

Potential of Deep Features for Opinion-Unaware, Distortion-Unaware, No-Reference Image Quality Assessment

Subhayan Mukherjee, Giuseppe Valenzise, Irene Cheng

▶ **To cite this version:**

Subhayan Mukherjee, Giuseppe Valenzise, Irene Cheng. Potential of Deep Features for Opinion-Unaware, Distortion-Unaware, No-Reference Image Quality Assessment. International conférence of Smart Multimedia 2019, Dec 2019, San Diego, CA, United States. pp.87-95, 10.1007/978-3-030-54407-2_8 . hal-03335535

HAL Id: hal-03335535
https://hal.science/hal-03335535

Submitted on 6 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1911.11903v1 [eess.IV] 27 Nov 2019

Potential of deep features for opinion-unaware, distortion-unaware, no-reference image quality assessment*

Subhayan Mukherjee[1][0000−0002−6479−3893], Giuseppe Valenzise[2], and Irene Cheng[1][0000−0001−9699−4895]

[1] University of Alberta, Edmonton AB T6G 2R3, Canada
{mukherje,locheng}@ualberta.ca
[2] CNRS - CentraleSupelec - Universit Paris-Sud, 91192 Gif-sur-Yvette Cedex, France
giuseppe.valenzise@l2s.centralesupelec.fr

Abstract. Image Quality Assessment algorithms predict a quality score for a pristine or distorted input image, such that it correlates with human opinion. Traditional methods required a non-distorted "reference" version of the input image to compare with, in order to predict this score. However, recent "No-reference" methods circumvent this requirement by modelling the distribution of clean image features, thereby making them more suitable for practical use. However, majority of such methods either use hand-crafted features or require training on human opinion scores (supervised learning), which are difficult to obtain and standardise. We explore the possibility of using deep features instead, particularly, the encoded (bottleneck) feature maps of a Convolutional Autoencoder neural network architecture. Also, we do not train the network on subjective scores (unsupervised learning). The primary requirements for an IQA method are monotonic increase in predicted scores with increasing degree of input image distortion, and consistent ranking of images with the same distortion type and content, but different distortion levels. Quantitative experiments using the Pearson, Kendall and Spearman correlation scores on a diverse set of images show that our proposed method meets the above requirements better than the state-of-art method (which uses hand-crafted features) for three types of distortions: blurring, noise and compression artefacts. This demonstrates the potential for future research in this relatively unexplored sub-area within IQA.

Keywords: Image Quality Assessment · Opinion Unaware · Distortion Unaware · No Reference · Deep Learning.

1 Introduction

Before the invention of digital cameras and other consumer grade digital imaging devices, the capture of images was quite limited. The time from capture to visualization was significant as in case of film cameras, we had to develop the photos

* Supported by NSERC Discovery Grant and DND Supplement.

in a dark room with chemical solutions. However, nowadays, with the advent of digital photography, coupled with the explosion in bandwidth of transmission channels like the Internet and social media, a tremendous volume of images are being captured and shared. Curating this huge volume of visual data that is being captured, stored, transmitted and viewed is a challenging task. Transmission of visual content occupies a large amount of Internet bandwidth. To meet the in-time transmission constraints limited by hardware resources, images and videos are usually processed and compressed before transmission and storage. Quality reductions happen as a trade-off between limited hardware resources and visual fidelity. Thus, automatic quality assessment methods are desirable to estimate the human-perceived quality measure, to replace subjective human perception. The distortions can be wide and varied; A common type of distortion is noise. Noise affects images captured using all types of sensors (optical or otherwise) and can even get injected into the image signal during transmission, for example through a telecommunications channel like television. The image quality research domain has focused on quality assessment of natural images and videos, since that is the dominant form of images we deal with everyday.

Existing IQA methods can be classified into three categories, based on the amount of the information that is available to the method: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) methods [10]. FR quality assessment requires both access to the distorted images or videos as well as the clean references. RR utilizes the limited information, depending on the actual situation, regarding the reference, rather than the full reference itself, together with the distorted images. NR approaches usually perform automatic quality assessment of the images or videos using only the distorted sources. We focus specifically on the No-Reference quality assessment of images. We can draw a parallel with the Human Visual System (HVS) which has the ability to distinguish between natural and distorted scenes based on few visual memories learned while the human brain processes visual information in various ways. Most NR-IQA algorithms follow the two step process (1) feature extraction and (2) quality prediction [15]. The schematic diagram of the same is shown in Fig.1.

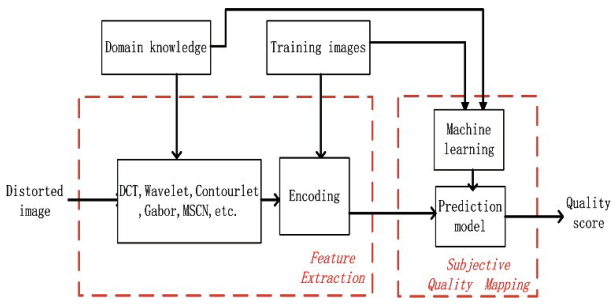


Fig. 1. General framework of NR-IQA algorithms.

Recent attempts at image quality assessment aim to mimic the response of the HVS which can mask certain artefacts depending on the location of the artefact in the image and the surrounding image content, brightness, contrast, etc. Such "perceptual" methods can be categorized into different classes based on availability of subjective rating scores for the distorted images, and knowledge about the type of possible distortions. Of those, we are interested in the specific case where the subjective rating scores are unavailable (opinion-unaware/OU) and the type of distortions is unknown (distortion-unaware or 'general'). Convolutional Neural Networks (CNNs) have been successfully used in different computer vision tasks, including style transfer, image generation, etc. Internal activations of deep convolutional networks, trained on high-level image classification tasks, have been found to be useful for natural feature representation, and are used to mimic human perception. In this work, we propose perceptual quality assessment using such "deep" features, and objectively evaluate its effectiveness.

1.1 Summary of Contributions

IQA has already been a quite active research area for several decades; Thus, our contribution is focused to a very specific sub-area (which has not received much attention) at the intersection of the following paradigms:

1. Using deep instead of hand-crafted features for image representation.
2. Using non-parametric instead of parametric approach to fit the distribution of pristine image features, and compare it to that of query image features.
3. Using unsupervised instead of supervised learning, thus removing the dependence on manual labelling (subjective scores) to train the IQA model.

The rest of the paper is organized as follows: Section 2 summarizes related IQA research. Section 3 describes how proposed method design differs from existing ones. Section 4 presents experimental results. Section 5 concludes the paper.

2 Background and Related Work

IQA methods may be broadly classified as reference or no-reference ("NR/blind") [10]. In simple terms, all the reference-based methods try to estimate some form of *distance* between the reference ("clean") image and the input/query image. The larger the distance, the greater the distortion score. No-reference methods predict the image quality based on the distorted image itself, without the need of a reference image. Most NR-IQA algorithms follow the two step process (1) feature extraction and (2) quality prediction [15]. NR-IQA algorithms are further categorized as (a) distortion-specific and (b) general / universal. The former assumes that the distortion-type is known, and employ distortion model(s) to predict one or more types of distortions in the image like noise, blur, blocking, ringing etc. to estimate its overall visual quality. The latter broadly assumes that natural scenes contain repeating patterns with a definite set of statistical properties called Natural Scene Statistics (NSS) in the spatial [8] or transformed [13]

domain, and distortions to natural images distort these properties in measurable ways. Researchers also explored more effective ways to characterize structural and contrast distortions by modelling the gradient magnitudes of natural images as Weibull distribution, in the works popularized as (IL)NIQE [9,16].

For score prediction, a popular approach is to fit the joint distribution of the feature vector and the associated opinion scores to a subset of the training data [13]. In this case, the score prediction amounts to maximizing the probability of opinion score of test data, given the test data feature vector. Other approaches quantify the distance in sparse feature space between reference and distorted images [12] in a manner that is both opinion-unaware and distortion-unaware. More recent works are based on machine learning, and extend to High Dynamic Range (HDR) images [5], though for this method the training is opinion-aware.

Very recent methods in the opinion/distortion unaware domain use (IL)NIQE features, but consider activations in pre-trained deep neural networks to select salient patches. They assign more weight to scores from those patches over others during score aggregation [17]. (IL)NIQE features can also reliably predict quality of multi-spectral images [18]. Few methods even tread the boundaries of opinion (un-)awareness and/or (no-)reference [11,6]. But even they do not operate at the intersection of paradigms outlined in Section 1.1, which motivates our research.

3 Proposed Method

Below, we briefly outline how the proposed method's functionality as visualised in Fig. 3 conforms to the intersection of paradigms outlined in Section 1.1:

3.1 Deep features for image representation

We use the local normalization non-linearity inspired by local gain control behaviors in biological visual systems in the 256-channel end-to-end compression architecture [2]. The CNN architecture used in the proposed method is shown in Fig. 2. The analysis transform block progressively down-samples the input image patch by a factor of 4,2,2 respectively, and uses the forward Generalized Divisive Normalization activation [2] for all except the last layer. The synthesis transform block progressively up-samples the output of the analysis transform block by a factor of 2,2,4 respectively, and uses the *inverse* Generalized Divisive Normalization activation [2] for all except the last layer. However, contrary to the authors in [2], our aim is not image compression; Hence, we remove the rate-distortion term from the loss function and re-train the network on the DIV2K dataset [1,4]. Random, overlapping 256×256 random patches are extracted from each 2K resolution color image. All patches are aggregated and shuffled into batches of 32 patches. Thus, the training is completely unsupervised, and does not use any subjective scores. Note that the CNN in Fig. 2 is used for feature extraction, but not score prediction, thereby making our proposed method opinion-*unaware*.

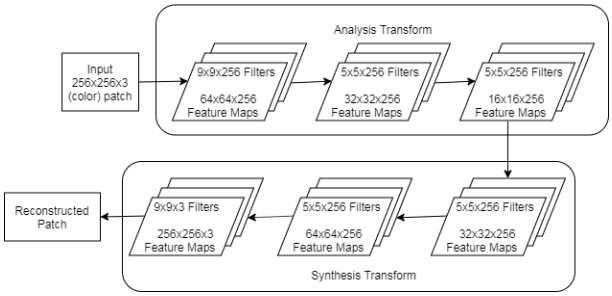


Fig. 2. Architecture of CNN used as feature extractor.

3.2 Non-parametric modelling of feature distribution

We use Kernel Density Estimation with Epanechnikov kernels to model any arbitrary-shaped distribution of the features in the encoded layer of the auto-encoder architecture (output of analysis transform block), for building the natural model. This choice was motivated by experiments which showed that the distribution of those features do not conform well to any well-known distribution. This is unlike other NR-IQA methods which mostly fit different types of parametric distributions to hand-crafted features for training and testing. The benefit of using a non-parametric approach is that we need very little information about the underlying distribution. In such scenarios, we cannot properly specify a parametric model. Thus, we can think of non-parametric models as much "broader" than parametric ones. Specifically, the kernel density estimator model usually just assumes that the probability density function of the *true* distribution from which the data are sampled satisfies 'smoothness' conditions like continuity or differentiability. Eq. 1 describes a kernel density estimator fitted to n observations X_1,...,X_n, where h (positive, chosen empirically) is the *bandwidth*, and K is the function representing the kernel, such that it outputs only positive values which sum (integrate) to 1 over the set of all observations (real numbers).

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \tag{1}$$

3.3 Opinion- and Distortion-unaware training and score prediction

After opinion-unaware training of the natural model, we predict the score for any given input image by comparing its distribution of encoded features with those of the natural model, using KL-Divergence. Thus, neither the training nor the testing phase uses any subjective scores or prior knowledge of any expected type(s) of distortions, making the proposed method opinion-and-distortion-*unaware*.

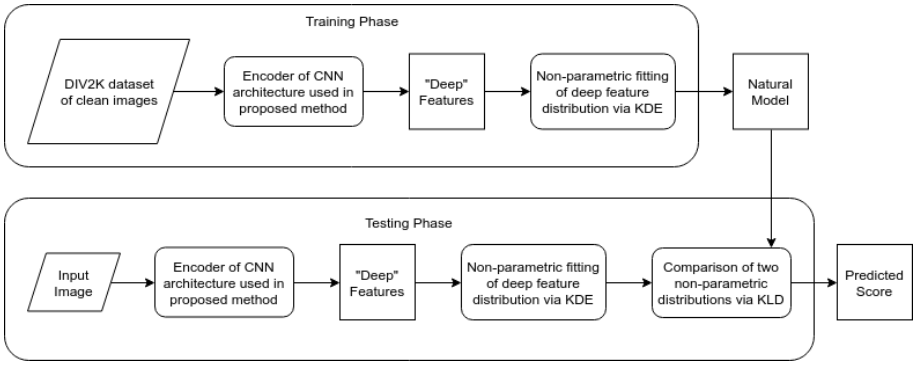


Fig. 3. Architecture of proposed method.

4 Results

4.1 Dataset

Two primary requirements for an IQA method [7] are monotonicity (monotonic increase in predicted scores with increasing degree of input image distortion), and consistency (consistent ranking of images with the same distortion type and content, but different distortion levels). To evaluate the same for our proposed method and compare against other state-of-the-art methods, we randomly selected 20 images from the General-100 dataset [3] and distorted them with five different levels of each of (Gaussian) blurring, (AWGN) noise and (JPEG-2000) compression artefacts. Thus, our test dataset had $20\times(1 \text{ clean} + 15 \text{ distorted}) = 20\times16 = 320$ images.

4.2 Context of comparison with state-of-art NR-IQA methods

We compare the proposed method with three state-of-art methods that operate under similar as well as relaxed constraints. As explained earlier, the proposed method is both opinion-unaware (OU) and distortion-unaware (DU) and we first compare it against another "(OU, DU)" method, NIQE [9]. Next, we compare with an opinion-*aware*, distortion-unaware "(OA, DU)" method, BRISQUE [8]. Lastly, we compare against an opinion-unaware, distortion-*aware* "(OU, DA)" method, PIQE [14]. Understandably, NIQE has similar constraints as proposed method, whereas BRISQUE and PIQE operate under relaxed constraints. Thus, it is easier for the last two methods to perform better than our proposed method, because they have more information available to them, although their application scenarios are much more limited than the proposed method, as explained earlier. To remind the reader, OA methods like BRISQUE require supervised training on subjective scores which are difficult to obtain and standardize, and suffers from generalization concerns. DA methods like PIQE have unpredictable performance for (combinations of) distortion types they haven not been designed to detect.

4.3 Comparison metrics

Pearson (Eq. 2), Kendall (Eq. 3) and Spearman (Eq. 4) correlation of predicted quality scores with distortion levels (0 through 5) were calculated. The average over all images and distortion types for all methods are reported in Table 1.

Pearson's correlation coefficient ρ_P measures the linear relationship between two variables X and Y, which have standard deviations σ_x and σ_y respectively, and co-variance $cov(X, Y)$. ρ_P can have a maximum value of $+1$ denoting perfect positive relationship and a minimum value of -1 denoting perfect negative relationship between X and Y, while a value of 0 indicates no relationship.

$$\rho_P = \frac{cov(X, Y)}{\sigma_x \sigma_y} \tag{2}$$

Kendall's correlation coefficient τ quantifies the degree of monotone relationship between two *ranked* variables X and Y, each having n observations. Total number of possible pairings of observations from two variables is $\binom{n}{2} = \frac{n(n-1)}{2}$. In some of those pairs, the order in which the observations are ranked are same for both variables ("concordant pairs", c) and in other pairs, the order in which the observations are ranked are different for both variables ("discordant pairs", d) such that $n = c + d$ and $S = c - d$ in Eq. 3. It follows that when $c = n$ and $d = 0$, $\tau = +1$ (perfect positive correlation), when $c = 0$ and $d = n$, $\tau = -1$ (perfect negative correlation), and when $c = d$, $\tau = 0$ (no correlation).

$$\tau = \frac{c - d}{c + d} = \frac{S}{\binom{n}{2}} = \frac{2S}{n(n-1)} \tag{3}$$

Spearman's correlation coefficient ρ_S measures the relationship between n observations of two *ranked* variables X and Y, where d_i is the pairwise difference of the variables' ranks. A value of $+1$ denotes perfect positive correlation, -1 indicates perfect negative correlation, and a 0 value indicates no correlation.

$$\rho_S = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{4}$$

Table 1. Performance comparison of proposed method with state-of-art NR-IQA methods operating under similar (NIQE) and relaxed constraints (BRISQUE and PIQE).

Method (constraints)	Pearson score	Kendall score	Spearman score	Time (sec.)
Proposed (OU, DU)	0.91	0.88	0.92	42.91
NIQE [9] (OU, DU)	0.85	0.85	0.89	0.04
BRISQUE [8] (OA, DU)	0.78	0.68	0.75	0.04
PIQE [14] (OU, DA)	0.89	0.90	0.94	0.06

Table 1 shows better correlation scores for proposed method against a state-of-art method operating under similar constraints (NIQE), and even one which

operates under relaxed constraints (BRISQUE). Another type of relaxed constraint method (PIQE) performs slightly better for distortions types it has been designed to detect. However, the proposed method's execution time is significantly higher than all compared methods, and thus it has room for improvement.

5 Conclusion and Future Work

We proposed a no-reference IQA method in an otherwise unexplored intersection of paradigms. We showed how biologically inspired activation in CNN layers can encode image patches in a reduced dimension, that captures the degree of distortion in the patch, without training on subjective opinion scores or assumption about possible distortion types. We showed via objective evaluation, the superior performance of our proposed method over the state-of-art. The next stage of our research will focus on improving the execution time of our proposed method, as well as further validating our proposed method on much larger datasets.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017)
2. Ballé, J., Laparra, V., Simoncelli, E.P.: End-to-end optimized image compression. In: Int'l. Conf. on Learning Representations (ICLR2017). Toulon, France (April 2017), available at http://arxiv.org/abs/1611.01704
3. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 391–407. Springer International Publishing, Cham (2016)
4. Ignatov, A., Timofte, R., et al.: Pirm challenge on perceptual image enhancement on smartphones: report. In: European Conference on Computer Vision (ECCV) Workshops (January 2019)
5. Kottayil, N.K., Valenzise, G., Dufaux, F., Cheng, I.: Blind quality estimation by disentangling perceptual and noisy features in high dynamic range images. IEEE Transactions on Image Processing **27**(3), 1512–1525 (March 2018)
6. Lin, K.Y., Wang, G.: Hallucinated-iqa: No-reference image quality assessment via adversarial learning. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
7. Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., Zhang, L.: Waterloo Exploration Database: New challenges for image quality assessment models. IEEE Transactions on Image Processing **26**(2), 1004–1016 (Feb 2017)
8. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing **21**(12), 4695–4708 (Dec 2012)
9. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (March 2013)
10. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference approaches to image and video quality assessment. In: Multimedia Quality of Experience (QoE), pp. 99–121. John Wiley & Sons, Ltd (Nov 2015)

11. Pan, D., Shi, P., Hou, M., Ying, Z., Fu, S., Zhang, Y.: Blind predicting similar quality map for image quality assessment. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
12. Priya, K.V.S.N.L.M., Channappayya, S.S.: A novel sparsity-inspired blind image quality assessment algorithm. In: 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP). pp. 984–988 (Dec 2014)
13. Saad, M.A., Bovik, A.C., Charrier, C.: A dct statistics-based blind image quality index. IEEE Signal Processing Letters **17**(6), 583–586 (June 2010)
14. Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: 2015 Twenty First National Conference on Communications (NCC). pp. 1–6 (Feb 2015)
15. Xu, S., Jiang, S., Min, W.: No-reference/blind image quality assessment: A survey. IETE Technical Review **34**(3), 223–245 (Apr 2016)
16. Zhang, L., Zhang, L., Bovik, A.C.: A feature-enriched completely blind image quality evaluator. IEEE Transactions on Image Processing **24**(8), 2579–2591 (Aug 2015)
17. Zhang, Z., Wang, H., Liu, S., Durrani, T.: Deep activation pooling for blind image quality assessment. Applied Sciences **8**(4), 478 (Mar 2018)
18. Zhou, B., Shao, F., Meng, X., Fu, R., Ho, Y.: No-reference quality assessment for pansharpened images via opinion-unaware learning. IEEE Access **7**, 40388–40401 (2019)