetV2-S are removed and the vector from the last CNL as depicted in Figure 1 is flattened. The flattened output is passed through a layer consisting of 1024 neurons which serves as a bottleneck to substantially decrease the dimension of the features while preserving the information content. Lastly, the 1024 features of the image are processed by a dense layer holding a solitary neuron to create a conclusive prediction. The framework of the model that has been proposed is illustrated in Figure 1.

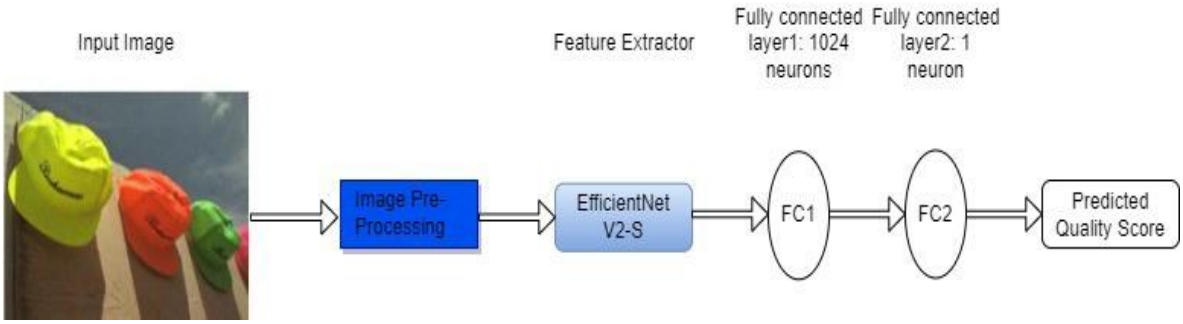

Fig. 1: The proposed model for BIQA

The algorithm for the proposed BIQA system is presented below:

Algorithm 1: BIQA System

Input: Image from the TID2013 dataset

Output: Quality predicted metric

1. Resize the input image [384x384]

2. Normalize the pixel values in the range [0 1]

3. **Model Training:** Suppose for Initial image size I_0 and regularization R_0^k and resultant image size I_e and regularization R_e^k. Epoch T and S stages. Progressive learning with adaptive regularization [32] is applied.

Input to the pre-trained CNN:
Several steps in training T and levels S.
for $i = 0$ to $S - 1$:
Image size: $I_i \leftarrow I_0 + (I_e - I_0) \cdot \frac{i}{S-1}$
Regularization: $R_i \leftarrow \left\{R_k^i = R_k^0 + (R_k^e - R_k^0) \cdot \frac{i}{S-1}\right\}$
The model is trained for $\frac{T}{S}$ steps with I_i and R_i.
end for

4. **Dense Layer Setup:** 1024 neurons and the final neuron gives a quality prediction.

The Feature Extraction-Prediction Set Up

The proposed method makes use of the pre-trained CNN EfficientnetV2- S [14] model for feature extraction trained on Imagenet-1k (ILSVRC-2012-CLS) and Imagenet-21k databases [33], which is the least complex model and has equivalent precision in making forecasts to that of another variation. EfficientNet systematically studies model scaling in 3 ways i.e. width, the number of layers, or depth and resolution of input to learn fine grain features of the input. To balance the precision of prediction and the complexity of networks, regression is performed using the EfficientNetV2-S structure as a feature extractor. The pre-trained weights of the model can be used to extract high-quality features from images, which can subsequently be used as classifier inputs.

Adjustable parameters include convolution kernel size, filter count, and MBConv module number, which vary across these eight variations. The architectural difference between EfficientNetV2 & EfficientNetV2-S is shown in Table 1.

Table 1. Architecture of EfficientNetV2 & EfficientNetV2-S (**These layers (MBConv with Fused-MBConv) were modified in EfficientNetV2 & EfficientNetV2-S architectures)*

Stage	Operator	Stride	#Channels-V2	#Channels-V2-S	#Layers
0	Conv3x3	2	24	24	1
1	*Fused-MBConv1, k3x3	1	24	24	2
2	*Fused-MBConv4, k3x3	2	48	48	4
3	*Fused-MBConv4, k3x3	2	64	64	4
4	MBConv4, k3x3, SE0.25	2	128	128	6
5	MBConv6, k3x3, SE0.25	1	160	160	9
6	MBConv6, k3x3, SE0.25	2	272	*256*	15
7	Conv1x1 & Pooling & FC	-	1792	*1280*	1

The TID2013 dataset's distorted images are used for experimentation first by splitting the "train" and "target" arrays with the train-test-split function (sklearn library), that is used to split arrays into train and test sets. In this instance, the data is divided into 80:20 training: evaluation sets. The split will always be the same because the random state option is set to 42. Then the training data is trained on the Efficient-Net V2 S architecture. The model has already been taught to identify salient details in pictures. It takes an image as input with shape (384, 384, 3) and returns a feature vector representation of the image. The feature extractor can be used as a part of a larger model, allowing the use of the pre-trained weights of the EfficientNet-V2 model and fine-tuning them on a new dataset. In this case, setting the trainable attribute of a layer to False prevents the layer's weights from being altered during training; this is helpful when employing a pre-trained model as a feature map without updating the learned features. By keeping the pre-trained model's weights fixed, we can reduce the likelihood of overfitting, as the model will be able to devote its full attention to mastering the new task rather than trying to fit itself into the new data. Because of this, the pre-trained model may be used as a feature extractor with a fixed set of features, and only the top-level layers of the model need to be trained. The first dense layer, the FC layer with 1024 neurons, is a bottleneck layer that reduces the feature vector output by the feature extractor layer. This layer will learn to incorporate pre-trained model characteristics for the current task. The count of parameters in this stage is equal to the sum of the 1280 neurons in the previous layer plus any biases added (1024). So in this case, the number of parameters is (1024 * 1280) + 1024 = 1311744. The dense layer is fully connected to a neural network. The input to the layer is a 2D tensor with shape (None, 1280), where the first dimension is the batch size, and the feature count constitutes the second dimension (1280). The output of the layer is another 2D tensor with shape (None, 1024), for which the second dimension represents the total neurons in this layer. The second dense layer has the same properties, but the number of neurons is only 1. This is called the output layer, and it is used for the final prediction.

For the model's compilation, we employ Adam optimization. The loss function governs the model's efficiency for training. For regression issues, the usual loss function is a mean squared error which calculates the expected-actual average squared difference. The model learns with 200 epochs with single image as the batch, so the model updates the weights after every training data sample. Table 2 below shows the parameters used for the proposed model.

Table 2. Parameters and respective values for the BIQA transfer learning with EfficientNet

Parameter	Value
Input Image size	384x384x3
Input Normalization range	[0 1]
Epoch for training	200
Batch size	1
Loss function	MSE (Mean Squared Error)
Optimizer	Adam
Train: Test Ratio	80: 20%

4. Results and Discussion

The EfficientVNet2-S model is trained on this 80% training data which has the right combination of MBconv and fused MBconvnet in its architecture improving the training speed. In EfficientVNet2-S progressive learning with adaptive regularization is used where the network is trained with images of small size and low regularization factor so that the model can understand basic representations quite quickly with little effort. Further, higher regularization factors are adopted by gradually increasing the image dimension. So, this regularization is adjusted adaptively, leading to good values of correlation coefficients.

For the task of determining the quality of the IQA measures, SROCC and PLCC [34] are utilized. The performance of NR-IQA algorithms can be evaluated for their effectiveness on the measured degree to which the ground mean opinion scores closely match the scores predicted in a standard comparison database.

SROCC is defined using expression (1), Considering P and Q as two datasets

$$\text{SROCC} = \frac{\sum_{i=1}^{m}(P_i-\hat{P})(Q_i-\hat{Q})}{\sqrt{\sum_{i=1}^{m}(P_i-\hat{P})^2}\sqrt{\sum_{i=1}^{m}(Q_i-\hat{Q})^2}} \quad (1)$$

Where $\hat{P}$ and $\hat{Q}$ represents the middle ranks. Now, PLCC is defined by expression (2),

$$\text{PLCC} = \frac{\sum_{i=1}^{m}(P_i-\bar{P})(Q_i-\bar{Q})}{\sqrt{\sum_{i=1}^{m}(P_i-\bar{P})^2}\sqrt{\sum_{i=1}^{m}(Q_i-\bar{Q})^2}} \quad (2)$$

Where $\bar{P}$ and $\bar{Q}$ represent the average of P and Q and P_i and Q_i represent the i^{th} elements of P and Q respectively.

Table 3 presents the distortion-wise SROCC and PLCC scores of various distortions in the TID2013 dataset. These scores indicate the correlation between the subjective IQSs assigned by human observers and the objective quality measures computed for each distortion type. Higher correlation coefficients indicate a stronger relationship between subjective IQSs and objective IQSs, suggesting that the objective measures are able to evaluate the perceptual quality effectively of the distorted images.

Table 3. Distortion-wise SROCC and PLCC score of TID2013

Type of Distortion	SROCC	PLCC	Type of Distortion	SROCC	PLCC
Masked noise	0.939	0.969	Local block-wise distortions	0.445	0.416
Additive noise in color components	0.939	0.964	Non-eccentricity pattern noise	0.609	0.849
Additive Gaussian noise	0.928	0.947	Mean shift	0.567	0.681
Spatially correlated noise	0.917	0.960	Change of color saturation	0.884	0.899
High-frequency noise	0.978	0.980	Contrast change	0.923	0.975
Impulse noise	0.972	0.968	Comfort noise	0.934	0.988
Quantization noise	0.917	0.905	Lossy compression of noisy images	0.846	0.931
Gaussian blur	0.967	0.970	Multiplicative Gaussian noise	0.967	0.989
Image denoising	0.928	0.975	Chromatic aberrations	0.890	0.981
JPEG compression	0.824	0.961	Color quantization with dither	0.851	0.938
JPEG2000 Compression	0.961	0.978	Sparse sampling and reconstruction	0.989	0.967
JPEG2000 transmission errors	0.895	0.927	All	0.938	0.945
JPEG transmission errors	0.923	0.952			

To delve deeper into the nuances of our suggested method, we present the results of various kinds of distortion and compare them to other NR-IQA models. Table 4 displays the outcomes of tests conducted on the TID2013 database. It clearly shows that our technique is the most accurate across 16 distortions (66% subsets) out of 24 different kinds of distortions. In contrast, for certain kinds of distortions, like Mean Shift and Local Block-Wise Distortions, smaller SROCC values are obtained. It is noteworthy that our methodology demonstrates substantial enhancements in performance for certain distortion types that are relevant to noise e.g., Sparse Sampling and Reconstruction, High-frequency noise, gaussian blur) and compression-related distortion types (e.g. JPEG 2000 compression, JPEG transmission errors). The SROCC and PLCC scores provide an understanding of the performance achieved in objective quality measures for individual types of image distortions in the TID2013 dataset.

The plot of the actual image scores and the predicted score is shown in Figure 2. It is observed that the mean value of actual values coincides with the mean value of predicted metrics. The proposed BIQA scheme using the EfficientNet-V2 S fails to follow the actual metrics when the quality metrics are higher. For low-quality metric scores, the predicted scores nearly approach the actual scores. Figure 3 shows some of the images from the TID2013 Dataset with the actual score and the predicted score using the proposed BIQA scheme. It can be observed that the prediction accuracy is a function of image details. The image with more objects shows a large deviation in the decimal part of the predicted score while images with lower objects are predicted with minimum loss. Also, the system is affected by color variations in the input image. More color changes deviate the predicted value from the actual score.

Table 4: Comparison: Distortion-wise SROCC of methods so far using TID2013 with proposed BIQA Scheme

SROCC	CORNIA	FRIQUEE	M3	BRISQUE	MEON	DBCNN	HOSA	FPR	S. Lee	Proposed
Additive Gaussian noise	0.692	0.730	0.766	0.711	0.813	0.790	0.833	**0.953**	0.857	0.928
Additive noise in color component	0.137	0.573	0.56	0.432	0.722	0.700	0.551	0.897	0.661	**0.939**
Spatially correlated noise	0.741	0.866	0.782	0.746	0.926	0.826	0.842	**0.967**	0.481	0.917
Masked noise	0.451	0.345	0.577	0.252	0.728	0.646	0.468	0.876	0.858	**0.939**
High-frequency noise	0.815	0.847	0.900	0.842	0.911	0.879	0.897	0.934	0.886	**0.978**
Impulse noise	0.616	0.730	0.738	0.765	0.901	0.708	0.809	0.779	0.892	**0.972**
Quantization noise	0.661	0.764	0.832	0.662	0.888	0.825	0.815	**0.920**	0.869	0.917
Gaussian blur	0.850	0.881	0.896	0.871	0.887	0.859	0.883	0.833	0.869	**0.967**
Image denoising	0.764	0.839	0.709	0.612	0.797	0.865	0.854	**0.944**	0.871	0.928
JPEG compression	0.797	0.813	0.844	0.764	0.850	0.894	0.891	**0.923**	0.872	0.824
JPEG2000 Compression	0.846	0.831	0.885	0.745	0.891	0.916	0.919	0.923	0.917	**0.961**
JPEG transmission errors	0.694	0.498	0.375	0.301	0.7463	0.772	0.73	0.797	0.834	**0.923**
JPEG2000 transmission errors	0.686	0.660	0.718	0.748	0.716	0.773	0.710	0.752	**0.907**	0.895
Non-eccentricity pattern noise	0.200	0.076	0.173	0.269	0.116	0.270	0.242	0.559	0.564	**0.609**
Local block-wise distortions	0.027	0.032	0.379	0.207	0.500	0.444	0.268	0.265	0.690	**0.445**
Mean shift	0.232	0.254	0.119	0.219	0.177	-0.009	0.211	0.009	0.519	**0.567**
Contrast change	0.254	0.585	0.155	-0.001	0.252	0.548	0.362	0.699	0.780	**0.923**
Change of color saturation	0.169	0.589	-0.199	0.003	0.684	0.631	0.045	0.409	0.753	**0.884**
Multiplicative Gaussian noise	0.593	0.704	0.738	0.717	0.849	0.711	0.768	0.887	0.844	**0.967**
Comfort noise	0.617	0.318	0.353	0.196	0.406	0.752	0.622	0.830	0.838	**0.934**
Lossy compression of noisy images	0.712	0.641	0.692	0.609	0.772	0.860	0.838	**0.982**	0.932	0.846
Color quantization with dither	0.683	0.768	**0.908**	0.831	0.857	0.833	0.896	0.901	0.794	0.851
Chromatic aberrations	0.696	0.737	0.570	0.615	0.779	0.732	0.753	0.768	0.812	**0.890**
Sparse sampling and reconstruction	0.865	0.891	0.893	0.807	0.855	0.902	0.909	0.887	0.926	**0.989**

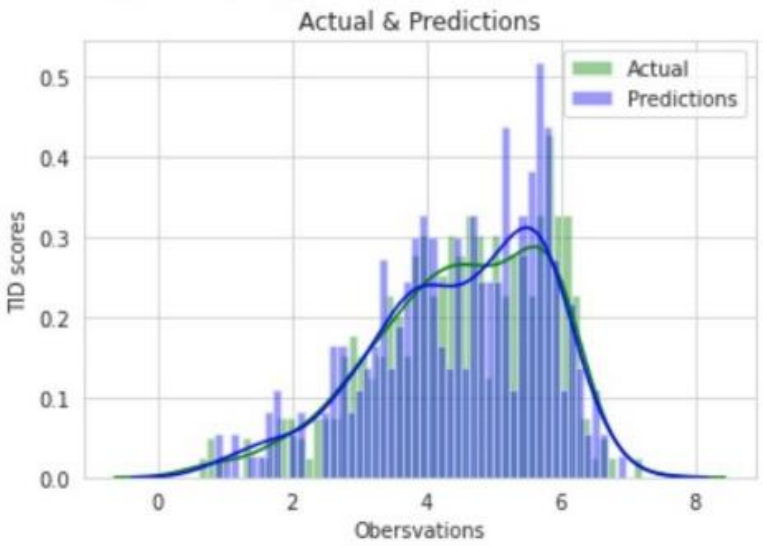

Fig. 2. Plot of Image Actual metrics and the metrics Predicted using the Proposed Approach.

Fig. 3. Images with their Actual and Predicted IQS

Table 5 shows the comparison of the proposed BIQA scheme concerning correlation coefficient metrics (SROCC and PLCC) with distinguished state-of-the-art works found in the literature for the TID2013 dataset. As seen from Table 5, the proposed BIQA scheme achieved a stronger relationship between subjective and objective IQSs obtaining higher values (0.938 and 0.945) of correlation coefficient metrics as compared to other methods. Thus, it shows that the objective measures possess the ability to evaluate the perceptual quality effectively for the distorted images.

Table 5. Quantitative Results of various BIQA schemes on TID2013 Dataset

Method	**SROCC**	**PLCC**
HyperIQA [12]	0.846	0.873
DACNN [30]	0.871	0.889
Se-Ho Lee and Kim [31]	0.877	0.894
DIIVINE [35]	0.654	0.549
BRISQUE [36]	0.604	0.694
CORNIA [37]	0.549	0.613
IL-NIQE [38]	0.523	0.673
HOSA [39]	0.688	0.764
DIQaM-NR [40]	0.835	0.855
TS-CNN [41]	0.783	0.824
CaHDC [42]	0.862	0.878
DB-CNN [43]	0.816	0.865
Proposed	**0.938**	**0.945**

5. Conclusion

We propose a simple but effective deep pre-trained convolutional neural network-based BIQA model using EfficientNet-V2 S for the TID2013 dataset. The network is evaluated with 80: 20 ratio concerning training: testing samples of the dataset distorted images for predicting the IQSs. EffcientNet-V2 S network is selected due to its ability to extract semantic relevant features from an source image to a much higher depth. The proposed BIQA scheme demonstrates state-of-the-art performance in through the SROCC and PLCC metrics. Out of 24 distortions, the proposed BIQA model is superior to other methods in 16 cases. The EfficientNet V2 S network is properly tuned by properly resizing and normalizing the input images. The last two-FCL are efficiently used to preserve the details of the image and the scores are predicted nearer to the actual scores as provided with the dataset. We achieved a stronger relationship between subjective IQSs and objective IQSs

while the objective measures are found to be effective in evaluating the perceptual quality of distorted images. The proposed scheme offers low time complexity and computational complexity. Further, the correlation between the predicted scores and the mean opinion scores of test images can be improved by increasing the iterations during training and fine-tuning other parameters such as batch size, kernel size, etc.

Other authentic and synthetic datasets can be evaluated using the proposed model. Also, a robust BIQA system can be constructed to mitigate the error between the actual score and the predicted score and improve the generalization ability of the network.

References

[1] Xue W., Zhang L., Mou X., and Bovik A., "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE Transaction on Image Processing, 2014, Vol. 23, pp. 684–695.

[2] Zhang L., Shen Y., and Li H., "VSI: A visual saliency-induced index for perceptual image quality assessment, "IEEE Transaction on Image Processing, 2014, Vol. 23, pp. 4270–4281.

[3] Chang H., Yang H., Gan Y., and Wang M., "Sparse feature fidelity for perceptual image quality assessment," IEEE Transaction on Image Processing, 2013, Vol. 22, pp. 4007–4018.

[4] Ma L., Li S., Zhang F., and Ngan K., "Reduced-reference image quality assessment using reorganized DCT-based image representation," IEEE Transaction on Multimedia, 2011, Vol. 13, pp. 824–829.

[5] Liu Y., Zhai G., Gu K., Liu X., Zhao D., and Gao W., "Reduced reference image quality assessment in free-energy principle and sparse representation," IEEE Transaction on Multimedia, 2018, Vol. 20, pp. 379–391.

[6] Wu J., Liu Y., Li L., and Shi G., "Attended visual content degradation based reduced reference image quality assessment," IEEE Access, 2018, Vol. 6, pp. 12493–12504.

[7] Zhu W., Zhai G., Min X., Hu M., Liu J., Guo G., and Yang X., "Multi-channel decomposition in tandem with free-energy principle for reduced reference image quality assessment," IEEE Transaction on Multimedia, 2019, Vol. 21, pp. 2334–2346.

[8] Gu K., Wang S., Zhai G., Ma S., Yang X., Lin W., Zhang W., and Gao W., "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," IEEE Transactions on Multimedia, 2016, Vol. 18, pp. 432–443.

[9] Zhu H., Li L., Wu J., Dong W., and Shi G., "MetaIQA: Deep meta-learning for no-reference image quality assessment," In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2022, pp. 14131–14140.

[10] Ma K., Liu W., Zhang K., Duanmu Z., Wang Z., and Zuo W., "End-to-end blind image quality assessment using deep neural networks," IEEE Transaction on Image Processing, 2018, Vol. 27, pp. 1202–1213.

[11] Bosse S., Maniry D., Miller K., Wiegand T. and Samek W., "Deep neural networks for no-reference and full-reference image quality assessment," IEEE Transactions on Image Processing, 2018, Vol. 27, pp. 206–219.

[12] Su S., Yan Q., Zhu Y., Zhang C., Ge X., Sun J., and Zhang Y., "Blindly assess image quality in the wild guided by a self-adaptive hyper network," In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020, pp. 3664–3673.

[13] Tan M. and Le Q, "Efficientnetv2: Smaller models and faster training," In International conference on machine learning, PMLR, 2021, pp. 10096–10106.

[14] Ponomarenko N., Jin L., Ieremeiev O., et al., "Image database TID2013: Peculiarities, results and perspectives," Signal Processing: Image Communication, 2015, Vol. 30, pp. 57–77.

[15] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng and Zhou Wang, "Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network," IEEE Transactions on Circuits and Systems for video Technology, January 2020, Vol. 30, No. 1.

[16] K. Ma et al., "Waterloo exploration database: New challenges for image quality assessment models," IEEE Transactions on Image Processing, February 2017, Vol. 26, No. 2, pp. 1004–1016.

[17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," International Journal of Computer Vision, June 2010 Vol. 88, No. 2, pp. 303–338.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Represent., 2015.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," In Proceeding of IEEE Conference of Computer Vision and Pattern Recognition, June 2009, pp. 248–255.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proc. IEEE Conf. Comput. Vis. Pattern Recognit., June 2016, pp. 770–778.

[21] Hu L., Peng J., Zhao T., Yu W., and Hu B., "A Blind Image Quality Index for Synthetic and Authentic Distortions with Hierarchical Feature Fusion," Applied Science, 2023, Vol. 13, 3591.

[22] Larson E. and Chandler D., "Most apparent distortion: Full-reference image quality assessment and the role of strategy," Journal of Electron. Imaging, 2010, Vol. 19, pp. 1–21.

[23] Ciancio A., Costa A., Silva E., Said A., Samadani R., and Obrador P., "No-reference blur assessment of digital pictures based on multi-feature classifiers, "IEEE Transactions on Image Processing, 2011, Vol. 20, pp. 64–75.

[24] Ghadiyaram D. and Bovik A., "Massive online crowd-sourced study of subjective and objective picture quality," IEEE Transactions on Image Processing, 2016, Vol. 25, pp. 372–387.

[25] Hosu V., Lin H., Sziranyi T. and Saupe D., "Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment," IEEE Transactions on Image Processing, 2020, Vol. 29, pp. 4041-4056.

[26] Lijing Lai, Jun Chu, and Lu Leng, "No-Reference Image Quality Assessment based on Quality Awareness Feature and Multi-task Training," Journal of Multimedia Information System, vol. 9, No. 2, June 2022, pp. 75-86.

[27] Feng C., Ye L., and Zhang Q., "Cross-Domain Feature Similarity Guided Blind Image Quality Assessment," Frontiers in Neuroscience, 2022, Vol. 5, 767977.

[28] He W. and Luo Z., "Blind Quality Assessment of Images Containing Objects of Interest," Sensors, 2023, Vol. 23, 8205.

[29] Ning Guo, Letu Qingge, YuanChen Huang, Kaushik Roy, YangGui Li and Pei Yang, "Blind Image Quality Assessment via Multi-perspective Consistency," International Journal of Intelligent Systems, Vol. 2023, Article ID 4631995, 14 pages.

[30] Pan Z., Zhang H., Lei J., et al., "DACNN: Blind image quality assessment via a distortion-aware convolutional neural network," IEEE Transactions on Circuits and Systems for Video Technology, 2022, Vol. 32, pp. 7518– 7531.

[31] Lee SH and Kim SW, "Dual-branch vision transformer for blind image quality assessment," Journal of Visual Communication and Image Representation, 2023, Vol. 94, 103850.

[32] Tan Ma and Le Q., "Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning," PMLR, 2019, pp. 6105–6114.

[33] Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., et al., "Imagenet large-scale visual recognition challenge. International Journal of Computer Vision, 2015, Vol. 115, No. 3, pp. 211–252.

[34] Xu L., Lin W., and Kuo C., "Visual Quality Assessment by Machine Learning," Springer Singapore, Imprint: Springer, 2015.

[35] Moorthy AK and Bovik AC, "Blind image quality assessment: From natural scene statistics to perceptual quality," IEEE Transactions on Image Processing, 2011, Vol. 20, No. 12, pp. 3350–3364.

[36] Mittal A., Moorthy AK, and Bovik AC, "No-reference image quality assessment in the spatial domain," IEEE Transactions on Image Processing, 2012, Vol. 21, pp. 4695–4708.

[37] Ye P., Kumar J., Kang L., et al. "Unsupervised feature learning framework for no-reference image quality assessment," IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1098–1105.

[38] Zhang W., Ma K., Deng D., et al. "Blind image quality assessment using a deep bilinear convolutional neural network," IEEE Transactions on Circuits and Systems for Video Technology, 2020, Vol. 30, pp. 36–47.

[39] Xu J., Ye P., Li Q., et al. "Blind image quality assessment based on high order statistics aggregation. IEEE Transactions on Image Processing, 2016, Vol. 25, No. 9, pp. 4444–4457.

[40] Kim J. and Lee S., "Fully deep blind image quality predictor," IEEE Journal on Selected Topics in Signal Processing, 2017, Vol. 11, pp. 206–220.

[41] Yan Q., Gong D., and Zhang Y., "Two-stream convolutional networks for blind image quality assessment," IEEE Transactions on Image Processing, 2019, Vol. 28, pp. 2200–2211.

[42] Wu J., Ma J., Liang F., et al., "End-to-end blind image quality prediction with a cascaded deep neural network," IEEE Transactions on Image Processing, 2020, Vol. 29, pp. 7414–7426.

[43] Zhang W., Ma K., Zhai G., et al., "Learning to blindly assess image quality in the laboratory and wild," In Proceeding of IEEE International Conference on Image Processing (ICIP), 2020, pp. 111–115.