Signal, Image and Video Processing (2020) 14:205–213
https://doi.org/10.1007/s11760-019-01543-z

ORIGINAL PAPER

An improved model for no-reference image quality assessment and a no-reference video quality assessment model based on frame analysis

Mukesh Kumar Rohil[1] · Neetika Gupta[1] · Prakash Yadav[1]

Received: 4 September 2018 / Revised: 3 June 2019 / Accepted: 17 July 2019 / Published online: 14 August 2019

Abstract

No-reference image quality assessment (NR-IQA) uses only the test image for its quality assessment, and as video is essentially comprised of image frames with additional temporal dimension, video quality assessment (VQA) requires a thorough understanding of image quality assessment metrics and models. Therefore, in order to identify features that deteriorate video quality, a fundamental analysis of spatial and temporal artifacts with respect to individual video frames needs to be performed. Existing IQA and VQA metrics are primarily for capturing few distortions and hence may not be good for all types of images and videos. In this paper, we propose an NR-IQA model by combining existing three methods (namely NIQE, BRISQUE and BLIINDS-II) using multi-linear regression. We also present a holistic no-reference video quality assessment (NR-VQA) model by exploring quantification of certain distortions like ringing, frame difference, blocking, clipping and contrast in video frames. For the proposed NR-IQA model, the results represent improved performance as compared to the state-of-the-art methods and it requires very low fraction of samples for training to provide a consistent accuracy over different training-to-testing ratios. The performance of NR-VQA model is examined using a simple neural network model to attain high value of goodness of fit.

Keywords No-reference image quality assessment · No-reference video quality assessment · Spatial artifacts · Temporal artifacts

1 Introduction

Various deformities are present in digital images while going through processes like acquisition, compression, transmission, reproduction, etc. These deformities mainly occur due to limitations of access devices, storage media, processing technologies and transmission equipment. Image distortions severely influence the ability of humans to excerpt and understand the information contained in images. Therefore, it becomes important to identify and measure image distortions in order to ensure, control and enhance image quality. In order to achieve the above objectives, many objective image quality assessment (OIQA) models have been developed and some of these are of substantial practical significance [1, 2]. Based on the availability of an original reference, image quality assessment (IQA) methods are categorized as full-reference (FR), reduced-reference (RR) and no-reference (NR) IQA methods. Most of the existing approaches fall under the category of FR-IQA, i.e., a complete reference image is assumed to be known. However, in many practical applications, the reference image is not available, making NR-IQA or 'blind' quality assessment approach desirable. Similarly, in cases when the reference image is only partially available to evaluate the quality of the distorted image, RR-IQA approach is adopted. However, FR-IQA and NR-IQA share the same limitation, i.e., features extraction from reference image becomes necessary for quality evaluation and hence adds to processing time [3].

NR-IQA can be divided into two categories, namely distortion-specific quality assessment and general-purpose

✉ Neetika Gupta
neetika.gupta@pilani.bits-pilani.ac.in

Mukesh Kumar Rohil
rohil@pilani.bits-pilani.ac.in

Prakash Yadav
prakashydv@gmail.com

[1] Department of Computer Science and Information Systems, Pilani Campus, Birla Institute of Technology and Science, Pilani, India

quality assessment. The former quantifies a specific distortion regardless of other factors and scores a distorted image accordingly. A number of NR-IQA methods following such an approach could be listed [4–10]. In a recent proposed NR-IQA metric [4], 32 natural scene statistics (NSS) features of the luminance relative order are initially extracted and then the quality score is predicted using a support vector regression. Gu et al. [5] proposed an NR-IQA model for accessing the perceptual quality of screen content pictures with big data learning. Chen and Bovik developed an NR-IQA method to quantify blur in an image [6]. Similarly, Zhu and Milanfar [7] focused on noise, Sazzad et al. [8] on JPEG2000 distortion, Sheikh et al. [9] on JPEG2000 by using NSS, Wang et al. [10] on JPEG, etc. Besides many advantages, distortion-specific approach limits its applicability with the fact that the type of distortion present in the image should be known in advance. Thus, the later approach, i.e., general-purpose quality assessment, based on training and learning is widely adapted for NR-IQA purpose. Examples include a two-step framework designed for distortion classification and distortion-specific quality assessment using several NSS features to implement a simple NR-IQA index named BIQI [11]. This method was later improved by using a series of NSS features in the wavelet domain to predict image quality (DIIVINE index) [12]. Saad et al. [13] proposed another efficient NR-IQA method named BLIINDS-II which extracts NSS features in the block discrete cosine transform (DCT) domain using a fast single-stage framework. In order to achieve better predictive performance with low computational complexity, Mittal et al. [14] proposed the BRISQUE index. Mittal et al. [15] proposed the NIQE NR-IQA index based on quality-aware collection of statistical features. The recent research in IQA using deep neural networks [16, 17] works well for only high training ratios, i.e., cases when 67% or 80% of the dataset is used as training samples. These methods, apart from being computational intensive, exhibit low performance for color images.

Over the years, research for distortion-specific video quality assessment techniques has advanced with an aim to design a universal measure for NR-VQA. Among many quantified artifacts of compression, blocking is one of the most common artifacts [18, 19]. A technique for evaluating jerkiness (also known as strobing [20]) in a video was given by Ong et al. [21] by finding the absolute frame differences between adjacent frames in a video. A non-application-specific NR-VQA model proposed by Keimel et al. [22] quantifies a multitude of factors to predict the overall video quality. In a study conducted by Saad et al. [23], an NR-VQA model based on the principles of natural video statistics is proposed where motion characteristics in a video are quantified by extracting block motion estimates and difference of the DCT coefficient between adjacent frames. Li et al. [24] proposed an NR-VQA metric based on the spatiotemporal natural video statistics in 3D discrete cosine transform (3D-DCT) domain. The recent work on VQA using convolutional neural networks (CNNs) [25] uses only low-performed FR metrics for label generation, and its performance depends on the chosen threshold value for the sample distribution.

In this study, we aimed at analyzing the image and video features that deteriorate the quality of a given image/video. While research in the field of NR-IQA and NR-VQA has progressed over the years by analyzing wide range of features that deteriorate image/video quality as discussed above, there still seems a need for developing a more generalized model for NR-IQA/VQA. Drawbacks that are to be resolved include accurate quantification of different distortion types and requirement of a minimum set of features to predict the perceived image/video quality. In this paper, we propose a way to analyze spatial artifacts in an image and spatial and temporal artifacts in a video frame for quantifying the distortions present in image/video using a generalized metric.

The rest of the paper is organized as follows. Section 2 describes the proposed model for NR-IQA. Section 3 explains the specifications and methodology used for the proposed model for NR-VQA. Section 4 describes the experiments performed and presents the results of the performance of the proposed method. Finally, Sect. 5 concludes the study and lists some future work.

2 Proposed model for no-reference image quality assessment

For the objective quality analysis of the test images, three existing NR-IQA methods are combined using multi-linear regression (MLR). The three methods are identified from the literature on image quality assessment, which are significantly different from each other; else, the distortion measurement will be redundant as the pooling of data will become erroneous. Many such models were identified in the literature; however, only NIQE, BRISQUE and BLIINDS-II are chosen because of their diverse footprint over quantifying different distortions.

1. Naturalness image quality evaluator (NIQE): It first constructs quality-aware collection of features. These features are computed as per the natural scene statistics (NSS) model. Comparative study conducted by Mittal et al. [15] shows that NIQE competes well with some of the best performing NR-IQA techniques like peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [26], BLIINDS-II, DIVINE, BRISQUE, etc., that requires training on large databases of human opinions of image

distortion. Mittal et al. [15] conclude that they have succeeded in creating a first of a kind blind IQA model that assesses image quality without knowledge of anticipated distortions or human opinions about them. The quality of the distorted image is expressed as a simple distance metric between the model statistics and those of the distorted image [15].

2. Blind/reference-less image spatial quality evaluator (BRISQUE): It uses NSS of locally normalized luminance coefficients to quantify possible losses of 'naturalness' in the image. BRISQUE is computationally less expensive than other blind image quality assessment algorithms because it does not require transforming the image into other domains. For wide range of transformations, BRISQUE [14] is proven to be statistically better than some of the FR-IQA methods such as PSNR and SSIM. Moreover, low computational complexity of BRISQUE makes it well suited for real-time applications. BRISQUE features are independent of the database and may also be used for distortion identification in images [14].
3. Blind image integrity notator using DCT statistics-II (BLIINDS-II) index: Given certain extracted features based on NSS model of image DCT coefficients, BLIINDS-II approach uses Bayesian inference model to assess image quality score. Some features that are indicative of perceptual quality are then formed by using estimated parameters of the model. Hence, BLIINDS-II adopts a simple probabilistic model for score prediction and requires minimum training. Given the extracted features from a distorted test image, the quality score that maximizes the probability of the empirically determined inference model is chosen as the predicted quality score of that image [13].

In the current work, multi-linear regression model is used to predict a single response variable Y which linearly depends upon three predictor variables (NIQE, BRISQUE and BLIINDS-II scores). Performance comparison of the three methods is made by evaluating the overall correlation with differential mean opinion score (DMOS). The combined model for blind estimation of image quality performs better than BRISQUE, which individually has the best performance among the three models. (The results are discussed in Sect. 4.)

3 Proposed model for no-reference video quality assessment

Original no-reference or blind video quality estimation remains to be the most researched field among all media quality estimation fields. There are presently no perceptual models of video distortion that may apply to the no-reference cases. However, a promising approach should consist of identifying the minimum set of features which influence the quality in most situations and examine their ability to predict perceived quality. The main problem is that certain algorithms look out for certain specific distortions only and with each additional distortion to monitor, the computational complexity goes up. Moreover, some distortions are content dependent, and this makes it difficult to come up with a general algorithm for NR-VQA.

Though it is difficult to assess the quality of a video without the availability of reference data, its applications have tremendous importance in the market. The most important application is to design flexible real-time control systems to monitor and deliver high-quality streams to consumers. Reliable metrics are important for enabling transparent and competitive ratings of quality of service (QoS), which would benefit both the consumers and the producers.

Therefore, in this study, to inspect some of the above-related issues, we have explored the quantification of certain distortions and identified a way to estimate these distortions more accurately to provide a robust NR-VQA model.

3.1 Ringing in a frame

The ringing effect is noticeable as simmers and ripples extending outwards from the edge up to the blocks which form the boundary along the edge.

There exist many methods of measuring ringing in an image; however, almost all methods transform the domain of the image. Fourier transform is popular among such transformations for transforming an image from spatial domain to the frequency domain. This is usually done to find which spatial frequencies are observable and which might be masked. Fourier transform therefore is the bottleneck to the computational complexity of all algorithms that measure ringing in an image. This is acceptable as far as it is limited for quality analysis to calibrate image compression algorithms or an image processing tool, but for large image volume or live video streams this becomes impractical. Keeping this constraint in mind, a method is designed for estimating ringing in an image without undergoing any actual transformation.

In this proposed method for identifying ringing effect in a frame, high-contrast boundaries are identified by thresholding the image with a suitable luminance threshold to generate a binary image. It is at these high-contrast edges that we need to identify 'splatter,' i.e., isolated pixels or groups of pixels that differ widely in their luminance to their surrounding pixel's luminance. For the final metric, only those regions are considered whose variance is greater than a certain threshold, calculated by averaging the variance in relatively smoother (low contrast) areas. The performance of the algorithm depends significantly on identifying smooth areas in the image.

3.2 Frame difference (ΔF)

Frame difference is a (very) rough estimate of the activity in the video. Faster temporal changes in pixel values lead to greater frame difference. However, it is video dependent and can vary drastically from one video to another. The LIVE database (discussed in Sect. 4) had no videos with scene change. All video contents refer to either panning of a natural scene or observing a smooth moving object.

Using frame difference as-it-is is a bad idea because it is highly video content dependent. However, a beneficial result of this experiment is the determination of ΔF as a strong metric for estimating scene changes on a large-scale video. Moreover, while using psycho-visual experiments, higher threshold of scene changes per second or per minute beyond which the quality of video deteriorates could also be estimated.

3.3 Blocking effect quantization

Blocking effect is the most popular video artifact. It is also among the simpler artifacts to be observed as its location is fixed in the spatial domain. Blocking is inherent in all lossy compression algorithms that use DCT. Wavelet transform is independent of this artifact, making JPEG2000 a better format quality-wise; however, it too suffers from ringing effect.

In order to quantify the blocking effect in an image or frame of a video, we first use an edge detection algorithm (sobel edge detection) to identify the perceptually sensitive areas. A single-degree differential is applied on the image in both horizontal and vertical dimensions. Depending on the block dimensions (4 or 8), we create a mask to highlight all differences at block boundaries. This makes the method computationally less intensive than actually parsing all block edges by looping. The boundary distortions are then enhanced by squaring, and root of the sums of these horizontal and vertical squared images is taken to obtain the final image. All block boundary pixel luminosities are then added to get an estimate of the blocking effect.

We consider two different block sizes for our experiment, 4×4 and 8×8. The block size 8×8 is used in MPEG-2 compressed videos and that is where the blocking effect is located; however, in the case of H.264 compressed MPEG-2/AVC videos, it supports multiple block sizes, 4×4 being among the popular ones. Though we observed that both have a high correlation (>90%) among them, we include both block sizes for completeness sake and to bring generality to our model.

3.4 Clipping

Clipping is the truncation in the number of bits of the luminance or chrominance components of the image values. It results in abrupt cutting of peak values at the top and bottom of the dynamic range, which leads to aliasing artifacts caused by the high frequencies at those discontinuities. The sharpness enhancing technique known as peaking can lead to clipping. In peaking, edges are enhanced by adding positive and negative overshoots to it, but if these values are beyond the limits of the dynamic range [(0.255) for 8 bit precision], then saturation occurs and pixels are clipped.

Clipping can be represented as the percentage of pixels having boundary values, e.g., either 0 or 255 for 8-bit precision. However, care must be taken at the margins where in some videos a blank line is introduced due to coding error. To take care of this, few pixels from the margin (say 3, reducing the image size by 6 in both directions) are ignored while quantifying clipping for a video frame.

3.5 Contrast

Contrast sensitivity, which largely depends upon the dynamic range of the luminance signal, is the ability to distinguish objects from the background. The perception of contrast is subjective because it depends upon other factors including the mental reference image of the objects and sometimes colors.

The following procedure is followed for contrast quantification:

1. Luminance histogram for the given image/video frame is computed.
2. The luminance histogram is then divided into half (vertically) so that luminosities less than half of the maximum luminosity lie on the one side of the histogram.
3. Difference between the cumulative luminance of each part is calculated and normalized by dividing the value with the average luminance of the image.

The features identified by this work are limited but diverse enough to provide an overall quality assessment of the video. Most generic algorithms quantify blocking effect to get an accuracy of around 80% [for example generalized block-edge impairment metric (GBIM) [27] which does not scale too well with videos].

In this study, four features are dedicated for estimating the quality of the video in spatial domain. Metrics to calculate the ringing effect in an image and the frame difference were explored but not considered as they did not scale too well in the case of videos. Therefore, the four features selected are: Blocking4 (blocking with block size 4×4), Blocking8 (blocking with block size 8×8), Clipping and Contrast.

Additionally two features are dedicated for estimating the quality of video in the temporal profile. Earlier in this paper (Sect. 3.2), an argument is presented regarding the unsuitability of inter-frame difference quantification for video quality estimation as it is heavily dependent upon video content and as LIVE video quality database (discussed in Sect. 4)

Table 1 MLR model for video quality assessment

Training ratio	Maximum correlation	Minimum correlation	Average correlation
0.5	0.4847	0.1733	0.26906
0.66	0.6299	0.2369	0.27458
0.75	0.6299	0.2369	0.28884
0.8	0.6707	0.2369	0.29423

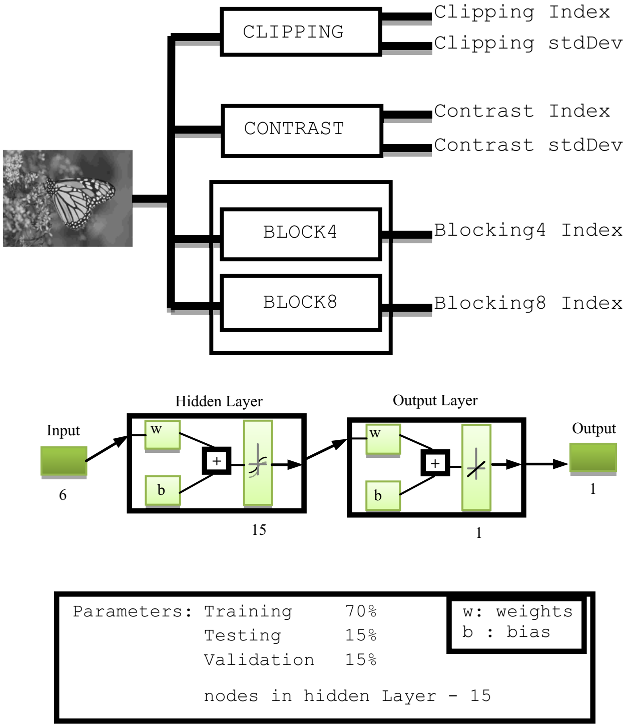


Fig. 1 No-reference video quality assessment model

restricts us to estimate quality of only smooth motion videos; therefore, the two features used for quantifying temporal distortion are: Clipping Standard Deviation [28] (observed as glaring/anti-aliasing in a video) and Contrast Standard Deviation [29] (observed as flickering in video).

If these features are treated independent of interactions among themselves, and a MLR model is developed to predict collective DMOS, the results are rather dismal even for training ratio 0.8 as shown in Table 1. The MLR analysis results are ineffectual, and this is intuitive as we have no knowledge of their interactions and is rather impractical to model such a scenario using MLR. So to fit such data, we use a neural network (NN) model [30] with one middle hidden layer of 15 units. The NN model is represented in Fig. 1.

4 Experimental setup and results

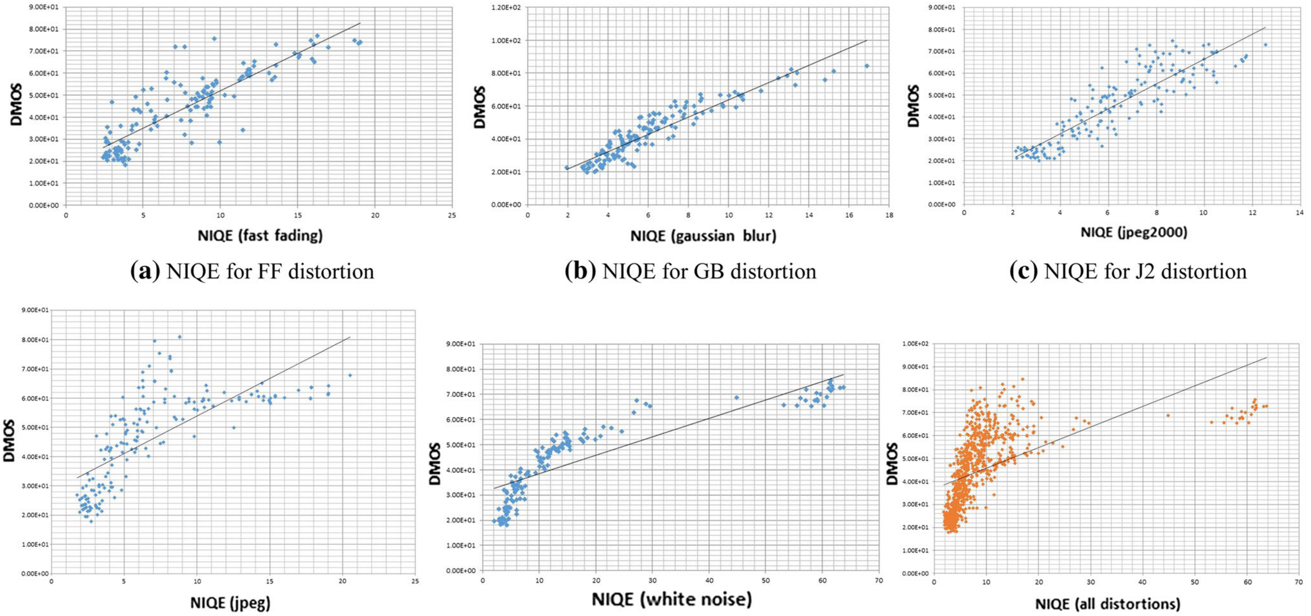
LIVE Image and Video Quality Assessment database details which are given in [31, 32], respectively, are used for the experiments. In LIVE Image Quality Assessment database [31], 29 high-resolution and high-quality color images are used as reference images. These images are distorted using five types of distortions; each type is provided separately and independent of each other: bit errors in JPEG2000 bit-streams (fast fading distortion) (FF)—145 images, Gaussian blur distortion (GB)—145 images, JPEG2000 compressed images (J2)—175 images, JPEG compressed images (J)—169 images, and white noise distortion (WN)—145 images. DMOS values for each of the 779 images are provided in MATLAB-compatible.mat files, and the value ranges between 0 and 100, where 0 indicates the bad quality and 100 indicates the good quality.

The LIVE Video Quality Assessment database [32] uses ten uncompressed high-quality videos [Blue sky (bs), Mobile and Calendar (mc), Pedestrian Area (pa), Par Run (pr), Riverbed (rb), Rush Hour (rh), SunFlower (sf), Shields (sh), Station (st) and Tractor (tr)] as reference videos with a wide variety of content. These reference videos were downsampled using various techniques to obtain the distorted videos in this database. A set of 150 distorted videos are created from these reference videos (15 distorted videos per reference) using four different distortion types: MPEG-2 compression, H.264 compression, simulated transmission of H.264 compressed bit-streams through error-prone IP networks and simulated transmission of H.264 compressed bit-streams through error-prone wireless networks. The mean and variance of the DMOS obtained from the subjective evaluations, along with the reference and distorted videos, are made available as part of the database. The DMOS scores for the LIVE Video Quality Assessment database lie in the [30, 82] range.

4.1 No-reference image quality assessment

In our experiments, it is found that NIQE performs well with most type of distortions but performs poorly for JPEG compression artifacts and white noise, which reduces its overall accuracy. Table 2 represents the results of its accuracy over five distortions of LIVE Image Quality Assessment database in terms of average DMOS correlation. Figure 2 represents six scatter plots of NIQE against five types of distortions respectively, with one additional plot for all distortion types. Figure 2a–c exhibits good performance of NIQE with FF, GB and J2 distortions, while from Fig. 2d–f we observe that the poor performance of NIQE with J and WN distortion destroys its overall estimation effort, leading to low overall accuracy of NIQE (0.541 DMOS correlation for all distortions).

Table 3 represents the results for BRISQUE accuracy for five distortion types. Figure 3 shows two scatter plots



(a) NIQE for FF distortion

(b) NIQE for GB distortion

(c) NIQE for J2 distortion

(d) NIQE for J distortion

(e) NIQE for WN distortion

(f) NIQE for all distortions

Fig. 2 Scatter plot for NIQE and DMOS for different distortions

Table 2 NIQE and DMOS correlation

Distortion	FF	GB	J2	J	WN	All
DMOS correlation	0.868	0.923	0.890	0.725	0.858	0.541

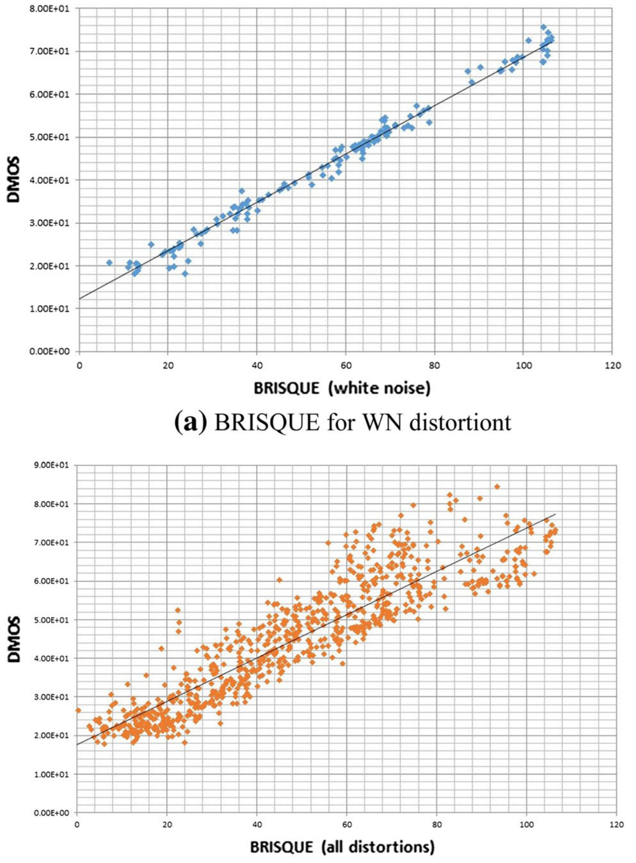
Table 3 BRISQUE and DMOS correlation

Distortion	FF	GB	J2	J	WN	All
DMOS correlation	0.935	0.968	0.931	0.897	0.991	0.910

determining the performance of BRISQUE with WN and all distortion types. In contrast to NIQE's poor performance on WN distortion, BRISQUE does exceptionally well (Fig. 3a).

It is also observed that compared to the other two blind models, BLIINDS-II is significantly slower, on the account of DCT transformations done by the algorithm. It is therefore not suitable for real-time streaming images. However, unlike the other two blind models, it provides information about DCT coefficients, which is ignored by faster algorithms. Table 4 and Fig. 4 signify overall good performance of BLIINDS-II for all distortion types.

The most noticeable result of our effort is a guaranteed lowest performance of no less than 89.5% (for high training ratio of 80% (Table 5) minimum correlation accounts to 0.8956, i.e., 89.5%). The average correlation coefficient of 91.6% is marginally better as it is 0.6% more than the best performing model (BRISQUE) (Table 3). Moreover, the standard deviation is also low (<0.75%, Table 5) compared to other models discussed so far in this study.



(a) BRISQUE for WN distortiont

(b) BRISQUE for all distortions

Fig. 3 Scatter plot for BRISQUE and DMOS for WN and all distortion types

Table 4 BLIINDS-II and DMOS correlation

Distortion	FF	GB	J2	J	WN	All
DMOS correlation	0.844	0.899	0.902	0.898	0.965	0.866

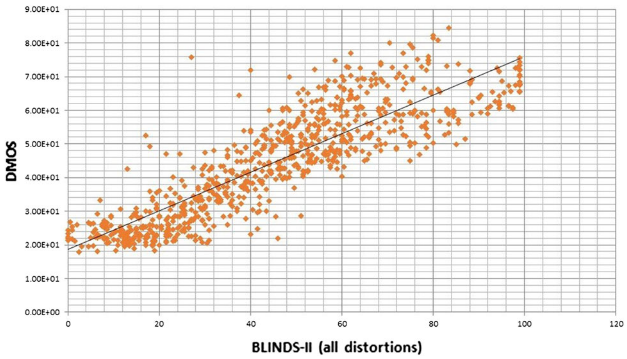


Fig. 4 Scatter plot for BLIINDS-II and DMOS for all distortion types

Table 5 Performance of the proposed NR-IQA model

Training ratio	0.5	0.66	0.75	0.8
Maximum correlation	0.9261	0.9291	0.9307	0.9351
Minimum correlation	0.9052	0.9018	0.9007	0.8956
Average correlation	0.9160	0.9164	0.9165	0.9160
Standard deviation	0.0048	0.0055	0.0063	0.0075

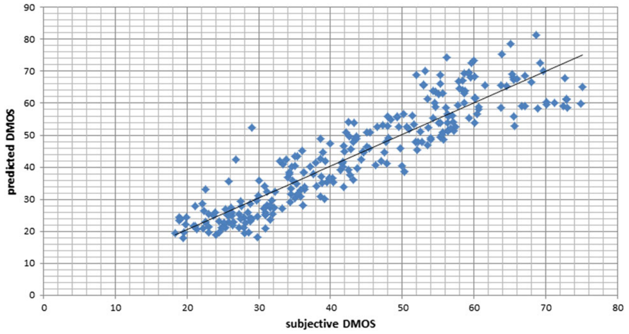


Fig. 5 Scatter plot for subjective and predicted DMOS for one iteration

Since iteration is done over 30 times, the scatter plot shown in Fig. 5 is a sample taken from one of those iterations (for arbitrarily chosen training ratio of 0.66) and it shows good correlation between the predictive and subjective DMOS values. The average correlation coefficient of 91.6% is marginally better, 0.6%, than the best performing metric (BRISQUE) (Table 3), 5% better than BLIINDS-II (Table 4) and 37.6% better than NIQE (Table 2) metrics.

What makes the combined model better than the other models is its requirement of very low fraction of samples for training to provide a consistent accuracy over many different training-to-testing ratios. Another advantage of the proposed NR-IQA model is that its performance is more or less same (row labeled 'Average Correlation' in Table 5) irrespective of the training ratio as long as the training ratio is 0.5 or above.

The marginal improvement in the proposed model with respect to BRISQUE NR-IQA metric and very high improvement for NIQE are attributed to nonlinear nature of the scatter plots shown in Fig. 2a–f, and the wide dispersion of the scatter plot is shown in Fig. 3b.

4.2 No-reference video quality assessment

Six metrics (Blocking4, Blocking8, Clipping, Clipping Standard Deviation, Contrast and Contrast Standard Deviation) are fed to the NN model with six nodes in the input layer and 15 nodes in the hidden layer. 70% data are used for training and 15% data are used for validation and testing each. The feed-forward network as implemented in MATLAB NN toolbox [27] with the default tan-sigmoid transfer function in the hidden layer and linear transfer function in the output layer is used. The network uses the default Levenberg–Marquardt algorithm [33] for training.

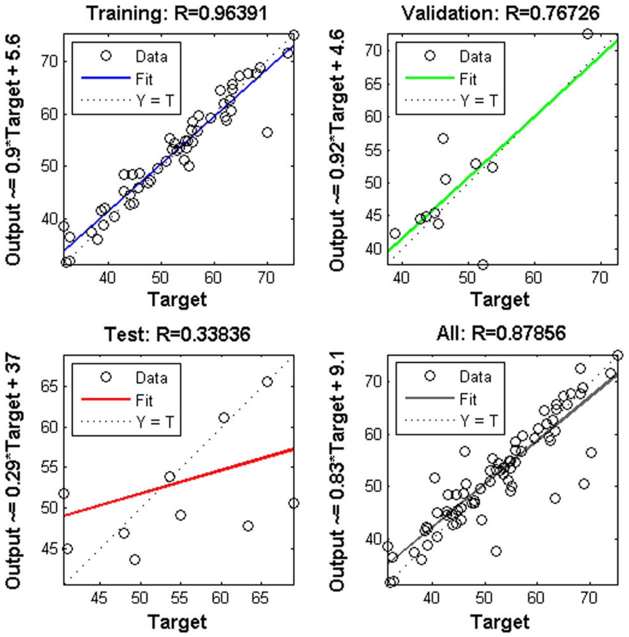
Figure 6 represents the corresponding graphs for a trial got from running the NN model for the proposed NR-VQA model. The regression plots represent the relationship between the Output (Y) (NN output) and the Target (T) value (desired output). The dashed line in the plots symbolizes the desired outcome (target value), and the solid line symbolizes the best fit linear regression line between the Output (Y) and the Target (T). For goodness-of-fit value R, if $R = 1$, an exact linear relationship exists between the Output (Y) and the Target (T). If $R = 0$, no linear relationship exists between the Output (Y) and the Target (T). The plots are shown for goodness of fit for training, validation and testing stage. It reports $R = 0.9639$ for training, $R = 0.7672$ for validation, $R = 0.3383$ for testing, and $R = 0.8785$ for all samples of dataset.

The result of the NN fitting is a goodness of fit $R = 0.8785$, a surprisingly good accuracy taking into account the limited number of features available to the model. Various other metrics for video quality assessment struggle to perform uniformly for all types of distortions, and probably our database too suffers from loss of generality.

We conclude this experiment's high accuracy in terms of goodness of fit, but the NR-VQA model still requires a wider selection of database videos to check its generality.

5 Conclusion and future work

For image NR-QA, we select three diverse NR-IQA models, namely NIQE, BRISQUE and BLIINDS-II, and perform multi-linear regression on their results to come up with a combined estimation of quality. The combined model for blind estimation of image quality performs better than BRISQUE,


Training: R=0.96391

Validation: R=0.76726

Test: R=0.33836

All: R=0.87856

Fig. 6 Results for a single training test case with 15 nodes in the hidden layer of the neural network. 70% data are used for training and 15% for testing and validation each

which individually has the best performance among the three models. Even a low fraction of samples for training provide a consistent accuracy over many different training-to-testing ratios; this makes the proposed model to execute better.

For video NR-QA, quantification for certain distortions present in video frames, namely ringing, frame difference, blocking, clipping and contrast, is examined. A neural network with one middle hidden layer of 15 units is used for fitting appropriate metrics that quantify these distortions in a video. Even though accuracy is attained in terms of goodness of fit $R = 0.8785$, the generality of the model, however, remains a drawback due to lack of availability of features that could fairly determine effects like ringing, freeze frame, etc.

Therefore, future work involves identifying more prominent and easily recognized features that affect image/video quality assessment and also determining a more efficient way of quantifying the perceived image/video quality. The proposed NR-VQA model can be improved by adding features that could be used for detecting and quantifying effects like ringing, freeze frame, etc., in video frames. Also, in the current scenario, until a nonlinear model is decided upon, neural networks serve perfectly in fitting the distortion quantifying features.

References

1. Lin, W., Kuo, C.J.: Perceptual visual quality metrics: a survey. J. Vis. Commun. Image Represent. **22**(4), 297–312 (2011)
2. Gao, X., Lu, W., Tao, D., Li, X.: Image quality assessment and human visual system. Vis. Commun. Image Process. **7744**, 77440Z-1–77440Z-10 (2010)
3. Kamble, V., Bhurchandi, K.M.: No-reference image quality assessment algorithms: a survey. Opt. Int. J. Light Electron Opt. **126**(11–12), 1090–1097 (2015)
4. Wang, T., Zhang, L., Jia, H.: An effective general-purpose NR-IQA model using natural scene statistics (NSS) of the luminance relative order. Sig. Process. Image Commun. **71**, 100–109 (2019)
5. Gu, K., Zhou, J., Zhai, G., Lin, W., Bovik, A.C.: No-reference quality assessment of screen content pictures. IEEE Trans. Image Process. **26**(8), 4005–4017 (2017)
6. Chen, M.J., Bovik, A.C.: No-reference image blur assessment using multiscale gradient. EURASIP J. Image Video Process. **1**, 1–11 (2011)
7. Zhu, X., Milanfar, P.: A no-reference sharpness metric sensitive to blur and noise. In: International Workshop on Quality of Multimedia Experience, pp. 64–69 (2009)
8. Sazzad, Z.M.P., Kawayoke, Y., Horita, Y.: No-reference image quality assessment for JPEG2000 based on spatial features. Sig. Process. Image Commun. **23**(4), 257–268 (2008)
9. Sheikh, H.R., Bovik, A.C., Cormack, L.K.: No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Trans. Image Process. **14**(11), 1918–1927 (2005)
10. Wang, Z., Bovik, A.C., Evans, B.L.: Blind measurement of blocking artifacts in images. In: Proceedings of the IEEE International Conference on Image Processing, pp. 981–984 (2000)
11. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Process. Lett. **17**(5), 513–516 (2010)
12. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans. Image Process. **20**(12), 3350–3364 (2011)
13. Saad, M., Bovik, A.C., Charrier, C.: Blind image quality assessment: a natural scene statistics approach in the DCT domain. IEEE Trans. Image Process. **21**(8), 3339–3352 (2012)
14. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
15. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Process. Lett. **22**(3), 209–212 (2013)
16. Bosse, S., Maniry, D., Muller, K.R., Wiegand, T., Samek, W.: Deep neural networks for no-reference and full-reference image quality assessment. IEEE Trans. Image Process. **27**(1), 206–219 (2018)
17. Bianco, S., Celona, L., Napoletano, P., Schettini, R.: On the use of deep learning for blind image quality assessment. SIViP **12**(2), 355–362 (2018)
18. Suthaharan, S.: Perceptual quality metric for digital video coding. IET Electron. Lett. **39**(5), 431–433 (2003)
19. Muijs, R., Kirenko, I.: A no-reference blocking artifact measure for adaptive video processing. In: Proceedings of European Signal Processing Conference, pp. 1–4 (2005)
20. Ou, Y.F., Ma, Z., Liu, T., Wang, Y.: Perceptual quality assessment of video considering both frame rate and quantization artifacts. IEEE Trans. Circuits Syst. Video Technol. **21**(3), 286–298 (2011)
21. Ong, E.P., Wu, S., Loke, M.H., Rahardja, S., Tay, J., Tan, C.K., Huang, L.: Video quality monitoring of streamed videos. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1153–1156 (2009)
22. Keimel, C., Oelbaum, T., Diepold, K.: No-reference video quality evaluation for high-definition video. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1145–1148 (2009)
23. Saad, M.A., Bovik, A.C.: Blind quality assessment of videos using a model of natural scene statistics and motion coherency. In: IEEE

Conference Record of the 46th Asilomar Conference on Signals, Systems and Computers, pp. 332–336 (2012)

24. Li, X., Guo, Q., Lu, X.: Spatiotemporal statistics for video quality assessment. IEEE Trans. Image Process. **25**(7), 3329–3342 (2016)
25. Zhang, Y., Gao, X., He, L., Lu, W., He, R.: Objective video quality assessment combining transfer learning with CNN. IEEE Trans. Neural Netw. Learn. Syst. (2019). https://doi.org/10.1109/TNNLS.2018.2890310
26. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
27. Wu, H.R., Yuen, M.: Generalized block-edge impairment metric (GBIM) for video coding. IEEE Signal Process. Lett. **4**(11), 317–320 (1997)
28. Turkowski, K.: Anti-aliasing through the use of coordinate transformations. ACM Trans. Graph. **1**(3), 215–234 (1982)
29. Farrell, J.E., Benson, B.L., Haynie, C.R.: Predicting flicker thresholds for video display terminals. Proc. SID **28**(4), 449–453 (1987)
30. Demuth, H., Beale, M.: Matlab Neural Network Toolbox User's Guide Version 6. The MathWorks Inc., Natick (2009)
31. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006)
32. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: Study of subjective and objective quality assessment of video. IEEE Trans. Image Process. **19**(6), 1427–1441 (2010)
33. Moré, J.J.: The Levenberg–Marquardt algorithm: implementation and theory. In: Watson, G.A. (eds.) Numerical Analysis. Lecture Notes in Mathematics, vol. 630, pp. 105–116. Springer, Berlin (1978)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.