Signal Processing: Image Communication 130 (2025) 117212

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Full-reference calibration-free image quality assessment

Paolo Giannitrapani *, Elio D. Di Claudio[1], Giovanni Jacovitti[1]

Department of Information Engineering, Electronics and Telecommunications (DIET), University of Rome "Sapienza", Via Eudossiana 18, I-00184 Rome, Italy

ARTICLE INFO

Keywords:
Image quality assessment
Psycho-visual calibration
Perceptual equivalence
Positional fisher information
Viewing distance

ABSTRACT

Objective Image Quality Assessment (IQA) methods often lack of linearity of their quality estimates with respect to scores expressed by human subjects and therefore IQA metrics undergo a calibration process based on subjective quality examples. However, example-based training presents a challenge in terms of generalization hampering result comparison across different applications and operative conditions. In this paper, new Full Reference (FR) techniques, providing estimates linearly correlated with human scores without using calibration are introduced. We show that on natural images, application of estimation theory and psychophysical principles to images degraded by Gaussian blur leads to a so-called canonical IQA method, whose estimates are linearly correlated to both the subjective scores and the viewing distance. Then, we show that any mainstream IQA methods can be reconducted to the canonical method by converting its metric based on a unique specimen image. The proposed scheme is extended to wide classes of degraded images, e.g. noisy and compressed images. The resulting calibration-free FR IQA methods allows for comparability and interoperability across different imaging systems and on different viewing distances. A comparison of their statistical performance with respect to state-of-the-art calibration prone methods is finally provided, showing that the presented model is a valid alternative to the final 5-parameter calibration step of IQA methods, and the two parameters of the model have a clear operational meaning and are simply determined in practical applications. The enhanced performance are achieved across multiple viewing distance databases by independently realigning the blur values associated with each distance.

1. Introductive notes

Image Quality Assessment (IQA) is based on *subjective* as well as *objective* methods.

The *subjective quality* of an image is defined as the average quality score assigned to it by a reference class of human subjects, usually expressed in the MOS (Mean Opinion Score) scale. Likewise, the *subjective quality loss* with respect to a pristine version of the same image considered of perfect quality (also referred to as reference or original image) is expressed using DMOS (Difference of Mean Opinion Score) units.

The *objective quality* consists instead of the algorithmic prediction of the subjective quality based on measurable image features, expressed as well in MOS/DMOS units. In the recent decades, many objective quality assessment methods have been proposed. Formulation of existing metrics was inspired by different criteria. Some metrics measure the similarity among image representations (reproduction fidelity) [1–4], possibly accounting for constraints suggested by more or less sophisticated models of the Human Visual System (HVS). Other classical metrics measure a loss of a somehow defined *visual information* caused by the image degradation, with explicit reference to a cognitive interpretation of the HVS role [5,6]. Metrics have mostly the form of scalar metrics, even if some vector metrics have been adopted to distinguish among different types of image degradation [7].

Methods requiring a complete representation of the reference image are referred to as Full Reference (FR) IQA methods, while methods using incomplete representations are referred to as Reduced Ref. [8]. Methods based only on the knowledge of degraded images are finally referred to as No Reference methods [9]. The present paper concerns FR methods.

The typical scheme of objective FR IQA methods includes (i) a local comparative analysis of corresponding details of the pristine and of the degraded images, followed by (ii) a pooling stage over the whole image; the result ζ is referred to as IQA *metric* (sometimes termed "objective quality" for short). The final step is (iii) the scoring function $\hat{m}(\zeta)$ converts the IQA metric ζ into the estimated subjective MOS/DMOS $\hat{m}$ (for short, the result is sometimes referred to as "subjective quality").

The scoring function is typically parametric and it models threshold and saturation phenomena. Parameters are adjusted by non-linear

* Corresponding author.
E-mail address: paolo.giannitrapani@uniroma1.it (P. Giannitrapani).
[1] Formerly with DIET

https://doi.org/10.1016/j.image.2024.117212
Received 31 December 2023; Received in revised form 2 August 2024; Accepted 11 September 2024
Available online 23 September 2024

regression using available empirical examples. The larger the number of parameters, the more accurate is the fitting of the specific training examples during calibration and the larger the variability across different sets of examples. The VQEG scoring function $\hat{m}_{VQEG}(\zeta)$ for a scalar metric consists of a logistic function whose typical form is [6,10]:

$$\hat{m}_{VQEG}(\zeta) = \beta_1 \left[\frac{1}{2} - \frac{1}{1+e^{\beta_2(\zeta-\beta_3)}}\right] + \beta_4\zeta + \beta_5 \tag{1}$$

where ζ is the IQA metric value, $\hat{m}(\zeta)$ is the estimated DMOS value [11], and β_i are parameters usually adjusted minimizing the Euclidean distance of empirical DMOS values from the estimated ones. In [12] a S-shaped function regulated by three parameters only is advocated.

Calibration compensates for the variability of methodologies employed for subjective quality example collection. Still, it is hard to find default values for calibration parameters in the available literature and this prevents the fair comparison of quality measurement across different applications [13].

The calibration parameters critically depend on many factors, namely protocols, computation of the MOS/DMOS values from raw data, different experimental settings such as display contrast, room illumination, and mostly they depend on the viewing distance [14]. In fact, the objective quality is calculated using the images reproduced on the display, while the subjective quality depends on the images perceived on the retina, whose scale is determined by the viewing distance. Unfortunately, the parametric forms of scoring functions generally lack an explicit dependence of this parameter.

Herein, we fill this gap in the literature by analyzing the role of the viewing distance in the case of *natural images*.

The main contributions of the paper are:

- We introduce a model of the Human Vision System (HVS) that accounts for pattern orientation selectivity and for 2D spatial frequency selectivity by means of a single, complex valued, Virtual Receptive Field (VRF) and accounts for the viewing distance.
- We propose a VRF based canonical IQA method for natural images in Gaussian blur, and show that any IQA metric can be decomposed into the cascade of a metric conversion rule and the canonical IQA method. The reason is that different types of degradations properly map onto perceptually equivalent level of normalized Gaussian blur.
- We demonstrate the performance improvement in the case of multiple viewing distances in the same dataset and the lower computational cost due to a low number of FLOPS and the need for only two parameters, providing comparisons with the state-of-the-art of both classic and deep learning based IQA methods.

The structure is organized as follows. In Section 2, an essential account of the employed HVS model is provided. In Section 3, the canonical IQA method for Gaussian blurred images is illustrated. In Section 4, it is shown how the canonical method can be used to linearize a-priori classical IQA methods for blur metrics. In Section 5, the extension to other image degradation types is discussed. In Section 6, the performance of the calibration-free methods are compared to the popular competitors on different databases. Implementation issues are discussed in Section 7. Finally, some remarks are outlined in Section 8 and conclusion is drawn in Section 9.

2. The virtual receptive field model of HVS

We present an accurate model useful for the evaluation of visual perception of blur in the case of natural images.

The *retina* in the foveal region is modeled as a distribution of light receptors whose position is individuated by the Cartesian coordinate pair $\mathbf{p} \equiv (x_1, x_2)$. The projection of an image on the retina generates a luminance component $I(\mathbf{p})$ (the chrominance components are ignored here). It is assumed that $I(\mathbf{p})$ is sampled by a grid of receptors whose average density is assumed as 60 samples per one visual degree. Since these receptors are not regularly placed on the retinal surface, invoking the generalized theory for non-baseband and nonuniform samples of [15], $I(\mathbf{p})$ is correctly represented provided that its 2D bandwidth does not exceed $(-30, +30)$ cycles/degree, i.e., $(-1/2, +1/2)$ cycles/arcmin. This assumed retinal resolution corresponds to the so-called *Snellen acuity*.

On the other hand, a display is formed by a rectangular array of pixels characterized by its *resolution* R expressed in pixel/mm, i.e., spaced by $d = \frac{1}{R}$ mm apart.

The image is projected onto the retina without loss of information and without redundancy if the spacing d matches the said 60/degree density of receptors on the retina. This occurs at a viewing distance δ_0 such that:

$$\delta_0 \times \tan(1) = \frac{1}{R} = d \tag{2}$$

where lengths are expressed in mm and angles in arcmins. This distance is referred here to as *nominal viewing distance*. In the ITU recommendations [16] this distance is referred to as the *Design Viewing Distance* (DVD), or *Optimal Viewing Distance* (OVD).

Specifically, for a display of height H (in mm) and characterized by L rows of square pixel *with reference to the center of the screen orthogonal to the line of sight*, we have (in mm):

$$\delta_0 = \frac{H}{L\tan(1)} = 3438 \times \frac{H}{L} \tag{3}$$

where 1 radian = 3438 arcminute.

At nominal viewing distance δ the spatial Fourier spectra of the images projected onto the retina are limited into the range $(-1/2, +1/2)$ cycles/arcmin or $(-30, +30)$ cycles/degree, according to the Nyquist criterion.

For a generic viewing distance δ, the spectrum of the retinal images lies within the band $\left(-\frac{1}{2}\frac{\delta}{\delta_0}, +\frac{1}{2}\frac{\delta}{\delta_0}\right)$ cycles/arcmin or $\left(-30\frac{\delta}{\delta_0}, +30\frac{\delta}{\delta_0}\right)$ cycles/degree both in horizontal and vertical directions. Thus, if the viewing distance is less than the nominal one, the bandwidth of the sensed image shrinks. If the distance exceeds the nominal one, the bandwidth broadens.

The receptive fields calculate a weighted sum of the luminance $I(\mathbf{p})$ in a neighborhood of $\mathbf{p}$, modeled by a spatial convolution, indicated by the symbol $*$, between $I(\mathbf{p})$ and the Virtual Receptive Field (VRF) $h(\mathbf{p})$:

$$\begin{aligned} y(\mathbf{p}) &= I(\mathbf{p}) * h(\mathbf{p}) \\ h(\mathbf{p}) &= Re\{h(\mathbf{p})\} + jIm\{h(\mathbf{p})\} \end{aligned} \tag{4}$$

The output image $y(\mathbf{p})$ will be referred to as *visual map* and $h(\mathbf{p})$ is the complex valued Point Spread Function, represented by the visual map of a single lighting point in the dark [17].

The role of the VRF model $h(\mathbf{p})$ is to extract the Positional Fisher Information (PFI) as a measure of the pattern localizability (see Luminance and Gradient layers of Fig. 2).

In the polar coordinates $r = \sqrt{x_1^2 + x_2^2}$ and $\varphi = tg^{-1}\frac{x_2}{x_1}$ the VRF model assumes the polar separable form:

$$h(r,\varphi) = \frac{r}{2\pi s_G^2} e^{-\frac{r^2}{2{s_G}^2}} e^{j\varphi} \tag{5}$$

where the parameter s_G is the *spread* of the VRF.

This VRF is a eigen-function with respect to 2D Fourier transformation [18], i.e., the Fourier transform of the VRF is the Virtual Neural Transfer Function:

$$H(\rho,\vartheta) = j2\pi\rho e^{j\vartheta} e^{-S_G^2\rho^2} \tag{6}$$

being $\rho = \sqrt{f_1^2 + f_2^2}$ and $\vartheta = tg^{-1}\frac{f_2}{f_1}$ the polar coordinates in the frequency domain.

According to the experimental data in [19], $H(\rho,\vartheta)$ is an accurate model of the Contrast Sensitivity Function (CSF) of the HVS [1], as shown in Fig. 1 (blue line).

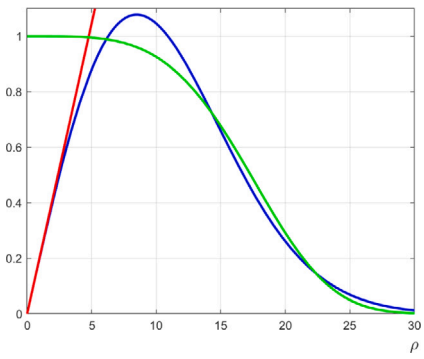

Fig. 1. The radial frequency response magnitude of the Virtual Neural Transfer Function $H(\rho,\vartheta)$ for $s_G = 2.5$ arcmins. The maximum is set at about 8.5 cycles/degree. It is the product of a linear component (red line) and a Gaussian-shaped component (green line).

The maximum along the radial frequency is set at about 8.5 cycles/degree. This does correspond to a spread $s_G = 2.5$ arcmin, and a contrast loss of ∼40 dB at the Nyquist frequency. The Virtual Neural Transfer Function is naturally interpreted as the cascade of two functions. The first one is the magnitude of $j2\pi\rho e^{j\vartheta}$ (red line in Fig. 1), that represents the *complex spatial gradient* operator [20] defined as

$$\nabla I(\mathbf{p}) \triangleq \frac{\partial I(x_1,x_2)}{\partial x_1} + j\frac{\partial I(x_1,x_2)}{\partial x_2} \tag{7}$$

This operator acts as an ideal edge extractor. The magnitude of its output $y(\mathbf{p})$ indicates the local edge strengths, while its phase indicates the edge orientations.

The second factor (green line in Fig. 1)

$$G(\rho,\vartheta) = e^{-S_G^2\rho^2} \tag{8}$$

represents a low-pass Gaussian-shaped filter acting as noise and aliasing suppressor. It is interpreted as the source of a *neural blur*.

Unlike models oriented to the study of visual acuity, employing specific stimuli localized both in space and in the spatial frequency domain, or even in both domains [21], the abstract, *functional model* of the receptive fields of the HVS presented here allows for a direct computation of the PFI. The model is oriented to the evaluation of the visual impact of blur in the vision of natural scenes. To this purpose, a generic image projected onto the retina is viewed as an element of the random set of natural images, characterized by stable statistical features. The model accounts for the *orientation selective behavior* and the spatial *frequency selective behavior* of the HVS according to the general principles of [22].

3. The canonical IQA method for a DMOS estimation on Gaussian blurred natural images

In this section, we introduce the canonical IQA method for a DMOS estimation only for natural images degraded by Gaussian blur.

Accurate pattern localization is a primary goal of living beings. We assume *as a principle* that the fine position of patterns in the observed images is determined by the HVS with the maximum allowable accuracy, given some macro-structural constraints. The maximum achievable accuracy when measuring the fine position of patterns in the background noise is deduced from Fisher Information on positional parameters (see the estimation theory in [23]): the Fisher Information inverse yields the minimum estimation variance.

As shown in Fig. 2 (Local averaging layer), the PFI of a *detail* of a pristine image, extracted by a window $w_{\mathbf{p}}(\mathbf{q})$, centered on $\mathbf{p}$ in presence of background white Gaussian noise with variance σ_V^2, is calculated in [24] as $\widetilde{\psi}(\mathbf{p}) = \frac{\widetilde{\lambda}(\mathbf{p})}{\sigma_V^2}$, where $\widetilde{\lambda}(\mathbf{p}) = \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 \left|\widetilde{y}(\mathbf{p}-\mathbf{q})\right|^2$ is the energy of the detail and

$$\widetilde{y}(\mathbf{p}) = \widetilde{I}_D\left(\mathbf{p}\frac{\delta}{\delta_0}\right) * h(\mathbf{p}) \tag{9}$$

is the visual map of a pristine (non-degraded) image on the display $\widetilde{I}_D(\mathbf{p})$ projected onto the retina ($\frac{\delta}{\delta_0}$ is the scale factor). Likewise, for the same detail of a degraded version of the image we have $\psi(\mathbf{p}) = \frac{\lambda(\mathbf{p})}{\sigma_V^2}$, where $\lambda(\mathbf{p}) = \sum_{\mathbf{q}} w_{\mathbf{p}}(\mathbf{q})^2 |y(\mathbf{p}-\mathbf{q})|^2$, and

$$y(\mathbf{p}) = \widetilde{I}_D\left(\mathbf{p}\frac{\delta}{\delta_0}\right) * b\left(\mathbf{p}\frac{\delta}{\delta_0}\right) * h(\mathbf{p}) \tag{10}$$

is the visual map of a blurred version of a pristine image on the display $\widetilde{I}_D(\mathbf{p}) * b(\mathbf{p})$ projected onto the retina.

Let us now calculate the PFI $\widetilde{\psi}$, ψ for the original and degraded images in the frequency domain. Here, a Gaussian blur applied to the observed image is described in polar frequency coordinates by the function

$$B(\rho,\vartheta) = e^{-s_B^2\rho^2} \tag{11}$$

where the parameter s_B will be referred to as the *spread* of the blur operator. Referring to the Fourier spectra of a generic detail, by the Parseval theorem we have[2]:

$$\psi(\mathbf{p}) = \frac{1}{\sigma_V^2}\int_0^{2\pi}\int_0^{+\infty}\left|D_{\mathbf{p}}(\rho)\right|^2 \rho^2 |G(\rho,\vartheta)|^2 |B(\rho,\vartheta)|^2 \rho\, d\rho d\vartheta$$

$$\widetilde{\psi}(\mathbf{p}) = \frac{1}{\sigma_V^2}\int_0^{2\pi}\int_0^{+\infty}\left|D_{\mathbf{p}}(\rho)\right|^2 \rho^2 |G(\rho,\vartheta)|^2 \rho\, d\rho d\vartheta$$

where $D_{\mathbf{p}}(\rho)$ is the energy spectrum of the detail $d_{\mathbf{p}}(\mathbf{q}) = w_{\mathbf{p}}(\mathbf{q})y(\mathbf{p}-\mathbf{q})$, $G(\rho,\vartheta)$ is defined in (8) and $B(\rho,\vartheta)$ in (11).

Let us now refer to *natural images*, modeled as elements of a random set whose ensemble radial spectral energy distribution decays as $\frac{1}{\rho^2}$, with the expected value [25]

$$E\left\{\left|D_{\mathbf{p}}(\rho,\vartheta)\right|^2\right\} = f(\vartheta)\frac{1}{\rho^2} \tag{12}$$

Therefore, for the random set of natural images, the expected value of the PFI in presence of blur is calculated as

$$\Psi = E\left[\psi(\mathbf{p})\right] = K\int_0^{+\infty}|G(\rho,\vartheta)|^2 |B(\rho,\vartheta)|^2 \rho\, d\rho \tag{13}$$

and, in the absence of blur,

$$\widetilde{\Psi} = E\left[\widetilde{\psi}(\mathbf{p})\right] = K\int_0^{+\infty}|G(\rho,\vartheta)|^2 \rho\, d\rho \tag{14}$$

where $K = \frac{1}{\sigma_V^2}\int_0^{2\pi} f(\vartheta)d\vartheta$.

From known results of integral calculus, we obtain the positional information of the blurred and pristine images in the closed form results:

$$\Psi = K\int_0^{+\infty} e^{-2(s_G^2+s_B^2)\rho^2}\rho\, d\rho = \frac{K}{4\left(s_G^2+s_B^2\right)} \tag{15}$$

$$\widetilde{\Psi} = K\int_0^{+\infty} e^{-2s_G^2\rho^2}\rho\, d\rho = \frac{K}{4s_G^2} \tag{16}$$

Now, leverage the following assumptions:

- The perceived increase of positional uncertainty is proportional to the objective one, considering the consistency of the position estimates in the 3D Euclidean space during the interaction of the subject with its surrounding world.
- The differential sensitivity to stimuli is inversely proportional to the size of the initial stimuli (Weber law [17]).

[2] It is assumed that the shape of the detail window is so smooth that it does not influence significantly its spectrum.

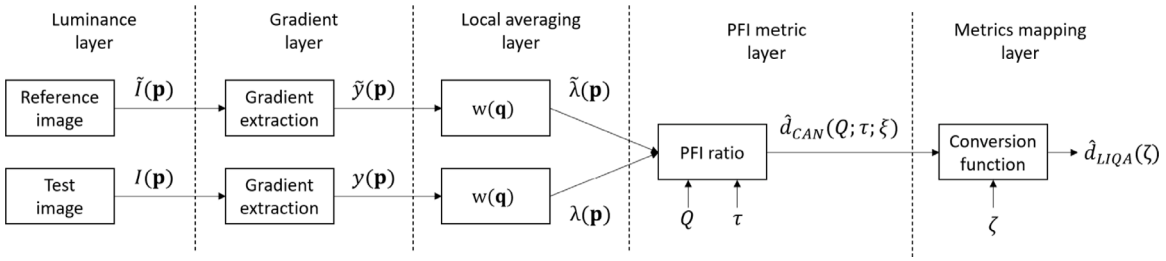

Fig. 2. The flow chart of the $\hat{d}_{LIQA}(\zeta)$ which shows the computational layers.

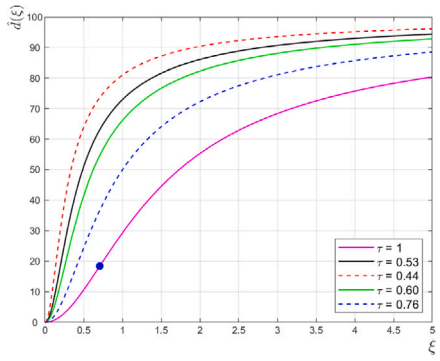

Fig. 3. The blur DMOS estimator $\hat{d}(\xi)$ versus the metric $\xi = s_B/s_G$. The blue dot indicates the point of maximum sensitivity of the HVS with respect to the variations in blur.

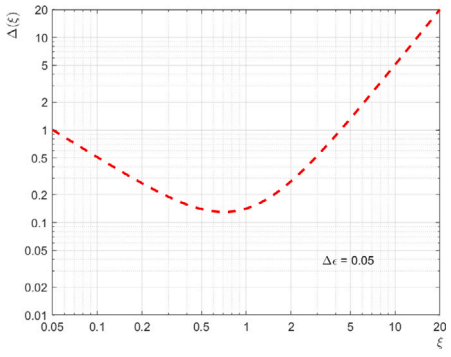

Fig. 4. The theoretical increment $\Delta(\xi)$ versus ξ for $\Delta\epsilon = 0.05$ in a log/log scale. The typical "dipper" shape is shown, with a minimum at $\xi = 0.707$, corresponding to $\hat{d}(\xi) = 18.4$.

The Gaussian blur increases the positional uncertainty. Therefore, the uncertainty increase $\epsilon(\xi) = \dfrac{\psi^{-1/2} - \tilde{\psi}^{-1/2}}{\psi^{-1/2}}$ provides a psychophysically motivated measure of the quality loss.

Introducing the *normalized blur* $\xi \triangleq \frac{s_B}{s_G}$, we have:

$$\epsilon(\xi) = \frac{\sqrt{s_G^2 + s_B^2} - \sqrt{s_G^2}}{\sqrt{s_G^2 + s_B^2}} = 1 - \sqrt{\frac{1}{1+\xi^2}}. \tag{17}$$

Hence, the percentage positional uncertainty increase

$$\hat{d}(\xi) = 100\left[1 - \sqrt{\frac{1}{1+\xi^2}}\right] \tag{18}$$

provides a theoretically grounded DMOS estimator.

In Fig. 3, plotting $\hat{d}(\xi)$, we see the magenta curve at the nominal viewing distance and we recognize that it exhibits a characteristic sigmoidal behavior, with a threshold for small input values and a saturation for large input values. The rating function complies with the typical "dipper" shape [26,27] of the incremental blur producing a Just Noticeable Difference (JND) of the perceived quality. The sensitivity with respect to the normalized spread ξ is calculated as

$$\frac{d\epsilon(\xi)}{d\xi} = \xi\left[\frac{1}{1+\xi^2}\right]^{\frac{3}{2}}; \tag{19}$$

and the increment $\Delta\xi$ necessary to produce a given increment $\Delta\epsilon(\xi)$ is approximated as:

$$\Delta\xi = \frac{1}{\xi}\left[1+\xi^2\right]^{\frac{3}{2}}\Delta\epsilon(\xi). \tag{20}$$

In Fig. 4, the typical "dipper" shape is shown. Specifically, its theoretical minimum occurs at $\xi = \sqrt{\frac{1}{2}}$, corresponding to $\hat{d}(\xi) = 18.4$ (see Figs. 3 and 4).

In addition to the physical parameter related to the objective viewing distance, to fit different dataset characteristics, we introduce a gain factor Q and set it to

$$Q = \frac{d_A}{100}\left[1 - \sqrt{\frac{1}{1+\xi_A}}\right]^{-1}, \tag{21}$$

where d_A is the DMOS of the *worst quality* image of the dataset (anchor image). Depending on the experiments, d_A may vary from 80 to 100 DMOS.

The above estimates refer to the nominal viewing distance, i.e., to the case where $\delta = \delta_0$. Accounting for the scaling of the retinal image due to the actual viewing distance, let us define the *normalized viewing distance* $\tau \triangleq \frac{\delta}{\delta_0}$ for notational simplicity. We assume that the spread of the VRF s_G is proportional to the viewing distance and the spread of the blur operator s_B is inversely proportional to the viewing distance. Thus, replacing $s'_G = \tau \cdot s_G$ and $s'_B = \frac{1}{\tau} \cdot s_B$ in (15) and (16), we modified the (17) as $\epsilon(\xi;\tau) = 1 - \sqrt{\frac{1}{1+\frac{\xi^2}{\tau^4}}}$.

Thus, we finally obtain the following theoretical formula for the prediction of the DMOS of *blurred natural images*:

$$\hat{d}_{CAN}(Q;\tau;\xi) = 100 \times Q \times \left(1 - \frac{1}{\sqrt{1+\frac{\xi^2}{\tau^4}}}\right) \tag{22}$$

referred to as *canonical estimate* of the DMOS of Gaussian blurred natural images. It is characterized by the *scoring gain* parameter Q and the dimensionless parameter τ (see PFI metric layer of Fig. 2). In Fig. 3, The model is shown at different viewing distances of experiments known in the literature and, in Fig. 5, it is possible to observe the variable behavior of the model under saturation, dependent on the image with the highest DMOS (worst quality).

Fig. 6 shows the canonical estimate of DMOS as a function of τ and ξ. It can be observed how subjective quality improves with increasing viewing distance and decreasing distortion. Consider, for instance, the subjective perception of blur: for a fixed value of ξ, as we move

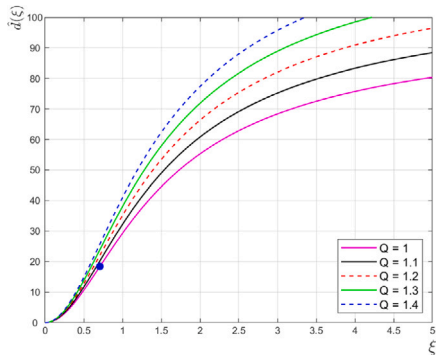

Fig. 5. The DMOS estimator $\hat{d}(\xi)$ versus the metric ξ (in centesimal units DMOS) at different values of Q.

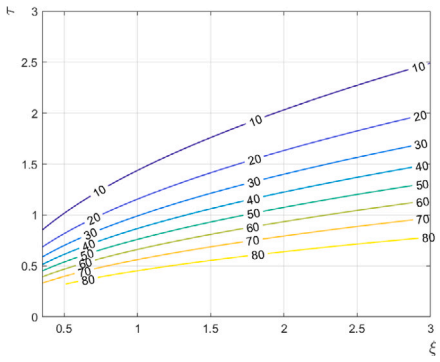

Fig. 6. The canonical DMOS estimator $\hat{d}(\xi)$ (in centesimal units DMOS) versus the metrics ξ and τ as a contour plot at the fixed value Q = 1.

Fig. 7. The specimen image "Iceroad" extracted from the database LIVE MD.

away from the image, subjective quality increases; furthermore, if we fix a viewing distance τ, subjective quality worsens with increasing distortion.

At this point, it is possible to use the canonical model to choose a typical image (image representative of the average perception of subjective quality for each blur level). We selected the specimen within the LIVE MD collection employed for statistical verification, whose images are *not shared by the other collections*. Fig. 7 shows the selected specimen image "Iceroad". In Fig. 8, it is shown the canonical model (continuous green line) and how the typical image (filled circles) is representative of an average image trend following the canonical model, for each blur level in the LIVE MD database. We obtained results with comparable performance using typical images from other databases.

4. From the canonical IQA method to general IQA methods

In the previous section, we proposed the canonical method for estimating the DMOS $\hat{d}_{CAN}$ of a Gaussian blurred natural image as a

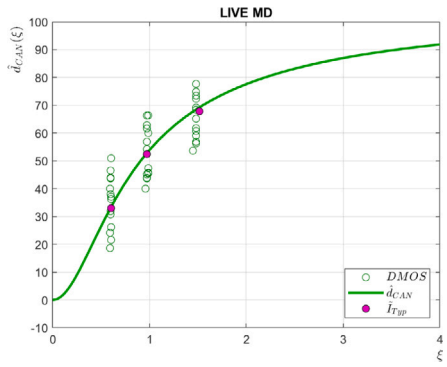

Fig. 8. The canonical DMOS estimator $\hat{d}(\xi)$ (green curve), and three blur levels of the specimen image (red points) for the LIVE MD database.

function of the blur IQA metric ξ. We now demonstrate how any IQA metric can be decomposed into a combination of a metric conversion rule and the canonical IQA method.

Let us consider now a generic IQA method based on a scalar metric ζ. We assume that ζ is *monotonically* related with the degree of Gaussian blur applied to images, i.e., with the canonical metric ξ: as blur increases, visual discomfort increases, just as it does with noise or jpeg distortion.

The method is founded on the conversion of the values of the metric ζ utilized by a generic IQA method into the corresponding value of the metric ξ (normalized blur) utilized in the canonical method, in the case of typical natural images. For these typical images, it is assumed that the canonical method and the generic IQA method do provide the same estimates $\hat{d}_{CAN}(\xi;\tau)$ and $\hat{d}_{IQA}(\zeta;\tau)$ respectively:

$$\hat{d}_{CAN}(\xi;\tau) = \hat{d}_{IQA}(\zeta;\tau) \tag{23}$$

This equality is obtained by searching for the function $\xi(\zeta;\tau)$. The function $\hat{d}_{CAN}(\xi;\tau)$ is available in an analytical closed form (22), whereas $\hat{d}_{IQA}(\zeta;\tau)$ is numerically calculable for a specific image. $\hat{d}_{IQA}(\zeta;\tau)$ are the empirical values of a chosen estimator (e.g., the Visual Information method (VIF) [5]) as the level of blur increases on a reference image, as shown in Section 4.1.

4.1. Building the conversion function

To solve for the conversion function $\xi(\zeta;\tau)$ we adopted a semi-numerical method, resorting to the *specimen* image $\tilde{I}_{Typ}$ sufficiently representative of natural images, with respect to *image content*, another factor affecting perceived quality loss in addition to Gaussian blur and viewing distance.

A suite of fifty Gaussian blurred versions of the specimen image spanning the whole DMOS range was generated from the reference pristine image, using a Gaussian filter with an incremental value of blur s_B. Thus, it is possible to calculate the value of the IQA metrics for each pair of reference and blurred images and obtain a conversion function to map the ζ values to $\xi = f(\zeta;\tau)$.

Based on these samples, to obtain continuous functions in ζ, the functions $f(\zeta;\tau)$ were defined using a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) Matlab® function. We remark that (i) the model is not trained on the data, (ii) preserves the claimed calibration-free nature, since the functions $f(\zeta;\tau)$ are fixed once for all.

Fig. 9 shows the functions $f(\zeta;\tau)\big|_{\tau=0.76}$ obtained on the specimen image "Iceroad" for the IQA metrics from the Visual Information method (VIF) [5], Multi-Scale Structural Similarity (MSSIM) [28], the

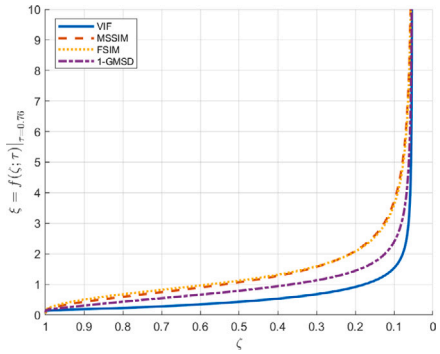

Fig. 9. The metric conversion function for the considered IQA methods. The scales of the metrics of the MSSIM, FSIM and GMSD, for $\tau = 0.76$, are purposely modified and aligned to the scale of the VIF metric to put into evidence their common asymptotic behavior.

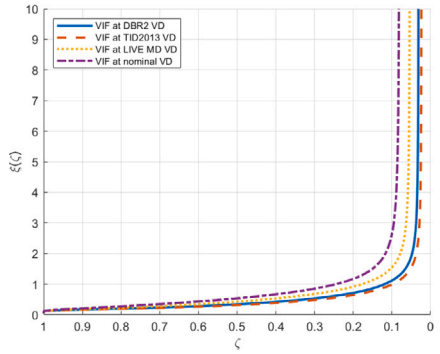

Fig. 10. The metric conversion function for the VIF method at different viewing distances: $\tau = 0.53$, $\tau = 0.44$, $\tau = 0.60$, and $\tau = 0.76$, corresponding respectively to the LIVE DBR2, TID2013, CSIQ and LIVE MD viewing distances, and $\tau = 1$ for the nominal viewing distance.

Feature Similarity Index (FSIM) [3] and the Gradient Magnitude Similarity Deviation (GMSD) [29].[3] The databases presented are useful only for experimental verification of the theoretical model.

Let us observe that the fundamental behavior of the functions $\xi = f(\zeta;\tau)$ is analogous for all the considered IQA metrics ζ. It compresses the equivalent blur values for highly degraded images and conversely expands them for slightly degraded images. In Fig. 10, it is shown how the functions change with the viewing distance in the VIF case.

4.2. Mapping from ζ to ξ

It is possible to map each value of the IQA estimates (for instance, the VIF calculated on each pair of reference-degraded images) with the conversion function and obtain the equivalent blur values. In Fig. 11, VIF values for blurred-only images of DBR2 are mapped to the conversion function to obtain blur equivalent values.

Once mapped onto the equivalent Gaussian blur axis $\xi = f(\zeta;\tau)$, a linear estimate $\hat{d}(\zeta)$ of the DMOS of blurred natural images, is given by the canonical rating function

$$\hat{d}(Q;\tau;\zeta) = 100 \times Q \times \left(1 - \frac{1}{\sqrt{1 + \frac{\xi(\zeta;\tau)^2}{\tau^4}}}\right). \tag{24}$$

[3] The whole set of functions is available to the reader in the supplementary data.

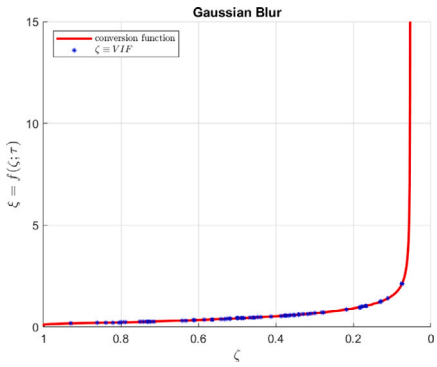

Fig. 11. The VIF values of blurred images (blue stars) of DBR2 mapped on the conversion function (continuous red line).

The metric in (24) represents an innovative model for accurately assessing the visual discomfort induced by the blurring of natural scenes. This 2-parameter a-priori linearized method, in which the viewing distance parameter plays a dominant role, is presented as an alternative to the 5-parameter VQEG model.

4.3. Extension of the method beyond Gaussian blur

The analysis conducted refer to the case of Gaussian blurred natural images, is of mainly theoretical interest. We here extend the analysis to images subject to different types of degradation, which are neatly distinct with respect to blur from a perceptual viewpoint [30].

IQA metrics are purposely designed to fit the subjective DMOS under different type of degradation [31]. This aspect is explicitly discussed in [5] where the concept of perceptual quality equivalence for some degradation viewed as mixtures of noise and blur is introduced. In [7] two metrics are jointly employed in a two-dimensional scoring function. Therein, it was outlined the distinct perception of detail loss due to blur and spurious details due to noise, and the perceptual quality equivalence among blur and noise is assumed as constant. See also [32], where a bijective mapping between the 2-D noise/blur space and the 3-D MSSIM space was defined. Therefore, the conversion function $\xi(\zeta;\tau)$ applies even if the values of the metric ζ are not only determined by the presence of blur, but by other factors that produce the same values of $\hat{d}_{CAN}(\xi) = \hat{d}_{IQA}(\zeta)$. For instance, in the VIF method the metric is computed as the contribution of many terms measuring losses of Shannon information, where the effect of noise is equivalent to blur.

Following these considerations, the method obtained in this way can be directly extended from blur to other distortions and is herein referred to as *Linearized IQA* (LIQA), whose estimates have the same form as the (24): $\hat{d}_{LIQA}(\zeta) \equiv \hat{d}(\zeta)$. The overall flow chart of the $\hat{d}_{LIQA}(\zeta)$ is shown in Fig. 2.

5. Performance evaluation

The proposed VRF-based LIQA method is calibration-free and it allows comparison of MOS/DMOS estimates across different applications and viewing distances.

To assess its performances, we present here a statistical performance comparison of some popular classical IQA methods already considered above, subject to final calibration according to VQEG recommendations, with their LIQA counterparts, which do not make use of empirical calibration. For the sake of completeness, we also present a comparison with the most advanced deep IQA methods and evaluate the performance on the same datasets in terms of linearity and absolute error, as well as computational cost.

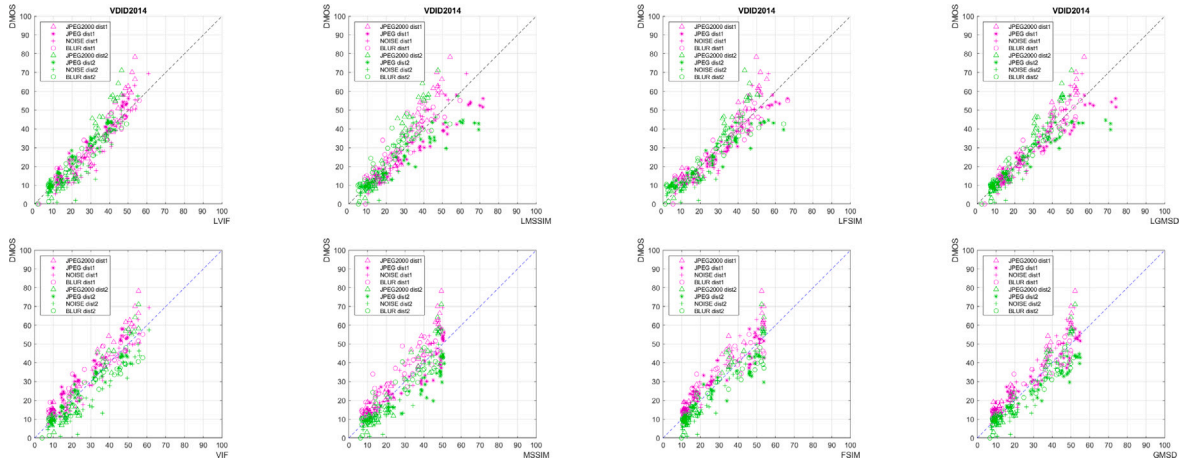

Fig. 12. The DMOS scatterplots versus the values predicted using respectively the theoretically **linearized IQA** method (upper row) and the conventional, empirically calibrated **IQA** method (lower row) for blur, noise, JPEG, JPEG2000 distortions for **VDID2014**. The first quartet of distorsions represents the data of the first session of the experiment (at τ_1, in magenta) and the second quartet represents the data of the second session (at τ_2, in green).

Table 1
Experimental verification for the **a-priori linearized IQA metrics** (left table) and **empirically calibrated IQA metrics** (right table) for VDID2014. Four distortions subsets: blur, noise, JPEG, JPEG2000. The viewing distance is $\tau_1 = 0.57$ for the first session of the experiment and $\tau_2 = 0.65$ for the second session.

A-priori linearized IQA metric	VDID2014			Empirically calibrated IQA metric	VDID2014		
	RMSE	SROCC	LCC		RMSE	SROCC	LCC
LVIF	6.1059	0.94482	0.94103	VIF	6.6367	0.91774	0.92393
LMSSIM	8.3164	0.92540	0.87787	MSSIM	7.7583	0.90094	0.89442
LFSIM	6.5765	0.94858	0.92554	FSIM	6.9776	0.92592	0.91577
LGMSD	6.5816	0.95366	0.92533	GMSD	6.8077	0.92717	0.91978

In the Multi-Scale Structural Similarity (MSSIM) [28] and the Feature Similarity Index (FSIM) [3], the dependence of their scoring function on the viewing distance is attenuated by adjusting a mix of metrics calculated at different scales.

The present comparative analysis focuses degradation types that appeared suitable for conversion into the canonical metric, namely, additive noise, JPEG and JPEG2000 compression, including some examples of *multiple degradations* [33].

To this purpose, VDID2014 database [14] is presented here to show the superior performance of the a-priori linearized IQA methods in the case of multiple distances in the same database. A single stimulus based judgments was adopted. Subjects were asked to provide their overall sensation of quality on a continuous quality scale from 0 to 1 during two consecutive sessions with two different viewing distances. In the original experiment, a preprocessing stage is applied to different IQA methods. Images are rescaled to emulate different viewing distances, and the corresponding metrics are combined using the criterion of optimizing the overall statistical performance. Here, the performance of the canonical method applied to raw data is shown.

Moreover, four different well-known collections of examples have been employed, namely the above indicated subsets of the LIVE DBR2, the TID2013, the CISQ and the LIVE MD databases. These databases are built using independent protocols and methods for calculating the DMOS, which is used here as a common measure of quality loss.

Specifically, in the LIVE DBR2 experiments used single stimulus based judgments, including reference images [34]. Subjects were asked to express a score among five quality levels with a slider.

In the TID2013 experiments, a tristimulus methodology [35] was adopted, where subjects simply selected the best between two degraded images in the presence of the reference image.

In the CSIQ experiments, subjects were asked to order the quality of four images simultaneously displayed on an array of monitors [36].

In the LIVE MD, a single stimulus [34] with hidden reference methodology was adopted.

These basic differences, along with diverse viewing conditions and protocols, ensure that the selected series of measurements, if they yield coherent results, will reflect the essential nature of the subjective quality rating phenomenon, independent of the way it was observed and measured.

For multiple viewing distance databases, as VDID2014, the DMOS estimates based on canonical model are slightly better than realignment with the logistic function, because it is possible to realign the blur values of different viewing distances independently, as shown in the scatterplots of empirical DMOS values versus the estimated ones of Fig. 12. An exception is the MSSIM, which, however, exhibits strong overfitting for images at high DMOS. The results of this comparative analysis are reported in Table 1. The selected statistical quality indices were the RMSE (indicative of the average distance between the actual DMOS values and the ones provided by the method), the SROCC (Spearman Rank Correlation Coefficient) which reveals the monotonicity between predicted and actual DMOS values, and the PLCC (Pearson Linear Correlation Coefficient) which measures the linearity of the DMOS mapping. Here, the PLCC and the SROCC indices improve by indicating higher linearity of L-estimators compared with classical IQA estimators.

We present a unique large database, called SUPERQUARTET, as the union of LIVE DBR2, TID2013 and CSIQ. We can work on data from different experiments as a single dataset. Fig. 13 shows the scatterplot of the four distortions present in each database (blur, noise, JPEG, JPEG2000) and Table 2 presents the performances of the a-priori linearized IQA metrics.

For single viewing distance databases, the scatterplots of empirical DMOS values versus the estimated ones are shown in Figs. 14–17, for direct visual comparison. The comparative analysis are reported in Table 3 (left table) for the calibration-free LIQA methods, and in the right table for the IQA methods after empirical calibration [6,10]. These tables indicate that the DMOS estimates based on theoretical modeling

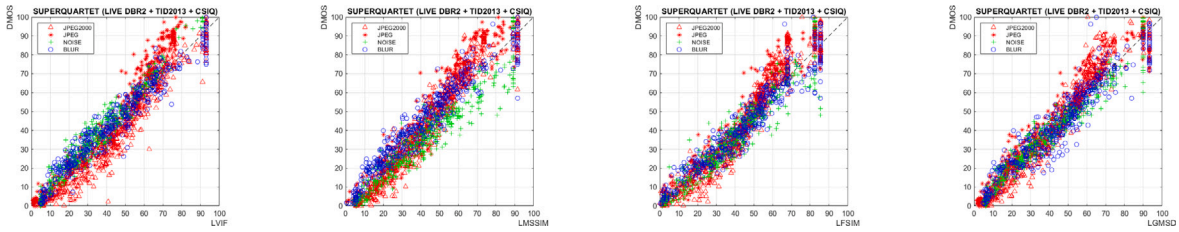

Fig. 13. The DMOS scatterplots versus the values predicted using the theoretically **linearized IQA** methods for blur, noise, JPEG, JPEG2000 distortions for SUPERQUARTET (LIVE DBR2 + TID2013 + CSIQ).

Table 2
Experimental verification for the **a-priori linearized IQA metrics** (left table) and **empirically calibrated IQA metrics** (right table) for SUPERQUARTET. Four distortions subsets: blur, noise, JPEG, JPEG2000.

A-priori linearized IQA metric	SUPERQUARTET			Empirically calibrated IQA metric	SUPERQUARTET		
	RMSE	SROCC	LCC		RMSE	SROCC	LCC
LVIF	7.8439	0.96006	0.95513	VIF	8.6157	0.94274	0.94356
LMSSIM	8.7550	0.95374	0.94260	MSSIM	9.2259	0.93997	0.93499
LFSIM	8.1944	0.96659	0.94999	FSIM	8.0509	0.95631	0.95090
LGMSD	7.2696	0.96956	0.96039	GMSD	7.6069	0.95899	0.95629

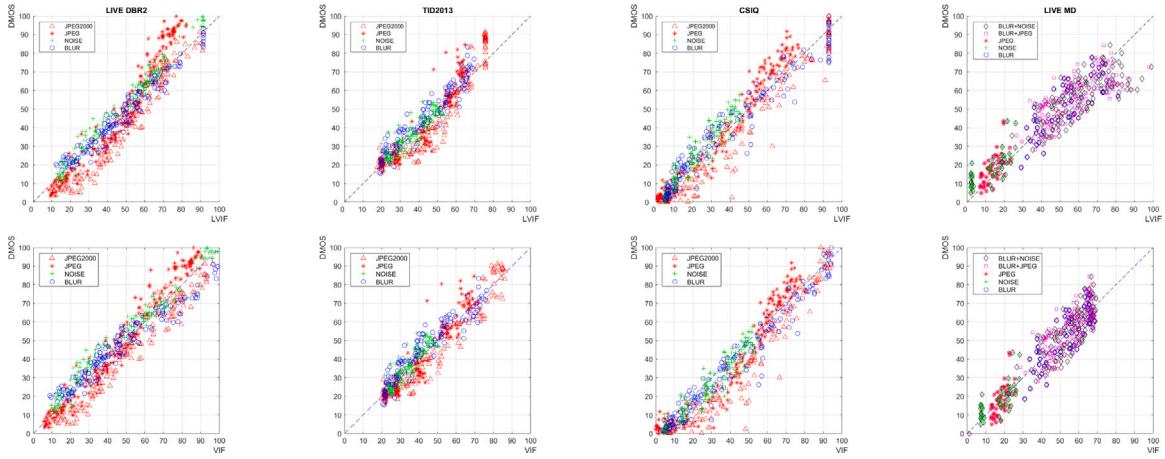

Fig. 14. The DMOS scatterplots versus the values predicted using respectively the theoretically **linearized VIF** method (upper row) and the conventional, empirically calibrated **VIF** method (lower row) for blur, noise, JPEG, JPEG2000 distortions for LIVE DBR2, TID2013 and CSIQ, and blur, noise, JPEG, blur+JPEG, blur+noise distortions for LIVE MD.

are statistically slightly more dispersed than the ones determined using empirical fitting to specific data, using the five parameter curve of (1). Fig. 18 shows the RMSE boxplots for each database, comparing each L-estimator to its empirically calibrated version. However, this advantage does depend on the adopted fitting model. For instance, the three-parameter curve of [37] would exhibit a larger RMSE in change of a better stability across applications.

In general, it is observed that the mutual positions of the clusters regarding the different degradations for the IQA methods tend to be preserved by the corresponding LIQA versions. Furthermore, a slightly more visible deviation from linearity of the LIQAs for the TID2013 dataset can be noted. This could be explained by the fact that the viewing distance in these experiments is not controlled, but left to the choice of the observers [35], whereas here it is determined by regression all over the subset of blurred image data. Finally, a numerical saturation occurs towards the worst quality, in correspondence of the vertical asymptotes of the conversion curve $\xi(\zeta;\tau)$, resulting from the conversion function used by the model (see Fig. 10).

In Table 4, the 95th percentile (95p) of residual magnitudes and the residual kurtosis provide further insight about the fitting error properties. The 95p values and kurtosis do not indicate troubling issues or clear differences in the use of the metrics, with an acceptable $95p \approx 2 \cdot RMSE$ and a moderate non Gaussianity of the residuals.

6. LIQA vs. DL-IQA: A performance analysis

Deep Learning IQA (DL-IQA) models can offer significant advantages due to their ability to learn more complex representations and better capture the nuances of human visual perception. Classical metrics like PSNR, SSIM, and MS-SSIM are well-studied and optimized for rapid calculation. This is crucial for real-time applications such as video streaming, where low latency is essential to ensure a good user experience. Moreover, the implementation of these metrics requires fewer computational resources compared to deep learning models, which often need specialized hardware like GPUs to run efficiently. When dealing with a massive amount of video data and the need to continuously assess the quality of these videos in real-time, classical metrics allow for greater scalability with limited computational resources.

We select two of our most performant estimators, LVIF and LGMSD, and compare them with some well-known deep learning-based estimators in the Full Reference scenario.

The TOPIQ-FR [38] employs a top-down approach for image quality assessment, directing the network to focus on locally distorted regions of semantic importance. It utilizes a CFANet that leverages multi-scale features and cross-attention mechanisms. This method builds upon ResNet50 and enhances performance by emphasizing semantically

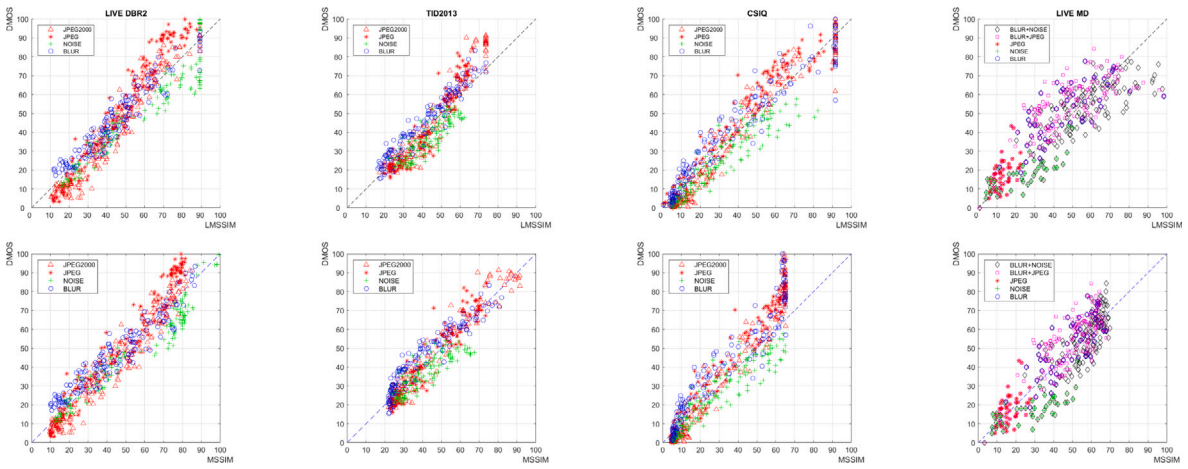

Fig. 15. The DMOS scatterplots versus the values predicted using respectively the theoretically **linearized MSSIM** method (upper row) and the conventional, empirically calibrated **MSSIM** method (lower row) for blur, noise, JPEG, JPEG2000 distortions for LIVE DBR2, TID2013 and CSIQ, and blur, noise, JPEG, blur+JPEG, blur+noise distortions for LIVE MD.

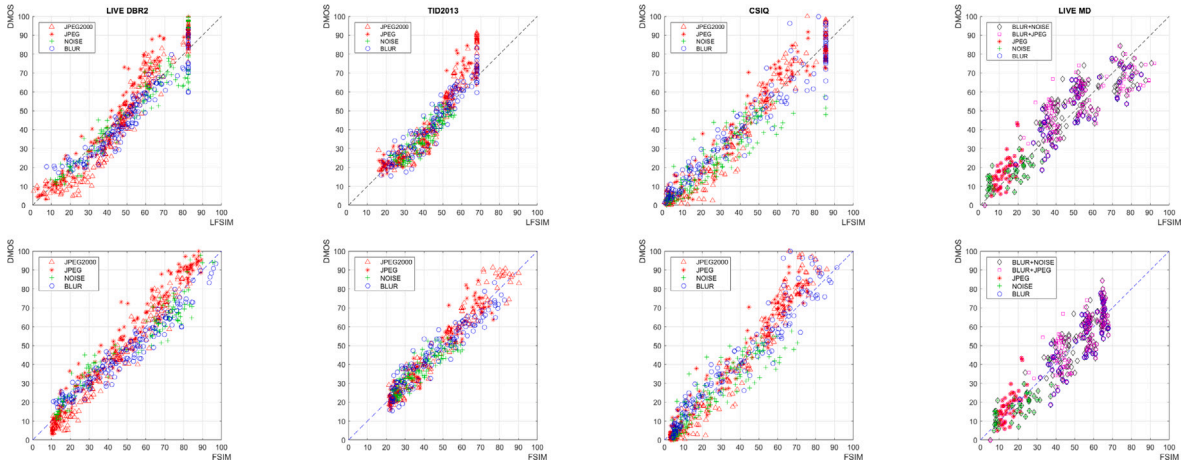

Fig. 16. The DMOS scatterplots versus the values predicted using respectively the theoretically **linearized FSIM** method (upper row) and the conventional, empirically calibrated **FSIM** method (lower row) for blur, noise, JPEG, JPEG2000 distortions for LIVE DBR2, TID2013 and CSIQ, and blur, noise, JPEG, blur+JPEG, blur+noise distortions for LIVE MD.

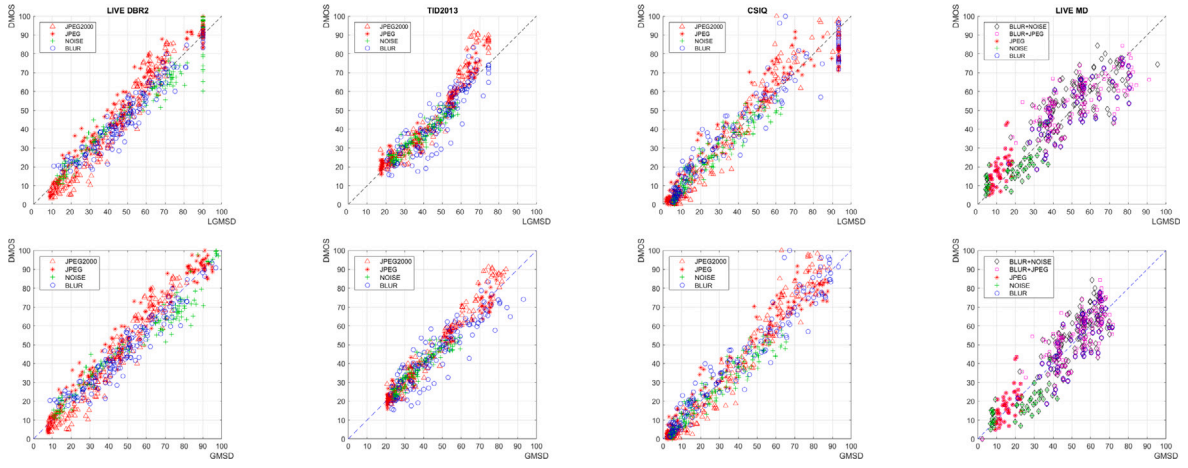

Fig. 17. The DMOS scatterplots versus the values predicted using respectively the theoretically **linearized GMSD** method (upper row) and the conventional, empirically calibrated **GMSD** method (lower row) for blur, noise, JPEG, JPEG2000 distortions for LIVE DBR2, TID2013 and CSIQ, and blur, noise, JPEG, blur+JPEG, blur+noise distortions for LIVE MD.

active regions, making it more efficient compared to state-of-the-art methods. TOPIQ-FR utilizes more than 20 parameters in its evaluation process. We also present its variant TOPIQ-FR-PIPAL trained on the PIPAL dataset (Perceptual Image Processing ALgorithms), specifically designed for image quality assessment, containing a wide range of distortions and algorithmic corrections [39].

The DISTS model [40] is a full-reference image quality model that uses a convolutional neural network to compare “texture similarity”

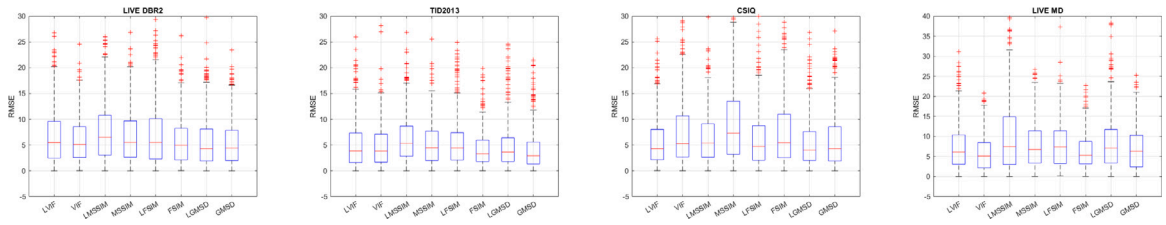

Fig. 18. The boxplot of the RMSE using respectively the theoretically **linearized IQA** methods and the conventional, empirically calibrated **IQA** methods.

Table 3
Experimental verification for the **a-priori linearized IQA metrics** (left table) and **empirically calibrated IQA metrics** (right table) for LIVE DBR2, TID2013, CSIQ and LIVE MD. Four distortions subsets for LIVE DBR2, TID2013 and CSIQ: blur, noise, JPEG, JPEG2000. Five distortions subsets for LIVE MD: blur, noise, JPEG, blur+JPEG, blur+noise. The viewing distances employed in the different databases were inferred by regression using the quality estimates for the subsets of blurred images, giving the following normalized values: $\tau = 0.53$ for LIVE DBR2, $\tau = 0.44$ for TID2013, $\tau = 0.60$ for CSIQ, and $\tau = 0.76$ for LIVE MD.

A-priori linearized IQA metric	Dataset	RMSE	SROCC	LCC	Empirically calibrated IQA metric	Dataset	RMSE	SROCC	LCC
LVIF	LIVE DBR2	8.6653	0.96536	0.95636	VIF	LIVE DBR2	7.3992	0.96535	0.96181
	TID2013	7.0170	0.92782	0.92789		TID2013	6.3344	0.92782	0.94021
	CSIQ	7.5789	0.95878	0.96336		CSIQ	9.5763	0.92283	0.93089
	LIVE MD	9.2065	0.92673	0.93604		LIVE MD	7.3096	0.92199	0.94277
LMSSIM	LIVE DBR2	9.5823	0.95323	0.93910	MSSIM	LIVE DBR2	8.4663	0.95307	0.94969
	TID2013	7.8579	0.91745	0.91394		TID2013	6.7297	0.91745	0.93224
	CSIQ	8.5463	0.95445	0.95326		CSIQ	13.1528	0.88089	0.86502
	LIVE MD	13.8323	0.87709	0.84556		LIVE MD	10.0285	0.87118	0.88922
LFSIM	LIVE DBR2	8.8302	0.96851	0.95149	FSIM	LIVE DBR2	7.2289	0.96846	0.96358
	TID2013	7.3283	0.95341	0.93291		TID2013	5.4113	0.95343	0.95673
	CSIQ	8.1802	0.96164	0.95810		CSIQ	10.5632	0.92742	0.91522
	LIVE MD	12.0365	0.91389	0.88788		LIVE MD	7.9413	0.90831	0.93207
LGMSD	LIVE DBR2	7.2629	0.96888	0.96478	GMSD	LIVE DBR2	6.9278	0.96888	0.96660
	TID2013	6.7091	0.95059	0.93579		TID2013	5.5523	0.95059	0.95440
	CSIQ	7.7122	0.96909	0.96232		CSIQ	8.1502	0.95490	0.95044
	LIVE MD	12.9031	0.90144	0.87629		LIVE MD	8.7612	0.89582	0.91666

Table 4
95p and Residual Kurtosis of the **a-priori linearized IQA metrics** (left table) and **empirically calibrated IQA metrics** (right table) for LIVE DBR2, TID2013, CSIQ, LIVE MD and VDID2014. Four distortions subsets for LIVE DBR2, TID2013, CSIQ and VDID2014: blur, noise, JPEG, JPEG2000. Five distortions subsets for LIVE MD: blur, noise, JPEG, blur+JPEG, blur+noise.

A-priori linearized IQA metric	Dataset	95p	Kurtosis	Empirically calibrated IQA metric	Dataset	95p	Kurtosis
LVIF	LIVE DBR2	16.99	3.17	VIF	LIVE DBR2	11.48	2.75
	TID2013	12.30	3.80		TID2013	9.65	4.21
	CSIQ	11.71	4.83		CSIQ	17.54	3.61
	LIVE MD	13.09	3.31		LIVE MD	13.07	2.63
	VDID2014	11.05	4.90		VDID2014	10.93	3.41
LMSSIM	LIVE DBR2	17.29	2.91	MSSIM	LIVE DBR2	14.90	2.93
	TID2013	13.58	2.95		TID2013	11.00	3.31
	CSIQ	14.14	3.96		CSIQ	19.40	3.83
	LIVE MD	21.03	4.30		LIVE MD	16.59	2.82
	VDID2014	15.85	3.99		VDID2014	12.67	3.71
LFSIM	LIVE DBR2	16.56	3.33	FSIM	LIVE DBR2	12.01	3.08
	TID2013	15.63	4.14		TID2013	9.72	3.67
	CSIQ	13.06	4.48		CSIQ	16.54	4.11
	LIVE MD	15.73	6.05		LIVE MD	12.79	2.56
	VDID2014	10.73	5.22		VDID2014	11.87	3.44
LGMSD	LIVE DBR2	12.58	3.64	GMSD	LIVE DBR2	10.88	3.29
	TID2013	12.24	4.80		TID2013	9.44	4.92
	CSIQ	13.76	5.68		CSIQ	13.26	4.66
	LIVE MD	15.40	5.71		LIVE MD	13.17	2.66
	VDID2014	11.26	6.47		VDID2014	10.64	3.48

and "structure similarity" between an original and a degraded image, optimizing the parameters to align with human image quality assessments. The model is designed to be insensitive to geometric transformations and has shown good performance in tasks such as texture classification and retrieval. DISTS employs more than 10 parameters, including two key parameters, α and β, which are structural and textural similarity parameters: α controls the balance between structural and textural similarities, and β normalizes the overall similarity.

Finally, we compare the performance with the LPIPS method [41], both in its base version using the AlexNet network, and in the LPIPS-VGG variant, utilizing the VGG (Visual Geometry Group) neural network for feature extraction. This benefits from VGG's ability to capture fine details due to its greater depth and complexity. VGG is often chosen because it balances depth and generalization capability effectively. PIEAPP [42] is also presented.

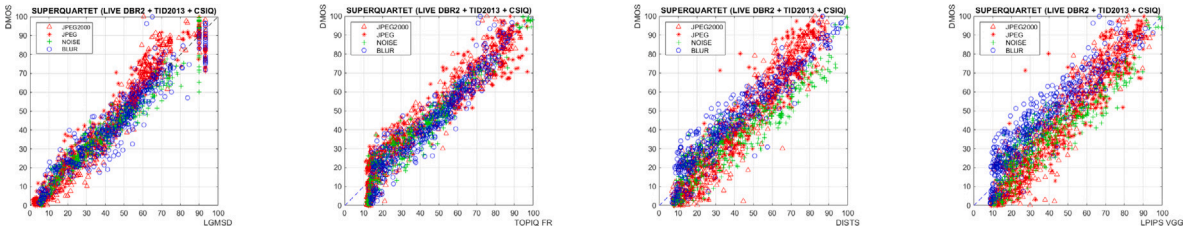

Fig. 19. The DMOS scatterplots versus the values predicted using the LGMSD, TOPIQ, DISTS and LPIPS methods for blur, noise, JPEG, JPEG2000 distortions for SUPERQUARTET (LIVE DBR2 + TID2013 + CSIQ). This dataset addresses the issue of varying viewing distances.

Table 5
Experimental verification of the **a-priori linearized IQA metrics** versus the **DL-IQA methods** for the SUPERQUARTET dataset. Four distortions subsets: blur, noise, JPEG, JPEG2000. The LIQA method LGMSD outperforms classical and deep learning based methods.

IQA metric	SUPERQUARTET		
	RMSE	SROCC	LCC
LVIF	7.8439	0.96006	0.95513
LGMSD	**7.2696**	**0.96956**	**0.96039**
TOPIQ FR	7.4752	0.96124	0.95782
TOPIQ FR PIPAL	10.1284	0.91518	0.92109
DISTS	9.5273	0.93066	0.93052
LPIPS	11.1168	0.91627	0.90408
LPIPS VGG	10.4791	0.90722	0.91527
PIEAPP	11.9807	0.90683	0.88764

For a fair performance comparison, the data obtained from the methods presented have all been aligned following VQEG guidelines [6, 10] (see (1)).

In Fig. 19, we show the TOPIQ-FR, DISTS, and LPIPS-VGG methods for the SUPERQUARTET dataset, and in Table 5, the performance of these and other DL-IQA methods is presented. We observe that, in the case of datasets with multiple viewing distances, LGMSD outperforms the DL-IQA methods.

In Fig. 20, we show the TOPIQ-FR, DISTS, and LPIPS-VGG methods for the LIVE DBR2, TID2013 and CSIQ datasets, for Gaussian blur, noise, JPEG and JPEG2000 distorsions. In Table 6, we show the performance of the different DL-IQA methods compared to our linearized LIQA ones. In the case of a single viewing distance, TOPIQ-FR outperforms, but LGMSD takes second place. DL-IQA methods employ millions of parameters, whereas the proposed LIQA method utilizes only two parameters.

Finally, for each dataset complete with all distortions, in Table 7, we present the linearity and rank ordering metrics. In blue bold, we highlight the best performances of the estimators and in black bold the second best. We note that in most cases, the LIQA estimators (last two columns on the right) exceed or equal the performance of the best DL-IQA, the TOPIQ FR (first column on the left).

With LIQA estimators, we only have computational cost calculations related to spline and canonical model on blur images. The piecewise polynomial is of degree d and N is the number of blur images to interpolate.

$$\text{FLOPS}_{\text{tot}} = \text{FLOPS}_{\text{spline}} + \text{FLOPS}_{\text{canonical}} =$$
$$2 \cdot d \cdot \text{N} + 8 \cdot \text{N} = (2 \cdot d + 8) \cdot \text{N}$$

From the TOPIQ article [38], we can estimate the number of FLOPS used in various DL-IQA estimators. The TOPIQ model employs approximately 19 giga FLOPS with around 35 million parameters. The DISTS [40] and LPIPS [41] models use approximately 62 giga FLOPS with around 12 million parameters. The PIEAPP model [42] utilizes approximately 155 giga FLOPS with around 80 million parameters.

In the context of IQA methods, classical methods are characterized by fixed parameters derived from mathematical operations and only a few trainable parameters, usually around five, typically from logistic regression. On the other hand, deep learning methods involve a significantly larger number of trainable parameters due to their reliance on neural networks architectures.

For instance, classical IQA methods like VIF, GMSD, and SSIM typically use fixed mathematical formulas and do not require extensive training. VIF (Visual Information Fidelity) employs fixed mathematical operations based on information theory and wavelet transforms without trainable parameters. Similarly, GMSD (Gradient Magnitude Similarity Deviation) involves fixed computations without trainable parameters, akin to VIF.

We also consider data rectification, given the VQEG logistics, with the number of iterations set to 250,000 (worst case). Assuming $m = 1000$ (the number of observations or data points that the logistic regression model is being fit to), the total FLOPS is:

$$\text{FLOPS} = (15 \cdot m + 5) \cdot 250{,}000 = 3.75 \text{ (G)}$$

In Table 8, we present the comparison of computational cost (FLOPS) and parameters used by LIQA methods with classical IQA and DL-IQA methods. The input image size is $3 \times 224 \times 224$ to ensure a fair comparison with DL-IQA methods. The FLOPS calculation for the metric represents the computational cost for processing each pair of images. The VQEG FLOPS refer to the entire final dataset, which is processed once. From the table we see how the number of necessary parameters is reduced to 2, compared to the 5 parameters suggested by the VQEG rectification. Furthermore, the low computational cost allows us to use IQA methods in real time while still maintaining high accuracy.

The direction towards hybrid models represents a natural evolution aiming to harness the best of both worlds: speed and reliability from classical methods, and precision from deep learning models. Moreover, combining traditional video quality metrics (such as PSNR, SSIM, MS-SSIM, FSIM, VIF) using machine learning techniques to weigh and fuse the results into an overall score that better reflects human perception is currently the most effective solution. In this context, LIQA methods can provide features to DL-IQA methods, leading to a reduction in the number of parameters, with particular attention to high accuracy in blur correction for the training of deep methods. As a further future development, it is possible to consider shifting the curves, and thus the model, to distances different from those used in the datasets, in spaces where DMOS and reference images are not present, working in Reduced Reference or No Reference environments [8,9]. Continuing

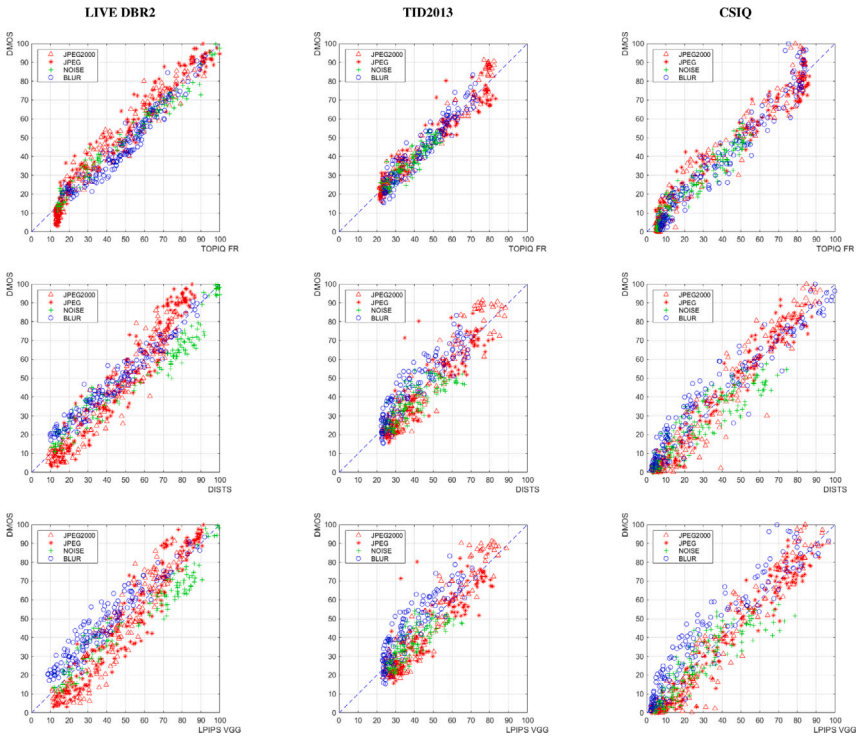

Fig. 20. The DMOS scatterplots versus the values predicted using TOPIQ-FR (upper row), DISTS (second row) and LPIPS-VGG (lower row) for blur, noise, JPEG, JPEG2000 distortions for LIVE DBR2 (first column), TID2013 (second column) and CSIQ (third column).

Table 6
Experimental verification of the **a-priori linearized IQA metrics** versus the **DL-IQA methods**. Four distortions subsets: blur, noise, JPEG, JPEG2000, for each single distance dataset LIVE DBR2, TID2013 and CSIQ. The best values are in bold blue, the second best in bold black. DL-IQA methods use millions of parameters, while the proposed LIQA method uses only two parameters. The number of parameters for the DL-IQA methods is expressed in millions (M).

IQA metric	Parameter	LIVE DBR2 quartet			TID2013 quartet			CSIQ quartet		
	no.	RMSE	SROCC	LCC	RMSE	SROCC	LCC	RMSE	SROCC	LCC
LVIF	**2**	8.6653	0.96536	0.95636	7.0170	0.92782	0.92789	7.5789	0.95878	0.96336
LGMSD	**2**	**7.2629**	**0.96888**	**0.96478**	**6.7091**	**0.95059**	**0.93579**	**7.7122**	**0.96909**	**0.96232**
TOPIQ FR	~35 M	**6.1015**	**0.97699**	**0.97420**	**5.0277**	**0.95797**	**0.96277**	**6.2881**	**0.96731**	**0.97493**
TOPIQ FR PIPAL	>35 M	9.6088	0.93748	0.93470	6.9727	0.90586	0.92706	9.3209	0.92414	0.94403
DISTS	~12 M	8.8388	0.94969	0.94503	8.0028	0.88691	0.90268	8.3521	0.94451	0.95532
LPIPS	~12 M	11.3597	0.91557	0.90742	7.9131	0.90441	0.90497	10.1163	0.94097	0.93372
LPIPS VGG	>100M	9.9120	0.92845	0.93035	8.7029	0.85383	0.88375	9.4743	0.92603	0.94212
PIEAPP	~80 M	11.6897	0.90935	0.90167	8.4116	0.90658	0.89187	9.5894	0.94837	0.94066

research on both approaches enables the development of more effective and versatile solutions for IQA.

7. Implementation issues

The computational steps for the implementation of the a-priori linearized methods are summarized as follows:

1. With reference to a typical natural image, stipulate a conventional DMOS for that image at some level of blur (*anchor* value) to determine the value of Q, using the canonical method.
2. Calculate the conversion function $\xi(\zeta;\tau)$ for the chosen conventional IQA method and for a set of viewing distance values of interest $\tau_1, \tau_2, \ldots, \tau_N$. To this purpose, the images are resampled for each viewing distance of interest.
3. Calculate the equivalent blur for the image of interest starting from the metric ζ determined by the application of the chosen conventional IQA and the associated conversion function $\xi(\zeta;\tau)$ for the wanted viewing distance.
4. Apply the rating function $\hat{d}_{LIQA}(\zeta)$ using the already determined values of Q and τ.

Since the costly step 2 is off-line, the computational burden for the estimation of the DMOS of a given degraded image reduces substantially to that required by the calculus of the metric ζ of the selected conventional IQA method.

8. Remarks

In the face of the concordant results of independent experiments based on different methods, protocols and viewing devices, one might still wonder how it is possible to obtain such accurate objective quality estimates linearly correlated with subjective judgments without a-posteriori calibration. The multiplicity of factors influencing the quality judgments pushes rather towards the use of empirical analogies, as done in classical and machine learning based IQA methods. To enlighten why the present unconventional calibration-free LIQA methods do constitute a viable approach, let us further evidence that these methods are not aimed to predict the subjective quality of a specific image, but rather the expected value of a random set of images characterized by a perceptual quality equivalent to Gaussian blur, which is regarded as *deterministic*. Its differences with respect to actual quality measurements are the effect of different unobserved factors. The good linear

Table 7
Experimental verification of the **a-priori linearized IQA metrics** versus the **DL-IQA methods** for the entire TID2013, CSIQ, LIVE DBR2 and LIVE MD datasets. Each best value distortion is highlighted in bold blue, the second best in bold black. We show the computational cost (FLOPS) and the number of parameters used by LIQA methods compared to classical IQA and DL-IQA methods. The FLOPS calculation for the metric refers to the computational cost for each pair of images processed. The values for classic or deep metrics are expressed in gigaflops (GFLOPS). The number of parameters for the DL-IQA methods is expressed in millions (M).

	TOPIQ FR	TOPIQ FR PIPAL	DISTS	LPIPS	LPIPS VGG	PIEAPP	LVIF	LGMSD
Parameter no.	~35 M	>35 M	~12 M	~12 M	>100 M	~80 M	**2**	**2**
GFLOPS	~19	>19	~62	~62	>62	~155	**8.876** 10^{-6}	**8.876** 10^{-6}

Datasets	Distortion type	TOPIQ FR		TOPIQ FR PIPAL		DISTS		LPIPS		LPIPS VGG		PIEAPP		LVIF		LGMSD	
		SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC
TID2013	Color additive noise	**0.85940**	**0.88616**	0.67623	0.69979	0.77524	0.77573	0.71814	0.70401	0.75052	0.74788	0.72645	0.73543	0.83099	0.86741	**0.86893**	**0.90584**
	Gaussian blur	**0.95781**	0.94712	0.91187	0.91419	0.92792	0.91942	0.95482	**0.94832**	0.91346	0.91529	0.89515	0.88314	**0.96529**	**0.95452**	0.91136	0.89622
	Gaussian white noise	**0.92780**	**0.92943**	0.81106	0.79340	0.83561	0.82461	0.80732	0.77647	0.82625	0.81714	0.85045	0.83958	0.89937	0.90233	**0.94605**	**0.94605**
	High frequency noise	**0.90949**	**0.95531**	0.82910	0.87718	0.86140	0.88995	0.83918	0.85558	0.85056	0.87439	0.87153	0.91126	0.89724	0.94620	**0.91640**	**0.95180**
	Impulse noise	**0.81325**	**0.79520**	0.60840	0.59069	0.67232	0.65150	0.66584	0.63559	0.61891	0.59871	0.76468	0.75597	**0.85420**	**0.84968**	0.76334	0.73752
	Masked noise	0.82655	0.84550	0.71230	0.78138	0.82612	0.85448	0.78751	0.82222	0.80759	0.83119	**0.82928**	**0.85959**	**0.84809**	**0.87843**	0.70855	0.68861
	Quantization noise	**0.86733**	**0.87277**	0.75364	0.74630	0.76881	0.76228	0.78646	0.77420	0.72987	0.72823	0.68225	0.66093	0.78650	0.77800	**0.90568**	**0.90582**
	Spatially correlated noise	**0.92012**	**0.91890**	0.86254	0.85819	0.85465	0.83945	0.78417	0.76617	0.81180	0.80685	0.80005	0.80804	0.88341	0.87633	**0.93485**	**0.92153**
	Block-wise distortions	0.19702	0.30319	0.38817	0.46027	0.33489	0.30580	0.45282	**0.51352**	0.52576	0.51080	0.15881	0.17683	**0.52957**	0.50336	**0.66305**	**0.62960**
	Chromatic aberrations	**0.89302**	0.92203	0.86589	0.95021	0.88625	0.95555	**0.89699**	0.92945	0.88542	**0.95942**	0.87867	**0.96144**	0.88548	0.94853	0.85166	0.91602
	Comfort noise	**0.93331**	**0.96060**	0.86779	0.91632	0.89312	0.89969	0.86773	0.86006	0.86985	0.89500	0.85927	0.91265	0.89502	0.90437	**0.92979**	**0.92543**
	Contrast change	0.58866	0.74896	0.59626	0.72227	0.47560	0.70351	0.43942	0.54380	0.35533	0.44257	**0.78930**	**0.85591**	**0.84816**	**0.88673**	0.32532	0.43384
	Image denoising	**0.94726**	**0.96406**	0.89383	0.93530	0.89007	0.92538	0.88555	0.90359	0.84314	0.89051	0.84322	0.87811	0.89321	0.88718	**0.95250**	**0.95287**
	Dither color quantization	**0.91863**	**0.92083**	0.80180	0.80875	0.81463	0.82240	0.79670	0.78004	0.79092	0.79277	0.88463	0.88149	0.84551	0.85975	**0.90982**	**0.91359**
	JPEG compression	**0.92369**	**0.96303**	0.88331	0.92695	0.88830	0.91495	0.89080	0.90589	0.87112	0.90253	0.84510	0.87099	0.91881	0.92737	**0.95001**	**0.97397**
	JPEG transmission errors	**0.91592**	**0.93322**	0.84367	0.87318	**0.91249**	0.88573	0.90487	0.88561	0.87416	0.88934	0.85169	0.86723	0.84164	**0.89794**	0.84011	0.77471
	JPEG2000 compression	**0.96366**	0.94797	0.92882	**0.95202**	0.93089	0.94130	0.92545	0.93467	0.92216	0.93213	0.94272	0.94964	0.95183	0.93798	**0.96558**	**0.95607**
	JPEG2000 transmission errors	**0.90369**	**0.90648**	0.88542	0.87749	0.86650	0.83445	0.81598	0.80886	0.81816	0.81025	0.85437	0.85859	0.87722	0.86847	**0.91347**	**0.88746**
	Lossy compression	**0.94948**	**0.95371**	0.91802	0.93107	0.92613	0.93125	0.90796	0.89874	0.91775	0.92247	0.86857	0.87879	0.92037	0.90062	**0.96306**	**0.94040**
	Mean shift	**0.81506**	**0.82059**	0.40003	0.46071	**0.80293**	**0.80866**	0.77628	0.80274	0.74932	0.66693	0.50405	0.49472	0.63070	0.60276	0.73561	0.76624
	Multiplicative Gaussian noise	**0.89490**	**0.89799**	0.76450	0.74778	0.78348	0.76232	0.72349	0.70283	0.75277	0.74788	0.82509	0.81782	0.84658	0.85800	**0.88805**	**0.88722**
	Non-eccentricity pattern noise	**0.82772**	**0.88367**	0.80788	0.84865	**0.84411**	**0.86185**	0.80460	0.80576	0.82694	0.84176	0.78298	0.79848	0.77113	0.76875	0.81427	0.79593
	Saturation change	**0.80788**	**0.80052**	0.47612	0.41774	0.79770	0.69480	**0.81322**	**0.79161**	0.77414	0.64074	0.69897	0.65795	0.16169	0.14065	0.19066	0.18733
	Sparse sampling	**0.95972**	0.94306	0.93447	**0.95867**	0.94066	0.94059	0.93815	0.94436	0.94281	0.95338	0.91400	0.93581	0.93589	0.92261	**0.96838**	**0.96408**
CSIQ	Gaussian white noise	**0.96041**	**0.96301**	0.86952	0.87738	0.92392	0.92232	0.92883	0.91415	0.92196	0.91569	0.94318	0.93977	0.95712	**0.95848**	**0.96789**	0.95846
	Gaussian blur	**0.97200**	**0.97329**	0.94857	0.96028	0.96023	0.96818	0.95002	0.93379	0.95368	0.95499	0.94827	0.92709	**0.97470**	**0.97409**	0.97125	0.96023
	JPEG compression	0.95119	**0.97886**	0.92813	0.96363	0.96370	0.97777	0.95144	0.96838	0.95631	0.96676	0.95000	0.97522	**0.97049**	**0.98225**	**0.96554**	0.96358
	Contrast decrement	**0.95136**	**0.95901**	0.92738	0.93159	**0.94791**	**0.94228**	0.94761	0.92431	0.91165	0.87544	0.94197	0.92855	0.93609	0.89651	0.90440	0.91735
	Additive pink gaussian noise	**0.96488**	**0.95918**	0.90333	0.90073	0.94172	0.93229	0.94522	0.93464	0.92539	0.91479	0.92559	0.91863	**0.95093**	**0.94208**	0.95064	0.93789
	JPEG2000 compression	0.96600	**0.97555**	0.92977	0.95319	0.95400	0.95908	0.93835	0.94964	0.95246	0.95506	0.95730	0.96737	**0.96731**	**0.97305**	**0.97177**	0.96860
LIVE DBR2	Gaussian blur	**0.97574**	0.96679	0.96495	0.95138	0.97117	0.97516	0.94480	0.90429	0.95011	0.95875	0.93058	0.89410	**0.98158**	**0.98714**	0.97484	**0.97930**
	Bit errors in JPEG2000 stream	**0.97028**	**0.97090**	0.96466	0.96699	0.96019	0.95859	0.95333	0.95294	0.95849	0.96124	0.95158	0.94491	**0.97728**	**0.97692**	0.96609	0.95633
	JPEG compression	0.97361	0.97659	0.96391	0.96391	0.96826	0.96884	0.96338	0.96204	0.96602	0.96292	0.95280	0.95671	**0.98569**	**0.98137**	**0.99059**	**0.98961**
	JPEG2000 compression	0.97138	0.97315	0.94187	0.94676	0.95180	0.95603	0.94719	0.95235	0.93976	0.93687	0.94525	0.94842	**0.97912**	**0.97509**	**0.98786**	**0.98101**
	Gaussian white noise	**0.98663**	**0.98750**	0.97189	0.96351	0.97394	0.95180	0.96993	0.93007	0.96847	0.96488	0.96912	0.94721	**0.98931**	**0.99311**	0.98461	0.97430
LIVE MD	Blur + JPEG	0.89785	0.92147	**0.90259**	**0.92789**	0.88377	0.88690	0.84737	0.89346	0.84184	0.84884	0.81801	0.87754	**0.89974**	**0.92963**	0.87355	0.86907
	Blur + Gaussian noise	0.89079	0.91163	**0.89874**	**0.93115**	0.79769	0.80659	0.76607	0.80842	0.76942	0.78714	0.78372	0.85523	**0.90560**	**0.93258**	0.86864	0.85771

Table 8
Computational cost (FLOPS) and parameter number used by LIQA methods compared to classical IQA and DL-IQA methods. The input image size is $3 \times 224 \times 224$ for a fair comparison with DL-IQA methods. The calculation of FLOPS for the metric refers to the computational cost for each pair of images processed. The VQEG FLOPS refer to the entire final dataset to be rectified, once. The values for classic or deep metrics are expressed in gigaflops (GFLOPS). The number of parameters for the DL-IQA methods is expressed in millions (M).

IQA metric	GFLOPS		Parameter
	Metric	VQEG	no.
LVIF	**8.876** 10^{-6}	–	**2**
LGMSD	**8.876** 10^{-6}	–	**2**
VIF	0.472	3.75	5
GMSD	0.001	3.75	5
TOPIQ FR	~19	3.75	~35 M
DISTS	~62	3.75	~12 M
LPIPS	~62	3.75	~12 M
PIEAPP	~155	3.75	~80 M

correlation of the theoretical estimates versus the empirical scores follows from the validity of the adopted general principles, namely the optimality of the HVS with respect to fine pattern localization, and the Weber law, as discussed in Section 3. Just as you are certain about the time that a stone takes to fall from a given height (believing in mechanics), you should also regard such linearity as a necessary consequence of the said principles.

As outlined in Section 4, the most critical point is the correct determination of a specimen original image representative of the universe of natural original images, beyond the spectral fall-off property invoked in the calculus of the expected average quality.

The specimen image covers the basic role of allowing conversion between conventional quality metrics and a perceptually quality equivalent Gaussian blur. In turn, the conversion rule constitutes the bridge for converting the estimated quality among different viewing distances. This possibility is precluded to conventional calibration-prone methods, also because of the scarcity of empirical datasets containing experiments for different viewing distances [43]. Among others, this feature allows straightforward prediction of the subjective image quality using jointly different databases, after equalization of the τ and of Q.

As far as the generalization of LIQA methods to different image degradation types is concerned, it is basically inherited from the properties of the corresponding IQA methods through the conversion formulas, as deduced in Section 4.3 and corroborated by the observation of the scatterplots of Section 5.

Extending the observations made regarding the choice of the specimen image, it is also outlined that, even though the quality equivalence invoked in Section 4 depends on empirical data through the settings of the IQA methods, these settings were established in a long series of past experiments and can be considered firmly consolidated.

Still, the methods considered in this paper cannot be strictly said "calibration-free" whenever the parameters Q and τ are unknown. In practice, they have a clear operative meaning and are simply determined in applications.

9. Conclusion

We have introduced a canonical estimator for a priori linearization of the blur metric and have demonstrated how any classical Image Quality Assessment (IQA) metric can be mapped to an equivalent blur metric. The study has provided a novel model for realistically estimating visual discomfort caused by the blur of natural scenes, taking into account the cognitive needs of subjects and the adaptive capability of the visual system.

We have presented a 2-parameter a-priori linearized model as an alternative to the 5-parameter scoring function suggested by VQEG, so as to remove the final linearization stage. The presented LIQA methods

appear to predict with a reasonable accuracy the subjective quality loss of natural images in the FR mode for a class of technically interesting degradations. By the comparative analysis of a degraded image against its pristine counterpart, LIQA methods provide quality loss estimates using only DMOS anchoring and viewing distance information.

Since they conform to an abstract model of real subjects retaining their essential, shared characteristics, it is expected that they will properly reflect the statistical opinion of large populations of observers. After all, this is the substantial goal of image providers.

Future advancements could involve adjusting the curves and consequently the model to different distances than those used in the datasets, operating in environments where DMOS and reference images are absent, specifically in Reduced Reference or No Reference settings. Moreover, LIQA methods can provide features to DL-IQA approaches, resulting in a decrease in the number of parameters while maintaining a focus on high precision in blur correction for training deep learning models.

CRediT authorship contribution statement

Paolo Giannitrapani: Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Elio D. Di Claudio:** Supervision, Writing – review & editing. **Giovanni Jacovitti:** Conceptualization, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.image.2024.117212.

References

[1] J. Mannos, D. Sakrison, The effects of a visual fidelity criterion of the encoding of images, IEEE Trans. Inform. Theory 20 (4) (1974) 525–536, http://dx.doi.org/10.1109/TIT.1974.1055250.

[2] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[3] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment (dec 13). URL http://sse.tongji.edu.cn/linzhang/IQA/FSIM/FSIM.htm.

[4] W. Xue, L. Zhang, X. Mou, A. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 684–695, http://dx.doi.org/10.1109/TIP.2013.2293423.

[5] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.

[6] VQEG, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Tech. Rep., Video Quality Experts Group, 2000, http://www.vqeg.org.

[7] E.D. Di Claudio, G. Jacovitti, A detail-based method for linear full reference image quality prediction, IEEE Trans. Image Process. 27 (1) (2018) 179–193.

[8] R. Gao, Z. Huang, S. Liu, QL-IQA: Learning distance distribution from quality levels for blind image quality assessment, Signal Process., Image Commun. 101 (2022) 116576, http://dx.doi.org/10.1016/j.image.2021.116576.

[9] C. Zhang, Z. Huang, S. Liu, J. Xiao, Dual-channel multi-task CNN for no-reference screen content image quality assessment, IEEE Trans. Circuits Syst. Video Technol. 32 (8) (2022) 5011–5025, http://dx.doi.org/10.1109/TCSVT.2022.3143321.

[10] VQEG, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II, Tech. Rep., Video Quality Experts Group, 2003, http://www.vqeg.org.

[11] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3441–3452, URL live.ece.utexas.edu/research/quality/subjective.htm.

[12] T.S.S.O. ITU, ITU-T Series J Cable Networks and Transmission of Television, Sound Programme and other Multimedia Signals, Measurement of the Quality of Service, Part 3, International Telecommunication Union, 2016.

[13] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723, http://dx.doi.org/10.1109/TAC.1974.1100705.

[14] K. Gu, M. Liu, G. Zhai, X. Yang, W. Zhang, Quality assessment considering viewing distance and image resolution, IEEE Trans. Broadcast. 61 (3) (2015) 520–531, http://dx.doi.org/10.1109/TBC.2015.2459851.

[15] H. Landau, Sampling, data transmission, and the Nyquist rate, Proc. IEEE 55 (10) (1967) 1701–1706, http://dx.doi.org/10.1109/PROC.1967.5962.

[16] T.S.S.O. ITU, Recommendation ITU-r BT.2022 general viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays, International Telecommunication Union, 2012.

[17] E.D. Di Claudio, P. Giannitrapani, G. Jacovitti, Predicting blur visual discomfort for natural scenes by the loss of positional information, Vis. Res. 189 (2021) 33–45, http://dx.doi.org/10.1016/j.visres.2021.07.018.

[18] E.D. Di Claudio, G. Jacovitti, A. Laurenti, Maximum likelihood orientation estimation of 1-D patterns in Laguerre–Gauss subspaces, IEEE Trans. Image Process. 19 (5) (2010) 1113–1125.

[19] F.W. Campbell, D.G. Green, Optical and retinal factors affecting visual resolution, J. Physiol. 181 (3) (1965) 576–593.

[20] M. Reisert, H. Burkhardt, Complex derivative filters, IEEE Trans. Image Process. 17 (12) (2008) 2265–2274.

[21] A.B. Watson, A.J. Ahumada Jr., A standard model for foveal detection of spatial contrast, J. Vis. 5 (9) (2005) http://dx.doi.org/10.1167/5.9.6, 6–6.

[22] J.G. Daugman, Six formal properties of two-dimensional anisotropic visual filters: structural principles and frequency/orientation selectivity, IEEE Trans. Syst. Man Cybern. 13 (1983) 882–887.

[23] H.L. Van Trees, Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise, Krieger Publishing Co., Inc., USA, 1992.

[24] A. Neri, G. Jacovitti, Maximum likelihood localization of 2-D patterns in the Gauss–Laguerre transform domain: theoretic framework and preliminary results, IEEE Trans. Image Process. 13 (1) (2004) 72–86, http://dx.doi.org/10.1109/TIP.2003.818021.

[25] A. Torralba, A. Oliva, Statistics of natural image categories, Netw., Comput. Neural Syst. 14 (3) (2003) 391–412, http://dx.doi.org/10.1088/0954-898X_14_3_302, PMID: 12938764.

[26] J.A. Solomon, The history of dipper functions, Atten. Percept. Psychophys. 71 (3) (2009) 435–443.

[27] L. O'Hare, P.B. Hibbard, Spatial frequency and visual discomfort, Vis. Res. 51 (15) (2011) 1767–1777, http://dx.doi.org/10.1016/j.visres.2011.06.002.

[28] Z. Wang, E. Simoncelli, A. Bovik, Multi-scale structural similarity for image quality assessment, in: Proc. of the 37th IEEE Asilomar Conference on Signal, Systems and Computer, Vol. 2, IEEE Comp. Soc., 2003, pp. 1398–1402.

[29] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 684–695, http://dx.doi.org/10.1109/TIP.2013.2293423.

[30] J.B. Martens, V. Kayargadde, Image quality prediction in a multidimensional perceptual space, in: Proc. of the 1996 IEEE Int. Conf. on Image Processing, Vol. 1, 1996, pp. 877–880.

[31] Z. Wang, A.C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26 (1) (2009) 98–117, http://dx.doi.org/10.1109/MSP.2008.930649.

[32] A.L. Caron, P.-M. Jodoin, Image multidistortion estimation, IEEE Trans. Image Process. 20 (12) (2011) 3442–3454, http://dx.doi.org/10.1109/TIP.2011.2159233.

[33] D. Jayaraman, A. Mittal, A.K. Moorthy, A.C. Bovik, Objective quality assessment of multiply distorted images, in: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers, ASILOMAR, 2012, pp. 1693–1697.

[34] T.S.S.O. ITU, Recommendation ITU-R BT.2012 Methodology for the Subjective Assessment of the Quality of Television Pictures, International Telecommunication Union, 2012.

[35] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. Jay Kuo, Image database TID2013: Peculiarities, results and perspectives, Signal Process., Image Commun. 30 (2015) 57–77, http://dx.doi.org/10.1016/j.image.2014.10.009.

[36] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 9 (1) (2010) 011006/1–21.

[37] T.S.S.O. ITU, ITU-T Recommendation P.10 Vocabulary for Performance and Quality of Service, Amendment 5, International Telecommunication Union, 2016.

[38] C. Chen, J. Mo, J. Hou, H. Wu, L. Liao, W. Sun, Q. Yan, W. Lin, TOPIQ: A top-down approach from semantics to distortions for image quality assessment, IEEE Trans. Image Process. 33 (2024) 2404–2418, http://dx.doi.org/10.1109/TIP.2024.3378466.

[39] G. Jinjin, C. Haoming, C. Haoyu, Y. Xiaoxing, J.S. Ren, D. Chao, PIPAL: A large-scale image quality assessment dataset for perceptual image restoration, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020, pp. 633–651.

[40] K. Ding, K. Ma, S. Wang, E.P. Simoncelli, Image quality assessment: Unifying structure and texture similarity, IEEE Trans. Pattern Anal. Mach. Intell. 44 (5) (2022) 2567–2581, http://dx.doi.org/10.1109/TPAMI.2020.3045810.

[41] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018.

[42] E. Prashnani, H. Cai, Y. Mostofi, P. Sen, PieAPP: Perceptual image-error assessment through pairwise preference, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018.

[43] X. Liu, M. Pedersen, J.Y. Hardeberg, CID:IQ – A new image quality database, in: A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass (Eds.), Image and Signal Processing, Springer International Publishing, Cham, 2014, pp. 193–202.