

Article

A Human Visual System Inspired No-Reference Image Quality Assessment Method Based on Local Feature Descriptors

Domonkos Varga 

Ronin Institute, Montclair, NJ 07043, USA; domonkos.varga@ronininstitute.org

Abstract: Objective quality assessment of natural images plays a key role in many fields related to imaging and sensor technology. Thus, this paper intends to introduce an innovative quality-aware feature extraction method for no-reference image quality assessment (NR-IQA). To be more specific, a various sequence of HVS inspired filters were applied to the color channels of an input image to enhance those statistical regularities in the image to which the human visual system is sensitive. From the obtained feature maps, the statistics of a wide range of local feature descriptors were extracted to compile quality-aware features since they treat images from the human visual system's point of view. To prove the efficiency of the proposed method, it was compared to 16 state-of-the-art NR-IQA techniques on five large benchmark databases, i.e., CLIVE, KonIQ-10k, SPAQ, TID2013, and KADID-10k. It was demonstrated that the proposed method is superior to the state-of-the-art in terms of three different performance indices.

Keywords: no-reference image quality assessment; quality-aware features; keypoint detector



Citation: Varga, D. A Human Visual System Inspired No-Reference Image Quality Assessment Method Based on Local Feature Descriptors. *Sensors* **2022**, *22*, 6775. <https://doi.org/10.3390/s22186775>

Academic Editor: Benoit Vozel

Received: 27 July 2022

Accepted: 5 September 2022

Published: 7 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the continuous development of imaging systems, the demand for innovative, objective image quality assessment (IQA) methods is growing. Since digital images are subjects of a variety of distortions and noise types during image acquisition [1], compression [2], reconstruction [3], and enhancement [4], image quality assessment has also many practical applications [5] in medical imaging [6], remote sensing imaging [7], monitoring the quality of streaming applications [8], or benchmarking image processing algorithms [9] under different distortions. Thus, objective IQA has been a subject of intensive research in the image processing community to replace subjective quality evaluation of digital images which is a time-consuming, expensive, and laborious process [10].

In the literature, objective IQA measures are traditionally divided into three branches [11], such as full-reference (FR), reduced-reference (RR), and no-reference (NR) IQA, with respect to the availability of the distortion-free—very often referred as reference—images. As the terminology implies, FR techniques evaluate the perceptual quality of a distorted image with full access to its reference image, while NR algorithms cannot rely on reference images. For RR methods, partial information about the reference images is available.

1.1. Contributions

The contributions of this work are as follows. Although the deep learning paradigm dominates the field of objective IQA [12,13], the interest in methods which simulates the sensitivity of HVS to statistical regularities and structures is also a hot research topic in the literature [14,15]. In our previous work [16], it was empirically corroborated that the statistics of local feature descriptors are quality-aware features. Here, we use systematically the statistics of local feature descriptors to compile a powerful feature vector for NR-IQA. Specifically, multiple HVS inspired filters are applied to the color channels of an input image to generate feature maps where the HVS sensitive statistical regularities are emphasized. Next, the statistics of local feature descriptors, such as KAZE [17] or BRISK [18], are

extracted from the feature maps as quality-aware features. Since there is a close connection between human visual perception and low-level visual features [19,20], various statistics of local feature descriptors is able to provide a powerful feature representation for NR-IQA. The effectiveness of the proposed method is empirically corroborated in tests recommended by the IQA community using publicly available large IQA benchmark databases, i.e., CLIVE [21], KonIQ-10k [22], SPAQ [23], TID2013 [24], and KADID-10k [25].

1.2. Structure of the Paper

The rest of this study is organized as follows. In Section 2, previous and related work are summarized. Section 3.1 provides an overview about the applied benchmark databases, gives the definition of the evaluation metrics, and defines the evaluation protocol. In Section 3.2, we introduce our proposed method. Section 4 presents the experimental results and a comparison to the state-of-the-art from various aspects. Finally, a conclusion is drawn in Section 5.

2. Related Work

NR-IQA algorithms can be classified into learning-free and learning-based categories. As the name indicates, learning-based methods rely on various machine and/or deep learning techniques to construct a model for perceptual quality estimation. Learning-free methods can be further divided into two groups, i.e., spatial [26] and spectral domain [27] based approaches. A common method [28] for score prediction involves fitting a portion of the training data to the joint distribution of the feature vector and the related opinion scores. Given the test data feature vector, the score prediction in this instance entails maximizing the likelihood of the test data opinion score. Other methods [29] that are both opinion- and distortion-unaware measure the separation in sparse feature space between the reference and distorted images. In contrast, Leonardi et al. [30] elaborated an opinion-unaware method that exploits the activation maps of pretrained convolutional neural networks [31] by considering the correlations between feature maps.

Classical machine learning based algorithms utilized for choice natural scene statistics (NSS) [32] and support vector regressors (SVR) [33]. According to our understanding, the evolution of the human visual system (HVS) has been driven by natural selection. Therefore, HVS assimilated a comprehensive knowledge about the regularity of our natural environment. In addition, researchers have pointed out [34] that certain image structure regularities deteriorate in the presence of noise and the deviation from them can be exploited for image quality evaluation. Classical methods utilizing NSS include for example BLIINDS-II [35], BRISQUE [36], CurveletQA [37], and DIIVINE [38]. Specifically, Saad et al. [35] constructed an NSS model from discrete cosine transform (DCT) coefficients of an image and fitted generalized Gaussian distributions on the coefficients to obtain their shape parameters which were used as quality-aware attributes and mapped onto quality scores with a trained SVR. In contrast, Jenadeleh and Moghaddam [39] proposed a Wakeby distribution statistical model to extract quality-aware features. Moorthy et al. [38] utilized steerable pyramid decomposition [40]—an overcomplete wavelet transform—using across multiple orientations and scales. In contrast, Mittal et al. [36] applied the spatial domain for NSS model construction. To be more specific, quality-aware features were derived from locally normalized luminance coefficients and mapped onto perceptual quality with a trained SVR. Liu et al. [37] proposed a two-stage framework incorporating a distortion classification and a quality prediction step. Further, quality-aware features were derived from the curvelet representation of the input image. Specifically, the authors empirically proved that the coordinates of the maxima given in log-histograms of the curvelet coefficients, the energy distributions, and the scale are good predictors of image perceptual quality. Based on the observation that image distortions can significantly modify the shapes of objects present in the image, Bagade et al. [41] introduced shape adaptive wavelet features for NR-IQA. In contrast, Jenadeleh et al. [42] boosted already existing NR-IQA features with features

proposed for image aesthetics assessment. Further, they demonstrated that aesthetic aware features are able to increase the performance of perceptual quality estimation.

Recently, deep neural networks, particularly convolutional neural networks (CNN) have gained a significant amount of attention in the literature due to their improved performance in many fields [43–46] compared to other approaches and paradigms. In NR-IQA, Kang et al. [47] applied first a CNN successfully. Namely, the authors implemented a traditional CNN which accepts image patches of 32×32 and predicts the patches' quality independently from each other. The entire image's perceptual quality was obtained by taking the arithmetic mean of the patches' quality scores. Similar to [47], Kim and Lee [48] trained a CNN on image patches but the patches' desired quality scores were determined by a traditional FR-IQA metric which restricts this method to the evaluation of artificially distorted images. Bare et al. [49] developed a network that operates on image patches similar to [47] but the patches target score is calculated from a traditional FR-IQA metric (feature similarity index [50]) similar to [48]. On the whole, the entire image's perceptual quality is estimated by the predicted feature similarity index [50] scores of the image patches. In contrast, Conde et al. [51] took a CNN backbone network and trained it using a loss function [52] which aims to minimize the mean squared error and maximize linear correlation coefficient between the predicted and ground-truth quality scores. Further, the authors applied several data augmentation techniques, such as horizontal flips, vertical flips, rotations, and random cropping. To handle images with different aspect ratios, Ke et al. [53] introduced a transformer [54] based NR-IQA model which applied a hash-based 2D absolute-position-encoding for embedding image patches extracted from multiple scales. In contrast, Zhu et al. [55] embedded the input images' original aspect ratios into the self-attention module of a swin transformer [56]. Sun et al. [57] introduced the distortion graph representation framework which contains a distortion type discrimination network aiming to discriminate between distortion types and a fuzzy prediction network for perceptual quality estimation. Liu et al. [58] introduced lifelong learning for NR-IQA to learn new distortion types without accessing to previous training data. First, the authors utilized a split-and-merge distillation strategy for compiling a single-head regression network. In the split phase, a distortion-specific generator was implemented for generating pseudo-features for unseen distortions. In the merge phase, these pseudo-features were coupled with pseudo-labels to distill knowledge about distortions.

A general, in-depth overview about the field of NR-IQA is out of the scope of this paper. For more details, we refer to the PhD thesis of Jenadeleh [59] and the book of Xu et al. [60] Besides natural images, there are other modalities, whose no-reference quality assessment are also investigated in the literature, such as stereoscopic images [61], light field images [62], or virtual reality [63].

3. Materials and Methods

3.1. Materials

In this part of the paper, the applied IQA benchmark databases and the evaluation protocol are discussed in detail.

3.1.1. Applied IQA Benchmark Databases

In this paper, we applied five publicly available IQA benchmark, such as CLIVE [21], KonIQ-10k [22], SPAQ [23], TID2013 [24], and KADID-10k [25], to evaluate and compare our proposed methods to the state-of-the-art. Specifically, CLIVE [21], KonIQ-10k [22], and SPAQ [23] contain unique quality labeled images with authentic distortions. The quality ratings were collected in a crowdsourcing experiment [64–66] for CLIVE [21] and KonIQ-10k [22], while the quality ratings were obtained in a traditional laboratory environment for SPAQ [23]. Further, CLIVE [21] and KonIQ-10k [22] contain images with fixed resolution. On the other hand, there is no fixed resolution in SPAQ [23] but the images have high resolution which varies around 4000×4000 . In contrast to CLIVE [21], KonIQ-10k [22], and SPAQ [23], TID2013 [24] and KADID-10k [25] consist of 24 and 25 reference images

whose perceptual quality are considered perfect, respectively. The quality labeled distorted images were produced artificially by an image processing tool from the reference images using different distortion types (i.e., JPEG compression noise, salt & pepper noise, Gaussian blur, etc.) at multiple distortion levels. The main properties of the applied IQA benchmark databases are summarized in Table 1. Further, the empirical distributions of quality scores are depicted in Figure 1.

Table 1. The main characteristics of the applied IQA benchmark databases.

Attribute	CLIVE [21]	KonIQ-10k [22]	SPAQ [23]	TID2013 [24]	KADID-10k [25]
#Distorted images	1162	10,073	11,125	3000	10,125
#Reference images	-	-	-	25	81
#Distortion types	-	-	-	24	25
#Distortion levels	-	-	-	5	5
Resolution	500 × 500	1024 × 768	~4000 × 4000	512 × 384	512 × 384
#Subjects	8100	1467	600	971	2209
#Annotations	1400	1,200,000	186,400	27,000	303,750
Scale of quality scores	0–100	1–5	0–100	0–9	1–5
Subjective methodology	crowdsourcing	crowdsourcing	laboratory	laboratory	crowdsourcing
Year	2017	2018	2020	2013	2019

3.1.2. Evaluation Protocol and Metrics

The assessment of NR-IQA algorithms involves the measurement of the correlation strength between the predicted scores and the ground-truth scores of an IQA benchmark database. As common in the literature, about 80% of images was used for training and the remaining 20% was used for testing in our experiments. Further, databases with artificial distortions were divided into training and test sets with respect to the reference images to prevent semantic content overlap between these two sets.

In this paper, the medians of Pearson linear correlation coefficients (PLCC), Spearman rank order correlation coefficient (SROCC), and Kendall rank order correlation coefficient (KROCC), which were measured over 100 random train-test splits, are given to characterize the performance of the proposed method and other examined state-of-the-art methods. However, there is a non-linear relationship between the predicted and the ground-truth scores. This is why, a non-linear logistic regression was applied before the computation of PLCC as advised by [67]:

$$Q_f = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(Q_p - \beta_3))} \right) + \beta_4 Q_p + \beta_5, \quad (1)$$

where Q_f and Q_p stand for the fitted and predicted score, respectively. Further, the regression parameters are denoted by β_i 's ($i = 1, \dots, 5$).

3.2. Proposed Method

The high-level overview of the proposed NR-IQA algorithm is depicted in Figure 2. It can be observed that the proposed method is built upon two distinct steps. In the first, training step, local features are extracted from a database of quality labeled, training images. Next, a regression model is trained based on them to obtain a quality model. This model is used to estimate the perceptual quality of a previously unseen image in the testing step.

Image distortions influence the human visual system's (HVS) sensitivity to local image structures, such as edges or texture elements [68]. For that reason, many NR-IQA methods have been proposed [69–71] in the literature to compile a quality model from them. However, edge information or local binary patterns are not always able to provide powerful feature representation for NR-IQA. Therefore, in this study, an application of local feature descriptors and HVS-inspired filters are investigated thoroughly to extract quality-aware local features and compile a powerful feature representation for NR-IQA.

Figure 3 depicts the general process of local quality-aware feature extraction. First, the input RGB image is filtered by a set of HVS inspired filters to create feature maps. On these maps, local keypoints are detected by local feature descriptors (such as SURF [72]). Finally, feature extraction is carried out from the neighborhoods of the detected keypoints.

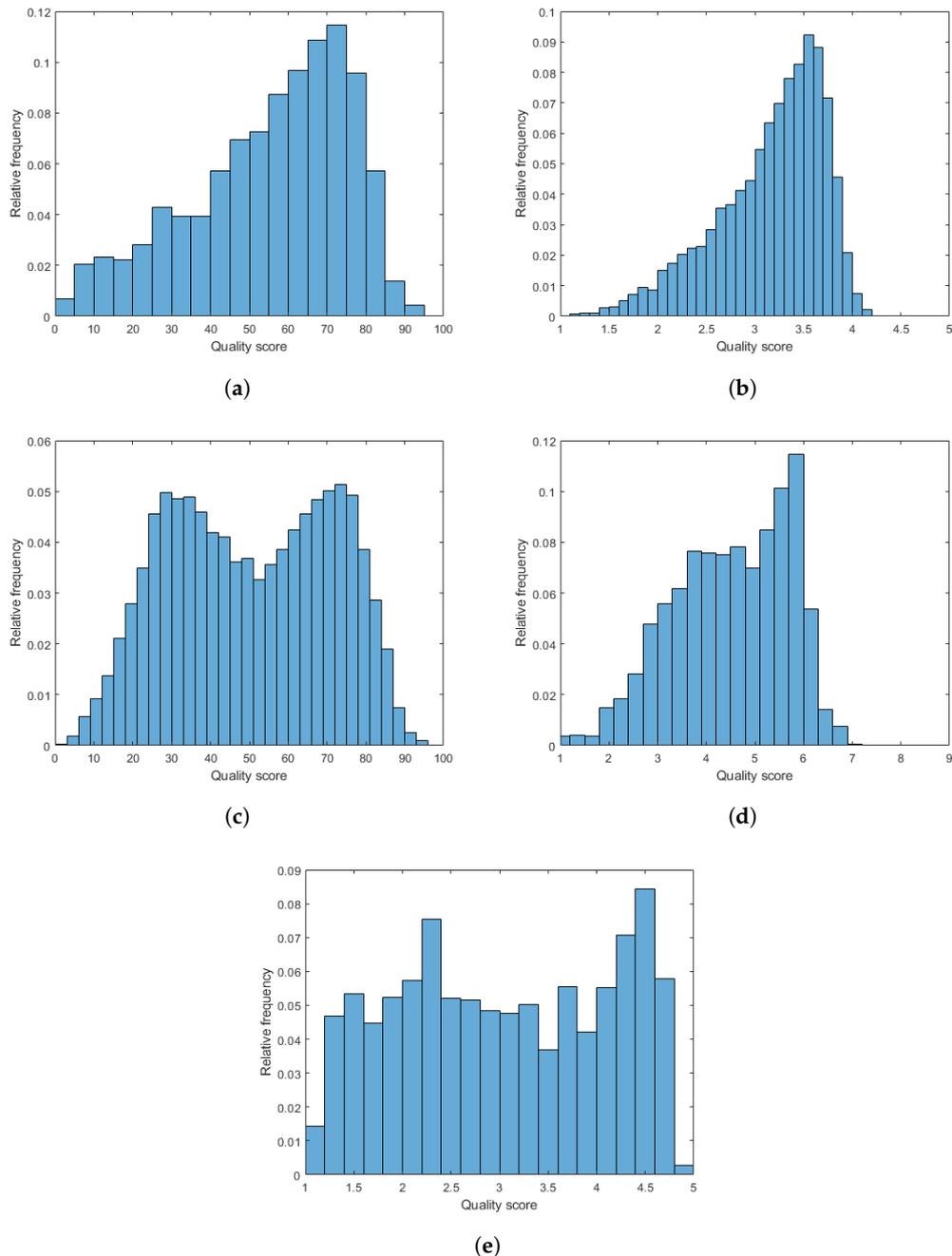


Figure 1. The empirical distributions of quality scores in the applied IQA databases. (a) CLIVE [21], (b) KonIQ-10k [22], (c) SPAQ [23], (d) TID2013 [24], (e) KADID-10k [25].

In the proposed method, an input RGB image is converted into YCbCr color space, since the chroma component is separated from the color information in YCbCr. The conversion from RGB to YCbCr was carried out using the following equation [73]:

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.2568 & 0.5041 & 0.0979 \\ -0.1482 & -0.2910 & 0.4392 \\ 0.4392 & -0.3678 & -0.0714 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \tag{2}$$

where R , G , and B stand for the red, green, and blue color channels, respectively. Subsequently, a color channel C_i is filtered using different HVS-inspired filters to obtain multiple feature maps. Since local feature descriptors treat images from the HVS's point of view, their statistics are able to provide quality-aware features [16]. Further, the applied HVS inspired filters emphasize those statistical regularities of a natural scene which are highly sensitive to image distortions from the perspective of HVS. Specifically, 5 statistical features are derived from each filtered color channels using the statistics of different local feature descriptors. In Sections 3.3–3.5, the compilation of HVS inspired feature maps are described. Next, the proposed quality-aware feature extraction from the feature maps is described in Section 3.6.

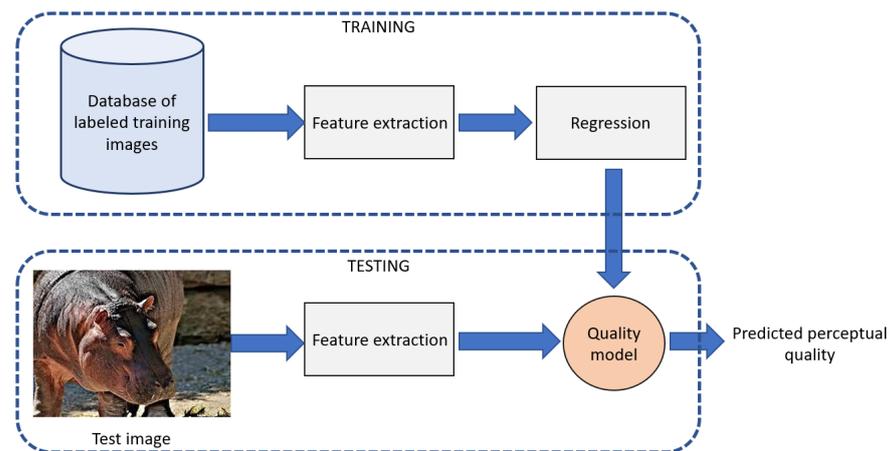


Figure 2. High-level overview of the proposed NR-IQA algorithm.

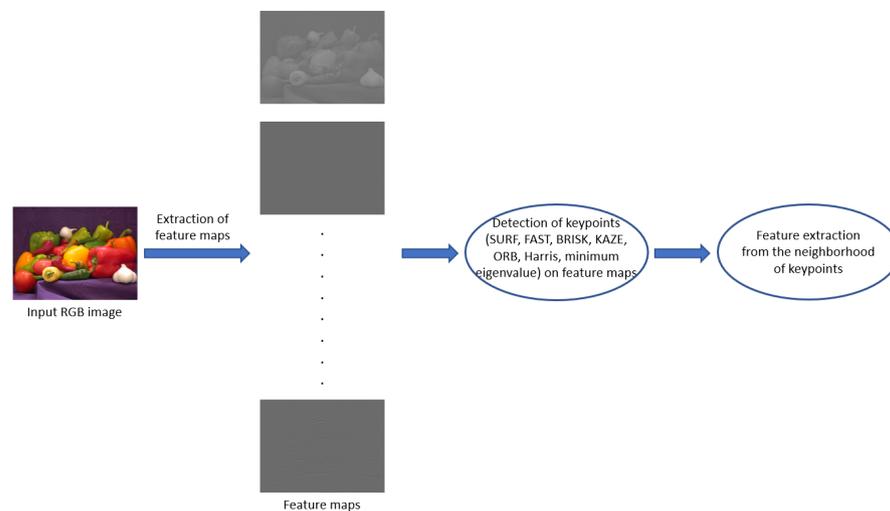


Figure 3. High-level overview of local feature extraction.

3.3. Bilaplacian Feature Maps

First, Bilaplacian feature maps were obtained using Bilaplacian filters. In [74], Gerhard et al. demonstrated that the HVS is highly adapted to statistical regularities of images. Further, zero-crossings [75] in an image occur where the gradient starts increasing or decreasing and help the HVS in interpreting the image. Using the idea of zero-crossings, Ghosh et al. [76] pointed out that the behaviour of the extended classical receptive field of retinal ganglion cells can be modeled as a combination of three zero-mean Gaussians at

three different scales which are equivalent to are the Bilaplacian of the Gaussian filter [77]. The $L(x, y)$ Laplacian of an $I(x, y)$ image can be expressed as:

$$L(x, y) = \frac{\partial^2 I(x, y)}{\partial x^2} + \frac{\partial^2 I(x, y)}{\partial y^2}. \quad (3)$$

Since a digital image is represented as a set of discrete pixels, discrete convolution kernels are used to approximate the Laplacian. In this paper, the following kernels are considered:

$$\mathbf{L}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \mathbf{L}_2 = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}, \mathbf{L}_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad (4)$$

$$\mathbf{L}_4 = \begin{pmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{pmatrix}, \mathbf{L}_5 = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}. \quad (5)$$

Bilaplacian kernels are obtained by convolving two Laplacian kernels:

$$\mathbf{L}_{ij}^2 = \mathbf{L}_i * \mathbf{L}_j, \quad (6)$$

where $*$ stands for the convolution operator. In our study, $\mathbf{L}_{11}^2, \mathbf{L}_{22}^2, \mathbf{L}_{33}^2, \mathbf{L}_{44}^2, \mathbf{L}_{55}^2, \mathbf{L}_{13}^2,$ and \mathbf{L}_{24}^2 Bilaplacian kernels were considered. Subsequently, a set of feature maps is derived from the input image by filtering with the Bilaplacian kernels the $Y, Cb,$ and Cr channels. To be more specific, $3 \times 7 = 21$ feature maps are obtained by filtering 3 color channels (Y, Cb, Cr) with 7 filters ($\mathbf{L}_{11}^2, \mathbf{L}_{22}^2, \mathbf{L}_{33}^2, \mathbf{L}_{44}^2, \mathbf{L}_{55}^2, \mathbf{L}_{13}^2, \mathbf{L}_{24}^2$).

3.4. High-Boost Feature Maps

High-boost filtering is used to enhance high-frequency image regions which the HVS is also sensitive for [78]. Similarly to the previous subsection, high-boost filtering is applied on the color channels of $Y, Cb,$ and Cr to strengthen high-frequency information. In this case, the convolution kernel is the following:

$$\mathbf{H} = \begin{pmatrix} -C & -C & -C \\ -C & 8C + 1 & -C \\ -C & -C & -C \end{pmatrix}, \quad (7)$$

where C is a constant value which controls the enhancement difference between a pixel location and its neighborhood. In our study, $C = 1$ was used. However, image distortions can occur at different scales. Therefore, a color channel was filtered 4 times in succession to obtain four feature maps from one channel. Since we have three channels, $3 \times 4 = 12$ feature maps were extracted in total applying high-boost filtering.

3.5. Derivative Feature Maps

In [74], Gerhard et al. draw the inference that the HVS is biased for processing natural images. Further, it has a large knowledge of statistical regularities in images. In [79], Li et al. demonstrated that derivatives and higher order derivatives are related to different statistical regularities of a natural scene. Therefore, there are good features for NR-IQA. For instance, higher order derivatives may be able to capture detailed discriminative information, while first order derivative information is typically related to the slope and elasticity of a surface. Second order derivatives intended to capture the geometric qualities associated to curvature [80]. Motivated by these previous works, we used the following convolution of two derivative kernels to filter Y, Cb, Cr color channels of an input image:

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & -1 \\ 1 & 0 & 1 \\ -1 & 1 & -1 \end{pmatrix} * \begin{pmatrix} 1 & -1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 1 \end{pmatrix}. \quad (8)$$

Similarly, we can define \mathbf{D}_2 , \mathbf{D}_3 , \mathbf{D}_4 , and \mathbf{D}_5 masks for 5×5 , 7×7 , 11×11 , and 13×13 sizes. Finally, the derivative feature maps are obtained by filtering Y , Cb , Cr color channels with \mathbf{D}_1 , \mathbf{D}_2 , \mathbf{D}_3 , \mathbf{D}_4 , and \mathbf{D}_5 . As a result, $3 \times 5 = 15$ derivative feature maps were obtained.

3.6. Feature Extraction

As one can see from the previous subsections, 21 Bilaplacian feature maps, 12 high-boost feature maps, and 15 derivative feature maps were generated which means in total $21 + 12 + 15 = 48$ feature maps. In the feature extraction step, $7 \times N$ keypoints are detected using 7 different local keypoint detectors, i.e., SURF (speed up robust features) [72], FAST (features from accelerated segment test) [81], BRISK (binary robust invariant scalable keypoints) [18], KAZE (Japanese word that means wind) [17], ORB (oriented FAST and rotated binary robust independent elementary features) [82], Harris [83], and minimum eigenvalue [84], in each feature map. For each keypoint, its $M \times M$ rectangular neighborhood with the keypoint's location as center point is taken. Further, each $M \times M$ rectangular block in each Bilaplacian, high-boost, and derivative feature maps are characterized by the mean, median, standard deviation, skewness, and kurtosis of the grayscale values found in the involved block. The skewness of a set of n elements is determined as

$$s = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^2}, \quad (9)$$

where \bar{x} is the arithmetic mean of all x_i elements. Similarly, the kurtosis can be given as

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}. \quad (10)$$

Using the statistics of the rectangular blocks, a given feature map is characterized by the arithmetic mean of all blocks' statistics. As a result, a $3 \times 7 \times 7 \times 5 = 735$ dimensional feature vector is obtained using Bilaplacian filters and the statistics of local feature descriptors, since 3 color channels were filtered with 7 Bilaplacian kernels and 7 different feature descriptors with 5 different statistics were applied. Further, a $3 \times 4 \times 7 \times 5 = 420$ dimensional feature vector is obtained using the high-boost filters, since 3 color channels were filtered with 4 high-boost kernels and 7 different feature descriptors with 5 different statistics were applied as in the previous case. Similarly, $3 \times 5 \times 7 \times 5 = 525$ dimensional feature vector is obtained from the derivative feature maps. By concatenating the fore-mentioned vectors, a $735 + 420 + 525 = 1680$ dimensional feature vector can be derived which can be mapped onto perceptual quality scores with a machine learning technique.

As an illustration, Figure 4 depicts a Y channel and its Bilaplacian feature maps with the detected FAST keypoints. From this illustration it can be seen that keypoints are accumulated around those regions which highly influence humans' quality perception. In Figure 5, it is illustrated that the location of keypoints on the feature maps is changing with respect to the strength of image distortion. As a consequence, it seems justified that the statistics of local feature descriptors on carefully chosen feature maps are quality-aware features.

3.7. Perceptual Quality Estimation

The quality model formally can be written as: $\mathbf{q} = G(\mathbf{F})$, where \mathbf{q} is a vector of quality scores, \mathbf{F} is a set of extracted feature vectors, and G denotes the quality model. Specifically, G can be determined by a properly chosen machine learning (regression) technique. In

this study, we made experiments with two different regression methods, i.e., support vector regressor (SVR) and Gaussian process regressor (GPR). In the followings, we denote the proposed methods by LFD-IQA-SVR and LFD-IQA-GPR with respect to the applied regression method. In the chosen codename, LFD refers to the abbreviation of local feature descriptors whose statistics were utilized as quality-aware features.

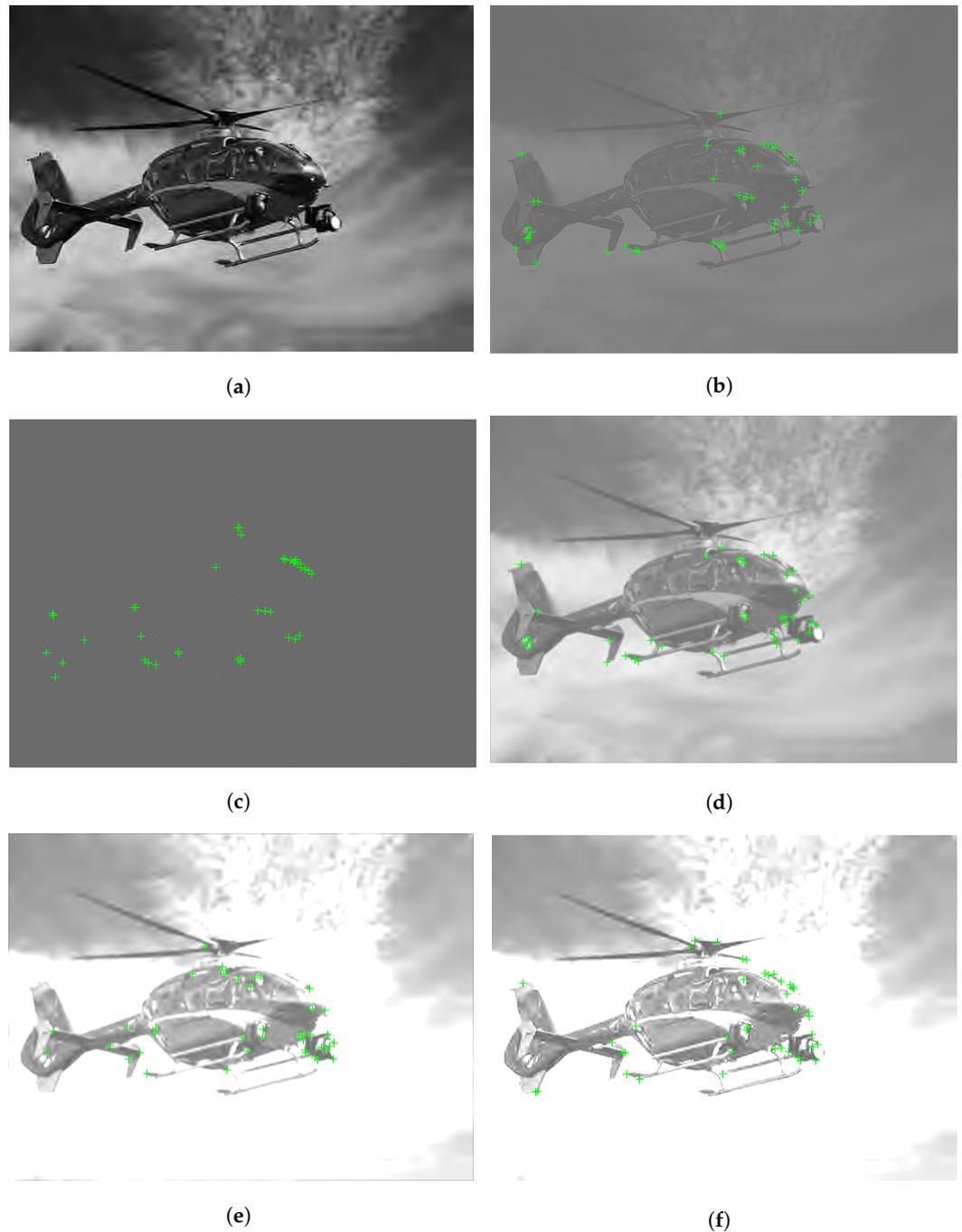


Figure 4. *Cont.*

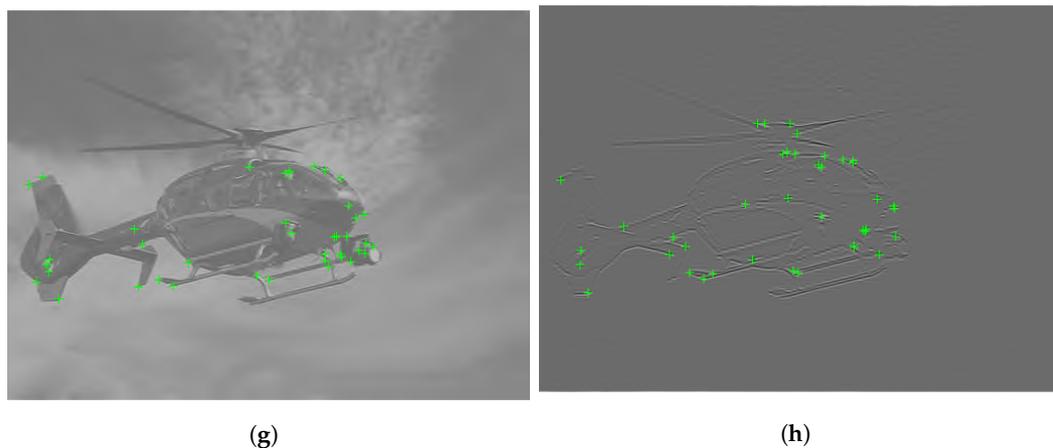


Figure 4. Illustration of Bilaplacian feature maps and detected FAST keypoints. (a) Y channel, (b) $Y * L_{11}^2$, (c) $Y * L_{22}^2$, (d) $Y * L_{33}^2$, (e) $Y * L_{44}^2$, (f) $Y * L_{55}^2$, (g) $Y * L_{13}^2$, (h) $Y * L_{24}^2$.

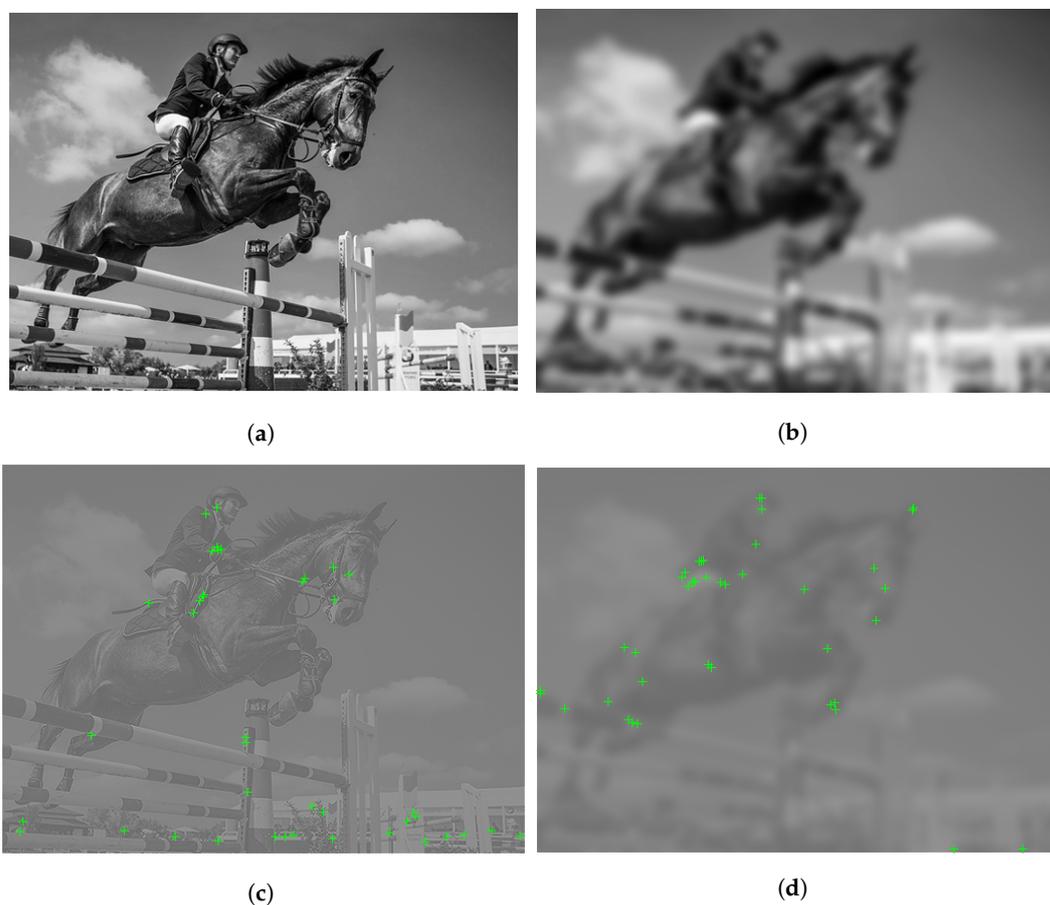


Figure 5. Illustration of Bilaplacian feature map and the detected FAST keypoints on two different distortion levels. (a) Luminance channel of pristine image ($Y_{pristine}$), (b) Luminance channel of distorted image ($Y_{distorted}$), (c) $Y_{pristine} * L_{11}^2$, (d) $Y_{distorted} * L_{11}^2$.

4. Experimental Results

In this section, our numerical experimental results are presented. First, an ablation study is carried out in Section 4.1 first to justify certain design choices of the proposed methods. In the following subsection, a comparison to several other state-of-the-art methods is presented using accepted publicly available benchmark databases and evaluation protocol described in Section 3.1. This comparison involves direct and cross-database tests as well as significance tests.

4.1. Ablation Study

In the proposed feature extraction methodology, there are two tunable parameters, i.e., the N number of detected keypoints for each feature feature descriptor and the $M \times M$ block size. In this ablation study, CLIVE [21] was utilized using the evaluation protocol given in Section 3.1 to determine an optimal value for these two parameters. To be more specific, we varied the number of detected keypoints from 1 to 55 and we experimented with 3 different block sizes, i.e., 3×3 , 5×5 , and 7×7 . The results for LFD-IQA-SVR and LFD-IQA-GPR are summarized in Figures 6 and 7, respectively. From these results, it can be seen that 5×5 -sized neighborhood is the optimal choice for both LFD-IQA-SVR and LFD-IQA-GPR. On the other hand, LFD-IQA-SVR achieves its best performance at 45 detected keypoints while LFD-IQA-GPR has its peak performance at 40 detected keypoints. Therefore, we applied 5×5 neighborhoods and 45 or 40 keypoints, respectively.

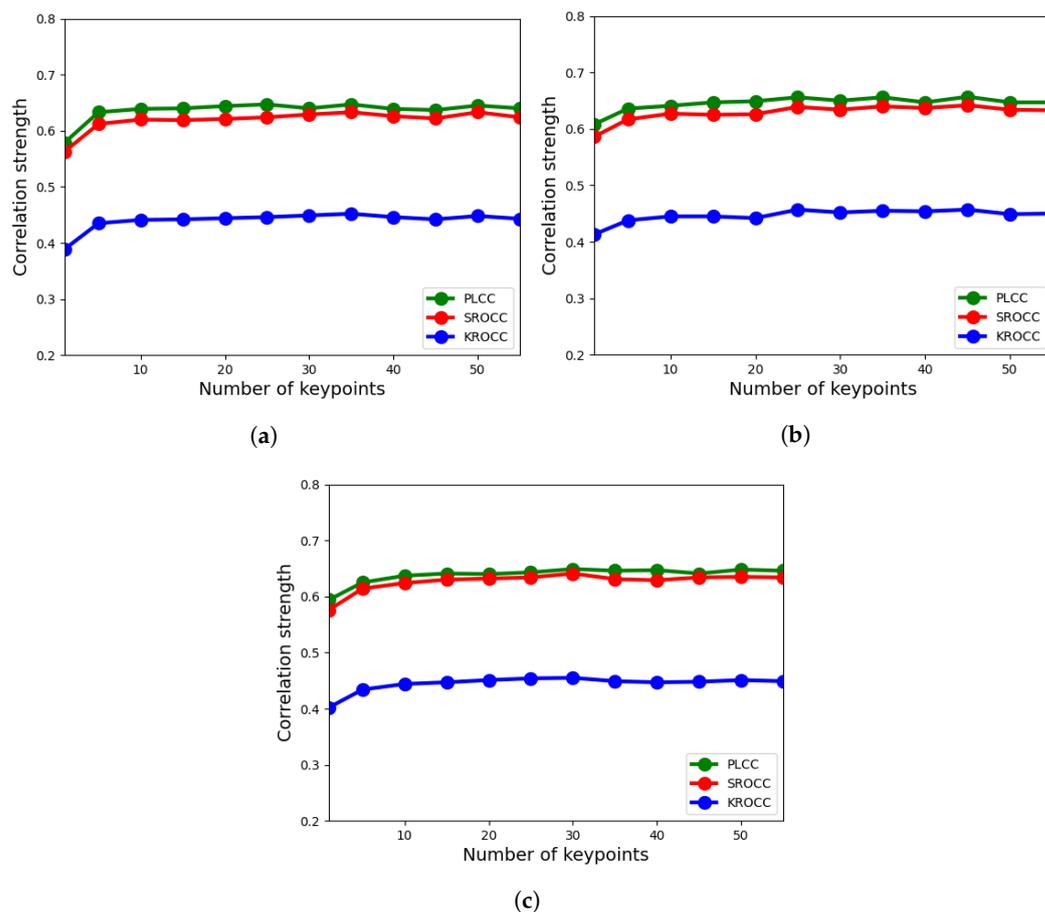


Figure 6. Ablation study for LFD-IQA-SVR: number of detected keypoints vs. correlation strength. (a) 3×3 block size, (b) 5×5 block size, (c) 7×7 block size.

The proposed methods were implemented and tested in MATLAB R2022a. To be more specific, the Computer Vision Toolbox's functions were utilized for the detection of keypoints and feature extraction, while the Statistics and Machine Learning Toolbox was used in the regression part of the proposed method.

4.2. Comparison to the State-of-the-Art

The proposed methods were compared to the following 16 state-of-the-art methods: BIQI [85], BLIINDS-II [35], BMPRI [86], BRISQUE [36], CurveletQA [37], DIIVINE [38], ENIQA [87], GRAD-LOG-CP [69], GWH-GLBP [70], IL-NIQE [27], NBIQA [88], NIQE [26], OG-IQA [71], PIQE [89], Robust BRISQUE [90], and SSEQ [91]. Excluding the training-free IL-NIQE [27], NIQE [26] and, PIQE [89], these methods were evaluated as the same way as the proposed methods. To provide a fair comparison, the same subsets of images were

selected in the random 100 train-test splits. Since IL-NIQE [27], NIQE [26] and, PIQE [89] are opinion unaware methods, they were tested on the applied benchmark databases in one iteration measuring PLCC, SROCC, and KROCC on the entire database without any train-test splits. In Tables 2 and 3, the median values measured over 100 random train-test splits for the considered and proposed NR-IQA methods on authentic distortions (CLIVE [21], KonIQ-10k [22], SPAQ [23]) are reported. Similarly, Table 4 summarizes the results on artificial distortions. From these results, it can be seen that the proposed LFD-IQA-SVR achieves the second best results in almost all cases, while the proposed LFD-IQA-GPR provides the best results for all databases in all performance metrics. In Table 5, the results measured on the individual databases are aggregated into direct and weighted averages of PLCC, SROCC, and KROCC. From these results, it can be concluded that the proposed methods are able to outperform all the other methods by a large margin. Further, the difference between the proposed and the other algorithms is larger in case of weighted averages. This indicates that the proposed methods tend to give a better performance on larger IQA databases. Figures 8 and 9 depict ground-truth versus predicted quality score scatter plots of the proposed methods determined on CLIVE [21], KonIQ-10k [22], and KADID-10k [25] test sets, respectively.

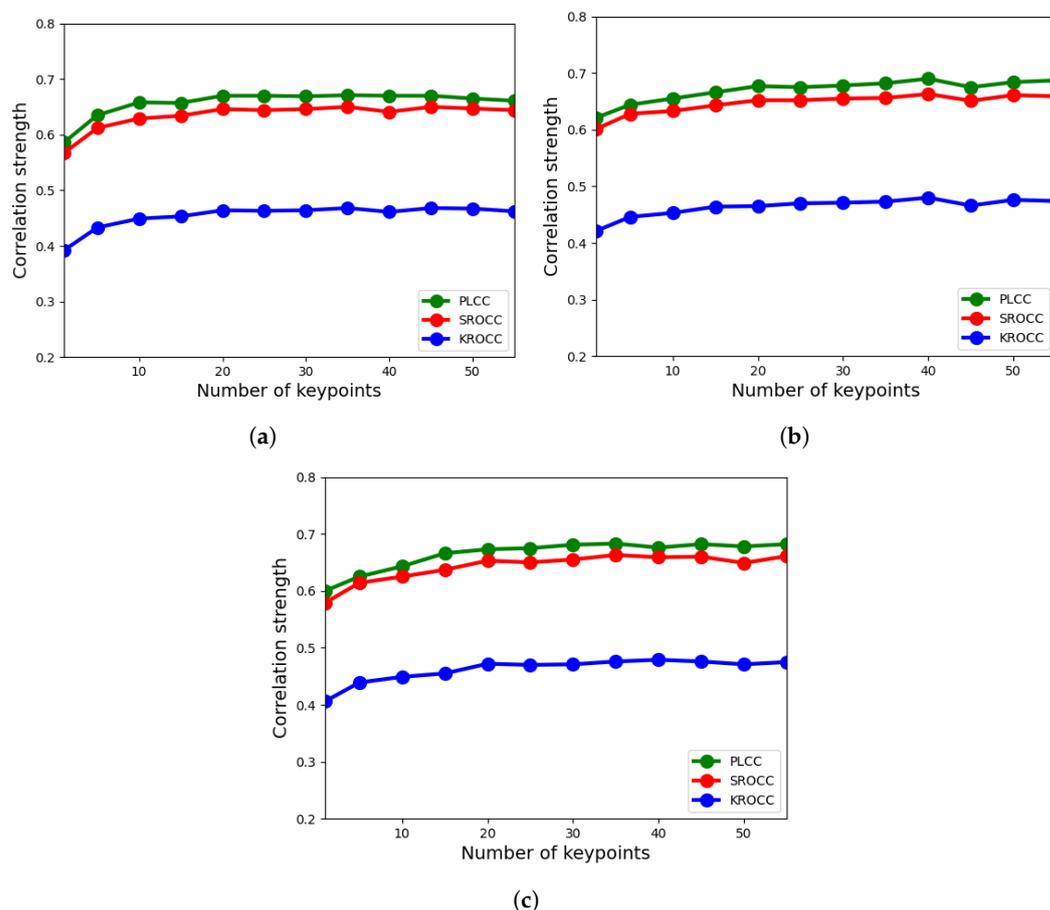


Figure 7. Ablation study for LFD-IQA-GPR: number of detected keypoints vs. correlation strength. (a) 3×3 block size, (b) 5×5 block size, (c) 7×7 block size.

To prove that achieved results summarized in Tables 2–4 are significant, the Wilcoxon rank sum test was applied [69,92]. To be specific, the null hypothesis was that two sets of 100 SROCC values produced by two different NR-IQA methods were sampled from continuous distributions with equal median values. In our tests, 5% significance level was applied. The results are summarized in Table 6 for LFD-IQA-SVR, while the results are shown in Table 7 for LFD-IQA-GPR. Here, symbol ‘1’ is used to denote that the proposed method is significantly better than the method in the row on the database in the column.

From the presented results, it can be clearly seen that the achieved result is significant compared to the state-of-the-art. As a consequence, the proposed HVS-inspired feature extraction method have proved to be more effective than the those of the examined state-of-the-art methods.

Table 2. Comparison to the state-of-the-art on CLIVE [21] and KonIQ-10k [22] databases. Median PLCC, SROCC, and KROCC values were measured over 100 random train-test splits. The best results are typed in bold, the second best results are underlined, and the third best results are typed in italic.

Method	CLIVE [21]			KonIQ-10k [22]		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
BIQI [85]	0.519	0.488	0.329	0.688	0.662	0.471
BLIINDS-II [35]	0.473	0.442	0.291	0.574	0.575	0.414
BMPRI [86]	0.541	0.487	0.333	0.637	0.619	0.421
BRISQUE [36]	0.524	0.497	0.345	0.707	0.677	0.494
CurveletQA [37]	0.636	0.621	0.421	0.730	0.718	0.495
DIIVINE [38]	0.617	0.580	0.405	0.709	0.693	0.471
ENIQA [87]	0.596	0.564	0.376	0.761	0.745	0.544
GRAD-LOG-CP [69]	0.607	0.604	0.383	0.705	0.696	0.501
GWH-GLBP [70]	0.584	0.559	0.395	0.723	0.698	0.507
IL-NIQE [27]	0.487	0.415	0.280	0.463	0.447	0.306
NBIQA [88]	0.629	0.604	0.427	0.771	0.749	0.515
NIQE [26]	0.328	0.299	0.200	0.319	0.400	0.272
OG-IQA [71]	0.545	0.505	0.364	0.652	0.635	0.447
PIQE [89]	0.172	0.108	0.081	0.208	0.246	0.172
Robust BRISQUE [90]	0.522	0.484	0.330	0.718	0.668	0.477
SSEQ [91]	0.487	0.436	0.309	0.589	0.572	0.423
LFD-IQA-SVR	<u>0.669</u>	<u>0.647</u>	<u>0.464</u>	<u>0.786</u>	<u>0.769</u>	<u>0.569</u>
LFD-IQA-GPR	0.696	0.667	0.480	0.801	0.775	0.577

Table 3. Comparison to the state-of-the-art on SPAQ [23] database. Median PLCC, SROCC, and KROCC values were measured over 100 random train-test splits. The best results are typed in bold, the second best results are underlined, and the third best results are typed in italic.

Method	PLCC	SROCC	KROCC
BIQI [85]	0.783	0.776	0.566
BLIINDS-II [35]	0.676	0.675	0.486
BMPRI [86]	0.739	0.734	0.506
BRISQUE [36]	0.726	0.720	0.518
CurveletQA [37]	0.793	0.774	0.503
DIIVINE [38]	0.774	0.756	0.514
ENIQA [87]	0.813	0.804	0.603
GRAD-LOG-CP [69]	0.786	0.782	0.572
GWH-GLBP [70]	0.801	0.796	0.542
IL-NIQE [27]	0.374	0.348	0.297
NBIQA [88]	0.802	0.793	0.539
NIQE [26]	0.264	0.310	0.206
OG-IQA [71]	0.726	0.724	0.594
PIQE [89]	0.211	0.156	0.091
Robust BRISQUE [90]	0.735	0.731	0.524
SSEQ [91]	0.745	0.742	0.549
LFD-IQA-SVR	<u>0.852</u>	<u>0.847</u>	<u>0.627</u>
LFD-IQA-GPR	0.869	0.864	0.664

Table 4. Comparison to the state-of-the-art on TID2013 [24] and KADID-10k [25] databases. Median PLCC, SROCC, and KROCC values were measured over 100 random train-test splits carried out with respect to the reference images. The best results are typed in bold, the second best results are underlined, and the third best results are typed in italic.

Method	TID2013 [24]			KADID-10k [25]		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
BIQI [85]	0.468	0.296	0.207	0.302	0.294	0.206
BLIINDS-II [35]	0.521	0.490	0.342	0.553	0.534	0.379
BMPRI [86]	0.692	0.583	0.422	0.555	0.534	0.382
BRISQUE [36]	0.565	0.411	0.289	0.426	0.398	0.276
CurveletQA [37]	0.560	0.471	0.337	0.471	0.442	0.316
DIIVINE [38]	0.521	0.487	0.340	0.429	0.436	0.307
ENIQA [87]	0.596	0.545	0.385	0.637	0.641	0.466
GRAD-LOG-CP [69]	0.662	0.627	0.454	0.590	0.570	0.415
GWH-GLBP [70]	0.315	0.357	0.245	0.302	0.285	0.196
IL-NIQE [27]	0.516	0.456	0.317	0.588	0.630	0.453
NBIQA [88]	0.695	0.628	0.459	0.646	0.615	0.446
NIQE [26]	0.263	0.277	0.184	0.302	0.338	0.228
OG-IQA [71]	0.564	0.452	0.321	0.527	0.447	0.314
PIQE [89]	0.491	0.364	0.255	0.289	0.237	0.201
Robust BRISQUE [90]	0.487	0.315	0.218	0.375	0.301	0.209
SSEQ [91]	0.615	0.520	0.373	0.454	0.434	0.304
LFD-IQA-SVR	0.637	0.645	0.470	0.845	0.838	0.640
LFD-IQA-GPR	0.705	0.669	0.492	0.857	0.848	0.654

Table 5. Direct and weighted average of PLCC, SROCC, and KROCC performance metrics. The best results are typed in bold, the second best results are underlined, and the third best results are typed in italic.

Method	Direct Average			Weighted Average		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
BIQI [85]	0.552	0.503	0.356	0.584	0.556	0.398
BLIINDS-II [35]	0.559	0.543	0.382	0.592	0.583	0.416
BMPRI [86]	0.633	0.591	0.413	0.647	0.623	0.434
BRISQUE [36]	0.590	0.541	0.384	0.615	0.582	0.417
CurveletQA [37]	0.638	0.605	0.414	0.658	0.633	0.431
DIIVINE [38]	0.610	0.590	0.407	0.631	0.618	0.424
ENIQA [87]	0.681	0.660	0.475	0.723	0.711	0.521
GRAD-LOG-CP [69]	0.670	0.656	0.465	0.691	0.678	0.491
GWH-GLBP [70]	0.545	0.539	0.377	0.588	0.578	0.403
IL-NIQE [27]	0.486	0.459	0.331	0.476	0.468	0.345
NBIQA [88]	0.709	0.678	0.477	0.734	0.710	0.495
NIQE [26]	0.295	0.325	0.218	0.292	0.340	0.229
OG-IQA [71]	0.603	0.553	0.408	0.629	0.590	0.442
PIQE [89]	0.274	0.222	0.160	0.255	0.221	0.159
Robust BRISQUE [90]	0.567	0.500	0.352	0.600	0.547	0.389
SSEQ [91]	0.578	0.541	0.392	0.598	0.577	0.421
LFD-IQA-SVR	0.758	0.749	0.554	0.807	0.799	0.596
LFD-IQA-GPR	0.786	0.765	0.573	0.827	0.811	0.616

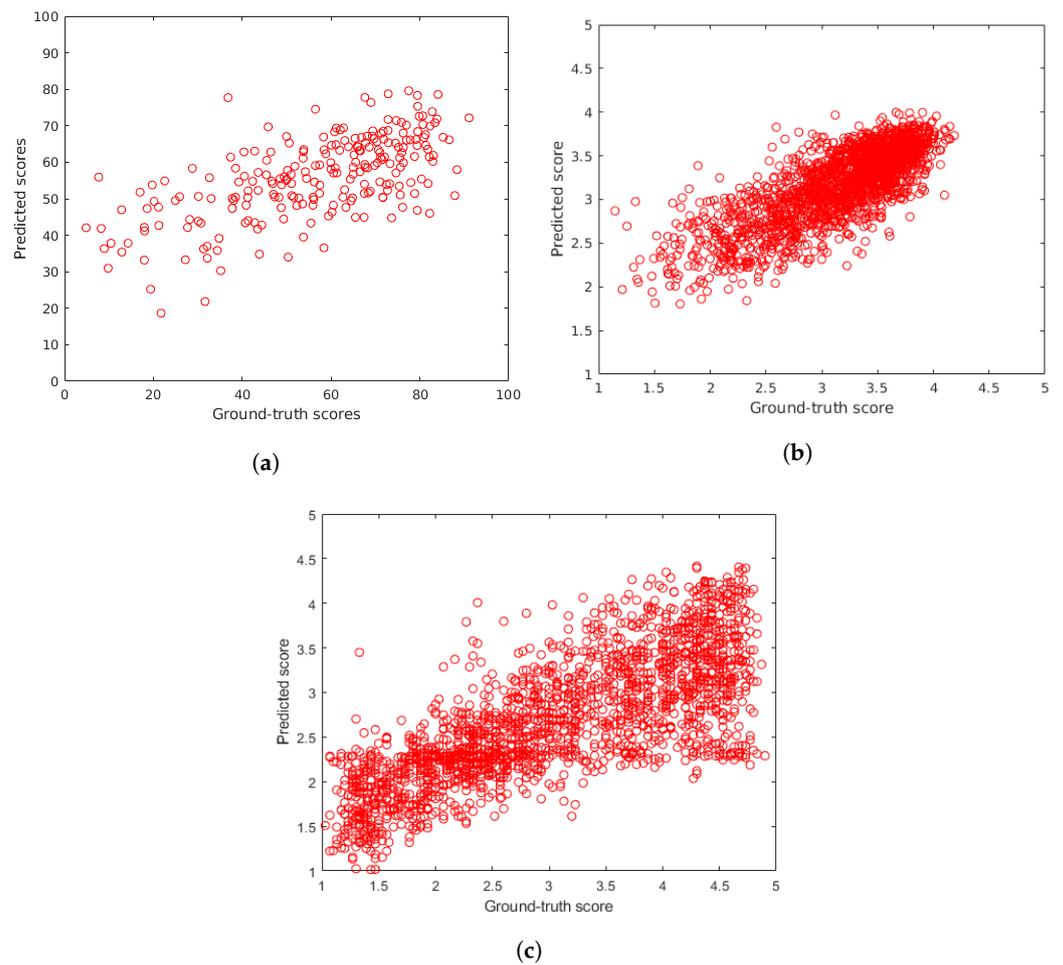


Figure 8. Ground-truth scores versus predicted scores using the proposed LFD-IQA-SVR method on (a) CLIVE [21], (b) KonIQ-10k [22], and (c) KADID-10k [25] test sets.

Table 6. Results of the two-sided Wilcoxon rank sum test. Symbol ‘1’ is used to denote that the proposed method—LFD-IQA-SVR—is significantly better than the method in the row on the database in the column.

Method	CLIVE [21]	KonIQ-10k [22]	SPAQ [23]	TID2013 [24]	KADID-10k [25]
BIQI [85]	1	1	1	1	1
BLIINDS-II [35]	1	1	1	1	1
BMPRI [86]	1	1	1	1	1
BRISQUE [36]	1	1	1	1	1
CurveletQA [37]	1	1	1	1	1
DIIVINE [38]	1	1	1	1	1
ENIQA [87]	1	1	1	1	1
GRAD-LOG-CP [69]	1	1	1	1	1
GWH-GLBP [70]	1	1	1	1	1
NBIQA [88]	1	1	1	1	1
OG-IQA [71]	1	1	1	1	1
Robust BRISQUE [90]	1	1	1	1	1
SSEQ [91]	1	1	1	1	1

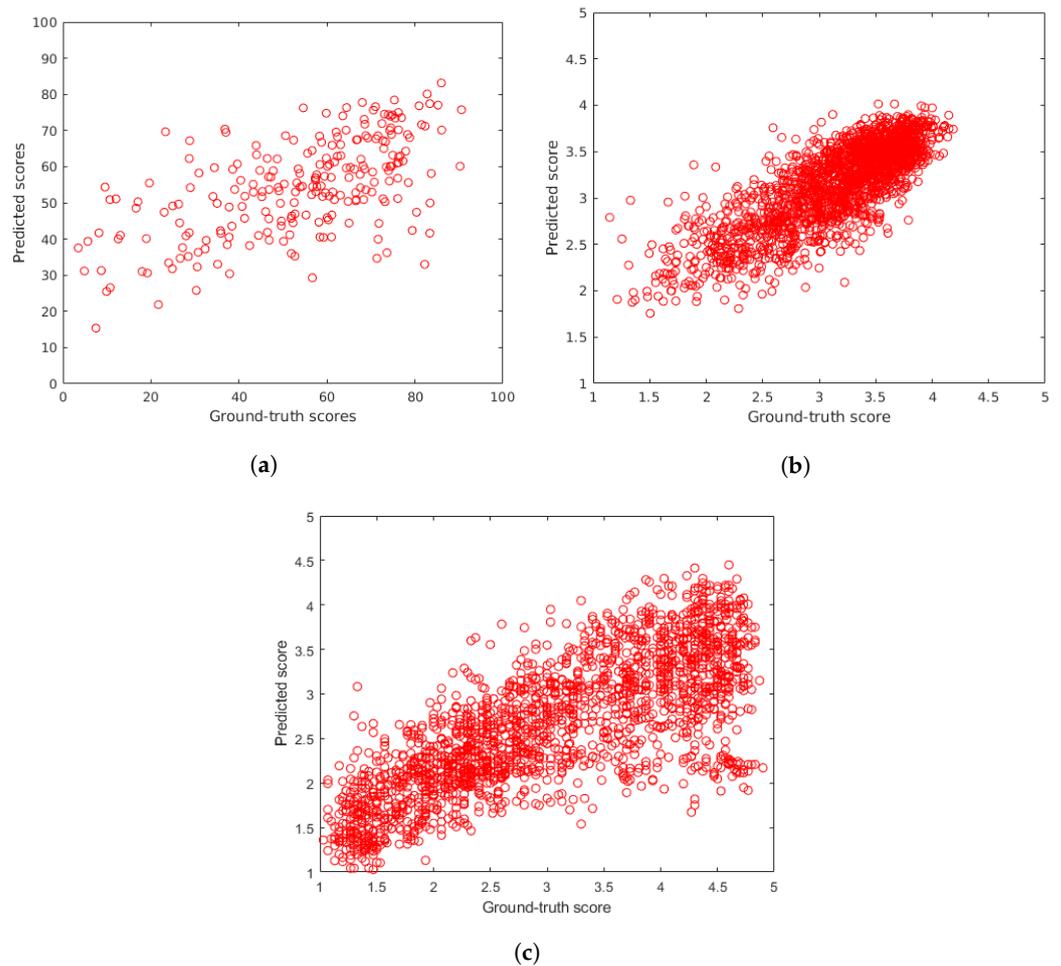


Figure 9. Ground-truth scores versus predicted scores using the proposed LFD-IQA-GPR method on (a) CLIVE [21], (b) KonIQ-10k [22], and (c) KADID-10k [25] test sets.

Table 7. Results of the two-sided Wilcoxon rank sum test. Symbol ‘1’ is used to denote that the proposed method—LFD-IQA-GPR—is significantly better than the method in the row on the database in the column.

Method	CLIVE [21]	KonIQ-10k [22]	SPAQ [23]	TID2013 [24]	KADID-10k [25]
BIQI [85]	1	1	1	1	1
BLIINDS-II [35]	1	1	1	1	1
BMPRI [86]	1	1	1	1	1
BRISQUE [36]	1	1	1	1	1
CurveletQA [37]	1	1	1	1	1
DIIVINE [38]	1	1	1	1	1
ENIQA [87]	1	1	1	1	1
GRAD-LOG-CP [69]	1	1	1	1	1
GWH-GLBP [70]	1	1	1	1	1
NBIQA [88]	1	1	1	1	1
OG-IQA [71]	1	1	1	1	1
Robust BRISQUE [90]	1	1	1	1	1
SSEQ [91]	1	1	1	1	1

In an other test, the generalization ability of the methods were examined. Namely, the algorithms were trained on the entire KonIQ-10k [22] database used as a training set and tested on the entire CLIVE [21] used as a test set. This process is called cross database test in the literature [93]. The results of the cross database are shown in Table 8. In this test, the

proposed methods are also the best performing ones. Namely, they are able to outperform the state-of-the-art by a large margin.

Table 8. Results of the cross database test. The examined and the proposed methods were trained on KonIQ-10k [22] and tested on CLIVE [21]. The best results are typed in bold, the second best results are underlined, and the third best results are typed in italic.

Method	PLCC	SROCC	KROCC
BIQI [85]	0.477	0.424	0.289
BLIINDS-II [35]	0.107	0.090	0.063
BMPRI [86]	0.453	0.389	0.298
BRISQUE [36]	0.509	0.460	0.310
CurveletQA [37]	0.496	0.505	0.347
DIIVINE [38]	0.479	0.434	0.299
ENIQA [87]	0.428	0.386	0.272
GRAD-LOG-CP [69]	0.427	0.384	0.261
GWH-GLBP [70]	0.480	0.479	0.328
NBIQA [88]	0.503	0.509	0.284
OG-IQA [71]	0.442	0.427	0.289
Robust BRISQUE [90]	0.516	0.481	0.327
SSEQ [91]	0.270	0.256	0.170
LFD-IQA-SVR	<u>0.567</u>	<u>0.561</u>	<u>0.390</u>
LFD-IQA-GPR	0.603	0.585	0.409

5. Conclusions

In this paper, a novel machine learning based NR-IQA method was introduced which applies an innovative quality-aware feature extraction procedure relying on the statistics of local feature descriptors. To be more specific, a sequence of HVS inspired filters were applied to Y , Cb , and Cr color channels of an input image to enhance those statistical regularities of the image to which the HVS is sensitive. Next, certain statistics of various local feature descriptors were extracted from each feature map to construct a powerful feature vector which is able to characterize possible image distortions from various points of view. Finally, the obtained feature vector is mapped onto perceptual quality scores with a trained regressor. The proposed method was compared to 16 state-of-the-art NR-IQA methods on five large benchmark IQA databases containing either authentic (CLIVE [21], KonIQ-10k [22], SPAQ [23]) or artificial (TID2013 [24], KADID-10k [25]) distortions. Specifically, the comparison involved the demonstration of three performance metrics on direct database tests, significance tests, and a cross database test. As shown, the proposed method is able to outperform significantly the state-of-the-art and provides competitive results. Future work involves a real-time GPU (graphical processing unit) implementation of the proposed method. Another direction of future research is to generalize the achieved results to other types of image modalities, such as stereoscopic or computer-generated images.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: In this paper, the following publicly available benchmark databases were used: 1. CLIVE: <https://live.ece.utexas.edu/research/ChallengeDB/index.html> (accessed on 16 April 2022); 2. KonIQ-10k: <http://database.mmsp-kn.de/koniq-10k-database.html> (accessed on 16 April 2022); 3. SPAQ: <https://github.com/h4nwei/SPAQ> (accessed on 16 April 2022); 4. KADID-10k: <http://database.mmsp-kn.de/kadid-10k-database.html> (accessed on 16 April 2022); 5. TID2013: <https://www.ponomarenko.info/tid2013.htm> (accessed on 16 April 2022). The source code of the proposed LFD-IQA no-reference image quality assessment method is available at: <https://github.com/Skythianos/LFD-IQA> (accessed on 1 September 2022).

Acknowledgments: We thank the academic editor and the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BRIEF	binary robust independent elementary features
BRISK	binary robust invariant scalable keypoints
CNN	convolutional neural network
DCT	discrete cosine transform
FAST	features from accelerated segment test
FR	full-reference
GPR	Gaussian process regressor
GPU	graphical processing unit
HVS	human visual system
IQA	image quality assessment
KADID	Konstanz artificially distorted image quality database
KROCC	Kendall rank order correlation coefficient
LIVE	laboratory for image and video engineering
NR	no-reference
NSS	natural scene statistics
ORB	oriented FAST and rotated BRIEF
PLCC	Pearson linear correlation coefficient
RR	reduced-reference
SPAQ	smartphone photography attribute and quality
SROCC	Spearman rank order correlation coefficient
SURF	speeded up robust features
SVR	support vector regressor
TID	Tampere image database

References

1. Williams, M.B.; Krupinski, E.A.; Strauss, K.J.; Breeden, W.K., III; Rzeszotarski, M.S.; Applegate, K.; Wyatt, M.; Bjork, S.; Seibert, J.A. Digital radiography image quality: Image acquisition. *J. Am. Coll. Radiol.* **2007**, *4*, 371–388. [[CrossRef](#)] [[PubMed](#)]
2. Ma, K.; Yeganeh, H.; Zeng, K.; Wang, Z. High dynamic range image compression by optimizing tone mapped image quality index. *IEEE Trans. Image Process.* **2015**, *24*, 3086–3097.
3. Flohr, T.; Stierstorfer, K.; Ulzheimer, S.; Bruder, H.; Primak, A.; McCollough, C. Image reconstruction and image quality evaluation for a 64-slice CT scanner with-flying focal spot. *Med. Phys.* **2005**, *32*, 2536–2547. [[CrossRef](#)]
4. Rahman, Z.; Jobson, D.J.; Woodell, G.A.; Hines, G.D. Image enhancement, image quality, and noise. In Proceedings of the Photonic Devices and Algorithms for Computing VII, San Diego, CA, USA, 31 July 2005; SPIE: Bellingham, WA, USA, 2005; Volume 5907, pp. 164–178.
5. Wang, Z. Applications of objective image quality assessment methods [applications corner]. *IEEE Signal Process. Mag.* **2011**, *28*, 137–142. [[CrossRef](#)]
6. Woodard, J.P.; Carley-Spencer, M.P. No-reference image quality metrics for structural MRI. *Neuroinformatics* **2006**, *4*, 243–262. [[CrossRef](#)]
7. Hung, S.C.; Wu, H.C.; Tseng, M.H. Integrating Image Quality Enhancement Methods and Deep Learning Techniques for Remote Sensing Scene Classification. *Appl. Sci.* **2021**, *11*, 11659. [[CrossRef](#)]
8. Lee, C.; Woo, S.; Baek, S.; Han, J.; Chae, J.; Rim, J. Comparison of objective quality models for adaptive bit-streaming services. In Proceedings of the 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), Larnaca, Cyprus, 27–30 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–4.
9. Torr, P.H.; Zissermann, A. Performance characterization of fundamental matrix estimation under image degradation. *Mach. Vis. Appl.* **1997**, *9*, 321–333. [[CrossRef](#)]
10. Chubarau, A.; Akhavan, T.; Yoo, H.; Mantiuk, R.K.; Clark, J. Perceptual image quality assessment for various viewing conditions and display systems. *Electron. Imaging* **2020**, *2020*, 67-1–67-9. [[CrossRef](#)]
11. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* **2020**, *63*, 211301. [[CrossRef](#)]
12. Shao, F.; Tian, W.; Lin, W.; Jiang, G.; Dai, Q. Toward a blind deep quality evaluator for stereoscopic images based on monocular and binocular interactions. *IEEE Trans. Image Process.* **2016**, *25*, 2059–2074. [[CrossRef](#)]

13. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
14. Freitas, P.G.; Da Eira, L.P.; Santos, S.S.; de Farias, M.C.Q. On the Application LBP Texture Descriptors and Its Variants for No-Reference Image Quality Assessment. *J. Imaging* **2018**, *4*, 114. [[CrossRef](#)]
15. Wu, J.; Lin, W.; Shi, G. Image quality assessment with degradation on spatial structure. *IEEE Signal Process. Lett.* **2014**, *21*, 437–440. [[CrossRef](#)]
16. Varga, D. No-Reference Quality Assessment of Authentically Distorted Images Based on Local and Global Features. *J. Imaging* **2022**, *8*, 173. [[CrossRef](#)] [[PubMed](#)]
17. Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE features. In *Computer Vision—ECCV 2012, Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 214–227.
18. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.
19. Liu, G.H.; Li, Z.Y.; Zhang, L.; Xu, Y. Image retrieval based on micro-structure descriptor. *Pattern Recognit.* **2011**, *44*, 2123–2133. [[CrossRef](#)]
20. Manjunath, B.S.; Ma, W.Y. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 837–842. [[CrossRef](#)]
21. Ghadiyaram, D.; Bovik, A.C. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. Image Process.* **2015**, *25*, 372–387. [[CrossRef](#)]
22. Lin, H.; Hosu, V.; Saupe, D. KonIQ-10K: Towards an ecologically valid and large-scale IQA database. *arXiv* **2018**, arXiv:1803.08489.
23. Fang, Y.; Zhu, H.; Zeng, Y.; Ma, K.; Wang, Z. Perceptual quality assessment of smartphone photography. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3677–3686.
24. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* **2015**, *30*, 57–77. [[CrossRef](#)]
25. Lin, H.; Hosu, V.; Saupe, D. KADID-10k: A large-scale artificially distorted IQA database. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–3.
26. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [[CrossRef](#)]
27. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [[CrossRef](#)] [[PubMed](#)]
28. Saad, M.A.; Bovik, A.C.; Charrier, C. A DCT statistics-based blind image quality index. *IEEE Signal Process. Lett.* **2010**, *17*, 583–586. [[CrossRef](#)]
29. Priya, K.M.; Channappayya, S.S. A novel sparsity-inspired blind image quality assessment algorithm. In Proceedings of the 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, USA, 3–5 December 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 984–988.
30. Leonardi, M.; Napoletano, P.; Schettini, R.; Rozza, A. No Reference, Opinion Unaware Image Quality Assessment by Anomaly Detection. *Sensors* **2021**, *21*, 994. [[CrossRef](#)] [[PubMed](#)]
31. Mishkin, D.; Sergievskiy, N.; Matas, J. Systematic evaluation of convolution neural network advances on the imagenet. *Comput. Vis. Image Underst.* **2017**, *161*, 11–19. [[CrossRef](#)]
32. Reinagel, P.; Zador, A.M. Natural scene statistics at the centre of gaze. *Netw. Comput. Neural Syst.* **1999**, *10*, 341. [[CrossRef](#)]
33. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [[CrossRef](#)]
34. Sheikh, H.R.; Bovik, A.C.; Cormack, L. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Trans. Image Process.* **2005**, *14*, 1918–1927. [[CrossRef](#)]
35. Saad, M.A.; Bovik, A.C.; Charrier, C. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.* **2012**, *21*, 3339–3352. [[CrossRef](#)]
36. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[PubMed](#)]
37. Liu, L.; Dong, H.; Huang, H.; Bovik, A.C. No-reference image quality assessment in curvelet domain. *Signal Process. Image Commun.* **2014**, *29*, 494–505. [[CrossRef](#)]
38. Moorthy, A.K.; Bovik, A.C. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Trans. Image Process.* **2011**, *20*, 3350–3364. [[CrossRef](#)]
39. Jenadeleh, M.; Moghaddam, M.E. BIQWS: Efficient Wakeby modeling of natural scene statistics for blind image quality assessment. *Multimed. Tools Appl.* **2017**, *76*, 13859–13880. [[CrossRef](#)]
40. Simoncelli, E.P.; Freeman, W.T.; Adelson, E.H.; Heeger, D.J. Shiftable multiscale transforms. *IEEE Trans. Inf. Theory* **1992**, *38*, 587–607. [[CrossRef](#)]
41. Bagade, J.V.; Singh, K.; Dandawate, Y.H. No reference image quality assessment with shape adaptive discrete wavelet features using neuro-wavelet model. *Multimed. Tools Appl.* **2022**, *81*, 31145–31160. [[CrossRef](#)]

42. Jenadeleh, M.; Masaeli, M.M.; Moghaddam, M.E. Blind image quality assessment based on aesthetic and statistical quality-aware features. *J. Electron. Imaging* **2017**, *26*, 043018. [[CrossRef](#)]
43. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3156–3164.
44. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [[CrossRef](#)]
45. El-Nouby, A.; Neverova, N.; Laptev, I.; Jégou, H. Training vision transformers for image retrieval. *arXiv* **2021**, arXiv:2102.05644.
46. Liu, W.; Liao, S.; Ren, W.; Hu, W.; Yu, Y. High-level semantic feature detection: A new perspective for pedestrian detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5187–5196.
47. Kang, L.; Ye, P.; Li, Y.; Doermann, D. Convolutional neural networks for no-reference image quality assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1733–1740.
48. Kim, J.; Lee, S. Fully deep blind image quality predictor. *IEEE J. Sel. Top. Signal Process.* **2016**, *11*, 206–220. [[CrossRef](#)]
49. Bare, B.; Li, K.; Yan, B. An accurate deep convolutional neural networks model for no-reference image quality assessment. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1356–1361.
50. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)] [[PubMed](#)]
51. Conde, M.V.; Burchi, M.; Timofte, R. Conformer and Blind Noisy Students for Improved Image Quality Assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 940–950.
52. Ayyoubzadeh, S.M.; Royat, A. (ASNA) An Attention-based Siamese-Difference Neural Network with Surrogate Ranking Loss function for Perceptual Image Quality Assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 388–397.
53. Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; Yang, F. MUSIQ: Multi-scale image quality transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 5148–5157.
54. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
55. Zhu, H.; Zhou, Y.; Shao, Z.; Du, W.L.; Zhao, J.; Yao, R. ARET-IQA: An Aspect-Ratio-Embedded Transformer for Image Quality Assessment. *Electronics* **2022**, *11*, 2132. [[CrossRef](#)]
56. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
57. Sun, S.; Yu, T.; Xu, J.; Zhou, W.; Chen, Z. GraphIQA: Learning distortion graph representations for blind image quality assessment. *IEEE Trans. Multimed.* **2022**. [[CrossRef](#)]
58. Liu, J.; Zhou, W.; Li, X.; Xu, J.; Chen, Z. LIQA: Lifelong blind image quality assessment. *IEEE Trans. Multimed.* **2022**, 1–16. [[CrossRef](#)]
59. Jenadeleh, M. Blind Image and Video Quality Assessment. Ph.D. Thesis; University of Konstanz: Konstanz, Germany, 2018.
60. Xu, L.; Lin, W.; Kuo, C.C.J. *Visual Quality Assessment by Machine Learning*; Springer: Singapore, 2015.
61. Zhou, W.; Chen, Z.; Li, W. Dual-stream interactive networks for no-reference stereoscopic image quality assessment. *IEEE Trans. Image Process.* **2019**, *28*, 3946–3958. [[CrossRef](#)]
62. Cui, Y.; Yu, M.; Jiang, Z.; Peng, Z.; Chen, F. Blind light field image quality assessment by analyzing angular-spatial characteristics. *Digit. Signal Process.* **2021**, *117*, 103138. [[CrossRef](#)]
63. Zhou, W.; Xu, J.; Jiang, Q.; Chen, Z. No-reference quality assessment for 360-degree images by analysis of multifrequency information and local-global naturalness. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1778–1791. [[CrossRef](#)]
64. Ribeiro, F.; Florencio, D.; Nascimento, V. Crowdsourcing subjective image quality evaluation. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 3097–3100.
65. Xu, Q.; Huang, Q.; Yao, Y. Online crowdsourcing subjective image quality assessment. In Proceedings of the 20th ACM International Conference on Multimedia, Nara, Japan, 29 October–2 November 2012; pp. 359–368.
66. Hosu, V.; Lin, H.; Saupé, D. Expertise screening in crowdsourcing image quality. In Proceedings of the 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), Cagliari, Italy, 28 May–1 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
67. Sheikh, H.R.; Bovik, A.C.; De Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128. [[CrossRef](#)] [[PubMed](#)]
68. Zhang, X.; Lin, W.; Xue, P. Just-noticeable difference estimation with pixels in images. *J. Vis. Commun. Image Represent.* **2008**, *19*, 30–41. [[CrossRef](#)]

69. Xue, W.; Mou, X.; Zhang, L.; Bovik, A.C.; Feng, X. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Trans. Image Process.* **2014**, *23*, 4850–4862. [[CrossRef](#)]
70. Li, Q.; Lin, W.; Fang, Y. No-reference quality assessment for multiply-distorted images in gradient domain. *IEEE Signal Process. Lett.* **2016**, *23*, 541–545. [[CrossRef](#)]
71. Liu, L.; Hua, Y.; Zhao, Q.; Huang, H.; Bovik, A.C. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Process. Image Commun.* **2016**, *40*, 1–15. [[CrossRef](#)]
72. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded up robust features. In *Computer Vision—ECCV 2006, Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 404–417.
73. Poynton, C.A. *A Technical Introduction to Digital Video*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1996.
74. Gerhard, H.E.; Wichmann, F.A.; Bethge, M. How sensitive is the human visual system to the local statistics of natural images? *PLoS Comput. Biol.* **2013**, *9*, e1002873. [[CrossRef](#)]
75. Marr, D.; Hildreth, E. Theory of edge detection. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1980**, *207*, 187–217.
76. Ghosh, K.; Sarkar, S.; Bhaumik, K. A possible mechanism of zero-crossing detection using the concept of the extended classical receptive field of retinal ganglion cells. *Biol. Cybern.* **2005**, *93*, 1–5. [[CrossRef](#)]
77. Ghosh, K.; Sarkar, S.; Bhaumik, K. Understanding image structure from a new multi-scale representation of higher order derivative filters. *Image Vis. Comput.* **2007**, *25*, 1228–1238. [[CrossRef](#)]
78. Patil, S.B.; Patil, B. Automatic Detection of Microaneurysms in Retinal Fundus Images using Modified High Boost Filtering, Line Detectors and OC-SVM. In *Proceedings of the 2020 International Conference on Industry 4.0 Technology (I4Tech)*, Pune, India, 13–15 February 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 148–153.
79. Li, Q.; Lin, W.; Fang, Y. No-reference image quality assessment based on high order derivatives. In *Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME)*, Seattle, WA, USA, 11–15 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
80. Huang, D.; Zhu, C.; Wang, Y.; Chen, L. HSOG: A novel local image descriptor based on histograms of the second-order gradients. *IEEE Trans. Image Process.* **2014**, *23*, 4680–4695. [[CrossRef](#)]
81. Rosten, E.; Drummond, T. Fusing points and lines for high performance tracking. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, Beijing, China, 17–21 October 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1–2. pp. 1508–1515.
82. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011*; IEEE: Piscataway, NJ, USA, 2011; pp. 2564–2571.
83. Harris, C.; Stephens, M. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, Manchester, UK, September 1988; The British Machine Vision Association and Society for Pattern Recognition, Durham, UK, 1988; Volume 15, pp. 10–5244.
84. Shi, J.; Tomasi. Good features to track. In *Proceedings of the 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 21–23 June 1994; IEEE: Piscataway, NJ, USA, 1994; pp. 593–600.
85. Moorthy, A.; Bovik, A. A modular framework for constructing blind universal quality indices. *IEEE Signal Process. Lett.* **2009**, *17*, 7.
86. Min, X.; Zhai, G.; Gu, K.; Liu, Y.; Yang, X. Blind image quality estimation via distortion aggravation. *IEEE Trans. Broadcast.* **2018**, *64*, 508–517. [[CrossRef](#)]
87. Chen, X.; Zhang, Q.; Lin, M.; Yang, G.; He, C. No-reference color image quality assessment: From entropy to perceptual quality. *EURASIP J. Image Video Process.* **2019**, *2019*, 77. [[CrossRef](#)]
88. Ou, F.Z.; Wang, Y.G.; Zhu, G. A novel blind image quality assessment method based on refined natural scene statistics. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 22–25 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1004–1008.
89. Venkatanath, N.; Praneeth, D.; Bh, M.C.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In *Proceedings of the 2015 Twenty First National Conference on Communications (NCC)*, Mumbai, India, 27 February–1 March 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6.
90. Mittal, A.; Moorthy, A.K.; Bovik, A.C. Making image quality assessment robust. In *Proceedings of the 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Pacific Grove, CA, USA, 4–7 November 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1718–1722.
91. Liu, L.; Liu, B.; Huang, H.; Bovik, A.C. No-reference image quality assessment based on spatial and spectral entropies. *Signal Process. Image Commun.* **2014**, *29*, 856–863. [[CrossRef](#)]
92. Zhang, Z.; Dai, G.; Liang, X.; Yu, S.; Li, L.; Xie, Y. Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment. *IEEE Access* **2018**, *6*, 11534–11543. [[CrossRef](#)]
93. Merdjanovska, E.; Rashkovska, A. Cross-Database Generalization of Deep Learning Models for Arrhythmia Classification. In *Proceedings of the 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, Opatija, Croatia, 27 September–1 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 346–351.