engineering proceedings

MDPI

Article

NR-IQA with Gaussian Derivative Filter, Convolutional Block Attention Module and Spatial Pyramid Pooling †

Jyothi sri Vadlamudi [1] and Sameeulla Khan Md [2,*]

[1] Electronics and Communication Engineering, Malla Reddy Engineering College, Hyderabad 500100, Telangana, India; jyothisriv@mrec.ac.in
[2] School of Electronics Engineering, VIT-AP University, Amaravati 522237, Andhra Pradesh, India
* Correspondence: sameeulla.k@vitap.ac.in
† Presented at the 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: https://sciforum.net/event/ecsa-11.

Abstract: Gaussian derivatives offer valuable capabilities for analyzing image characteristics such as structure, edges, texture, and features, which are essential aspects in the assessment of image quality. Present days Convolutional Neural Networks (CNN) gained its importance in all computer vision applications and also in image quality assessment domain. Because of these characteristics of gaussian derivative that performs a major role in assessing image quality, this work is carried by combining these characteristics with the CNNs to better extract the features for assesssing the quality of an image. While CNNs have demonstrated their ability to handle distortion effectively, they are limited in their capacity to capture features at different scales, making them inadequate in dealing with significant variations in object size. Consequently, the concept of spatial pyramid pooling (SPP) has been introduced to address this limitation in image quality assessment (IQA). SPP involves pooling the spatial feature maps from the highest convolutional layers into a feature representation of fixed length. Additionally, through the utilization of convolutional block attention module (CBAM) a module designed for the interpretation of images and local importance pooling (LIP) proposed method for No-reference image quality assessment has demonstrated improved accuracy, generalization, and efficiency compared to conventional (or) traditional IQA methods. .

Keywords: image quality assessment; no-reference; spatial pyramid pooling; local importance pooling; convolutional block attention module

Citation: Vadlamudi, J.s.; Md, S.K. NR-IQA with Gaussian Derivative Filter, Convolutional Block Attention Module and Spatial Pyramid Pooling. *Eng. Proc.* **2024**, *1*, 0. https://doi.org/

Academic Editor: Firstname Lastname

Published: 26 November 2024

1. Introduction

In the present era, digital images are commonly captured using a wide range of mobile cameras, subjected to compression utilizing both advanced and traditional methodologies [1,2], Which traverse diverse communication channels [3], and are stored across various devices. At every phase of image processing pipeline, there exists the potential for introducing unforeseen distortions, which can result in a degradation of perceptual quality. Consequently, the significance of image quality assessment (IQA) cannot be overstated, as it performs a significant role in overseeing image quality and ensuring the reliability of image processing procedures. IQA's principal aim is to quantitatively forecast the perceptual quality of digital images. Throughout the stages of content generation to consumption, digital images are prone to degradation. Distortions such as Gaussian white noise, Gaussian blur (GB), or blocking artifacts may be introduced at various stages, including image acquisition, transmission, storage, post-processing, or compression. An effective IQA algorithm plays a crucial role in objectively measuring quality of images retrieved from Internet and precisely evaluating the effectivness of image processing algorithms, encompassing tasks like image compression and super-resolution, based on human observer perspectives.

Image quality measures are categorized into three groups based on the presence of a reference or pristine image. While full-reference (FR) methods have the advantage of

having access to complete reference image, no-reference (NR) approaches lacks details about the reference image. Reduced-reference (RR) image quality assessment (IQA) lies between these two extremes since it only has access to collection of features extracted from reference image. As NR-IQA does not exploit any information about original image, it poses a more difficult challenge compared to FR-IQA and is considered one of the most demanding problems in IQA. NR-IQA finds wide-ranging applications since, in practical scenarios, reference images are often unavailable, making NR IQA particularly relevant.

2. Related Work

No-reference image quality assessment (NR-IQA) methods analyze image quality without a reference image by relying on natural image statistics, which tend to follow specific distributions. Distorted images deviate from these regularities. NIQE [4] assesses image quality by analyzing the deviation of mean subtracted contrast normalized (MSCN) coefficients from a Gaussian distribution. IL-NIQE extends this by using wavelet-based natural scene statistics (NSS) with support vector regression (SVR) to predict quality scores [5]. BLIINDS [6] uses DCT coefficients, while CDIIVINE [7] refines wavelet-based DIIVINE for better accuracy. BRISQUE [8] employs SVR in the spatial domain using NSS data.

Deep learning techniques in Image Quality Assessment (IQA) utilize deep neural networks to obtain visual features from an image. These features are subsequently employed to calculate a functional expression representing quality score of a distorted image. Harnessing, capabilities of deep neural networks, these methods can effectively capture complex patterns and relationships within the image data, enabling accurate and reliable quality estimation without relying on reference images. Kang [9] initiated CNN to create a IQA method. In their study, Kim [10] employed a technique based on image blocks to enhance the training dataset, this approach allows for improved learning and generalization capabilities of the model, ultimately leading to enhanced performance in image-related tasks.

3. Gaussian Filter and Gaussian Derivative

Gaussian filters are extensively employed as smoothing filters, with significant implications in edge detection within the human visual system. They are proved as highly valuable detectors for identifying edges and lines in various applications like noise supression, low-pass filtering, blurring etc. Further, Utilizing the derivatives of the Gaussian filter enables the simultaneous execution of noise reduction and edge detection. In traditional computer vision relying on manually designed image features, it has been shown that numerous visual tasks can be effectively tackled by computing image features and descriptors derived from Gaussian derivatives or their approximations as initial layer of image features [11–14]. It raises the question of whether Gaussian derivatives could serve as computational primitives for building deep networks.

The first-order Gaussian derivative provides insights into the variations in image intensity. It serves as a measure of the gradient, indicating the place at which pixel values alter and Second-order Gaussian derivatives provides details regarding the curvature or the rate of the gradient's change. In image processing, this connection is frequently associated with the identification of corners or locations with significant curvature.

The 1-D Gaussian function is defined as

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-x^2/2\sigma^2\right) \tag{1}$$

Here, σ represents the standard deviation, acting as a parameter that determines the width of the filter.

The 1st and 2nd order derivatives of 1-D Gaussian function are given as

$$G'(x) = \frac{\partial G(x)}{\partial x} = -\frac{x}{\sigma^2}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right) = -\frac{x}{\sigma^2}G(x) \tag{2}$$

$$\begin{aligned} G''(x) = \frac{\partial^2 G(x)}{\partial x^2} &= \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right) \\ &= \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right)G(x) \end{aligned} \tag{3}$$

A 2-D Gaussian filter is represented as follows

$$G(x,y) = \frac{1}{2\pi\sigma^2}e^{-\left(x^2+y2/2\sigma^2\right)} \tag{4}$$

The 1st and 2nd order derivatives of 2-D Gaussian function are given as

$$\begin{aligned} G'(x) &= \frac{\partial G(x,y)}{\partial x} = -\frac{x}{\sigma^2}G(x,y) \\ G'(y) &= \frac{\partial G(x,y)}{\partial y} = -\frac{y}{\sigma^2}G(x,y) \end{aligned} \tag{5}$$

$$G_{xx}(x,y;\sigma) \tag{6}$$

$$\begin{aligned} G_{xx}(x,y) &= \frac{\partial^2(G(x,y))}{\partial x^2} = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right)G(x,y) \\ G_{yy}(x,y) &= \frac{\partial^2(G(x,y))}{\partial y^2} = \left(\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2}\right)G(x,y) \end{aligned} \tag{7}$$

4. Spatial Pyramid Pooling Layer

Convolutional layers can accept inputs of any size, but their output sizes vary, while classifiers like SVM or softmax, and fully-connected layers, require fixed-length vectors. A common solution is the Bag-of-Words (BoW) approach [15], which pools features to generate fixed-length vectors. Spatial Pyramid Pooling (SPP) [16,17] extends BoW by pooling within local spatial bins, preserving spatial information and adapting to different image sizes.

To handle variable-sized images in deep networks, we replace the last pooling layer after the final two convolutional layers with an SPP layer. Within each bin, max pooling is performed on filter responses. This produces fixed-length vectors of size kM, where M is the number of bins and k is the number of filters. These vectors are then fed into the fully-connected layer, allowing the network to process images of any size while maintaining a consistent input for classification.

5. Convolutional Block Attention Module (CBAM)

CBAM [18] is a method designed for both channel and spatial dimensions to emphasize the relevant features. It consists of two modules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). These modules operate sequentially and serve distinct purposes. CAM focuses on learning the important content along channel axis, while SAM concentrates on capturing positional information along the spatial axis. By incorporating both modules, CBAM effectively highlights target features by attending to both channel-specific details and spatial relationships.

When provided with an intermediate feature map $F_m \in R^{C\times H\times W}$ as input (where C,H,W are dimension, height, width of the channel respectively), CBAM performs a two-step attention calculation. First, it caluculates channel attention map of 1-dimensional.

$M_c \in R^{C\times1\times1}$ followed by a spatial attention map of 2-dimensional $M_s \in R^{1\times H\times W}$. Attention procedure is represented as below.

$$F'_m = M_c(F_m) \otimes F_m \tag{8}$$

$$F''_m = M_s\left(F'_m\right) \otimes F'_m \tag{9}$$

The element-wise multiplication operation, denoted by $\otimes$, is used to perform the calculations in channel and spatial attention modules denoted by $M_c(\cdot)$ and $M_s(\cdot)$ respectively. Process of execution of CBAM can be summarized as follows:

- The input feature map F_m is multiplied by CAM, resulting in intermediate output F'_m.
- The refined feature F'_m is then forwarded as input to the SAM, which calculates the feature map F''_m.

Channel attention module: It accepts input from both average & max-pooled features. Next, these characteristics are fed into a Multilayer Perceptron (MLP) that shares weights among its layers. Resulting feature vectors are integrated through element-wise summation. Primary objective of CAM is to compensate for the limitations of channel attention. CAM is mathematically represented as follows:

$$M_c(F_m) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F_m)) + \mathrm{MLP}(\mathrm{MaxPool}(F_m))) \tag{10}$$

where σ denotes sigmoid activation function, MLP symbolizes shared fully connected layer, AvgPool denotes operation of average pooling, MaxPool denotes operation of maximum pooling, & '+' represents the element-wise addition.

Spatial attention module: Steps in calculating spatial attention involves using of average & max pooling operations across channel axis. Following that, the results are combined to provide an efficient feature descriptor. A spatial attention map is created by utilizing a convolution layer on a combined feature description.

$$\mathbf{M_s}(\mathbf{F_m}) \in \mathbf{R}^{H\times W} \tag{11}$$

by encoding instructions on what to highlight or diminish. Channel information from a feature map is gathered through the application of these two operations on pooling, resulting in the creation of two 2D maps:

$$\mathbf{F_m^s}_{\mathrm{avg}} \in \mathbb{R}^{1\times H\times W} \text{ and } \mathbf{F_m^s}_{\mathrm{max}} \in \mathbb{R}^{1\times H\times W} \tag{12}$$

One represents average-pooled features, while the other represents max-pooled features across the channel. These two sets are then combined and subjected to convolution through a convolutional layer, yielding 2D spatial attention map. SAM is formulated as below:

$$M_s\left(F'_m\right) = \sigma\left(\mathrm{conv}^{7\times7}\left(\left[\mathrm{AvgPool}\left(F'_m\right); \mathrm{MaxPool}\left(F'_m\right)\right]\right)\right) \tag{13}$$

where σ is sigmoid activation function and $conv^{7\times7}$ is convolutional kernel of 7×7.

6. Local Importance Pooling

To streamline non-local operations while preserving key characteristics, various subsampling methods like average and max pooling reduce computation and memory usage. However, in pixel-level tasks like IQA, inappropriate pooling can lead to loss of critical details, weakening the effectiveness of non-local strategies. This study aims to balance accuracy and computational efficiency in non-local enhancement by employing an adaptive down-sampling method. Inspired by Local Importance Pooling (LIP) [19], the proposed approach introduces a simplified non-local strategy for IQA. During training, LIP-based non-local blocks enhance discriminative features and discard irrelevant ones, making it

particularly effective for tasks requiring fine detail while reducing computational costs. Following are the main contributions of proposed work.

- Proposed a Gaussian Derivative Convolutional Layer to extract features such as texture,edge and structural content that are useful for Image quality assessment and introduced spatial pyramid pooling to overcome the limitations imposed by the size constraints of the features.
- In the proposed model we have introduced CBAM an attention module to effectively highlight features by considering both channel and spatial relationships and utilized local importance pooling in place of Max pooling and average pooling to enhance features while discarding irrelevant ones. Proposed method is shown in Figure 1.

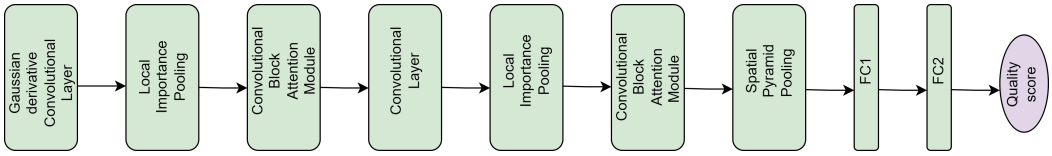

Figure 1. Proposed Architecture.

7. Materials and Methods

This section offers a comprehensive outline about databases & experimental results.

7.1. IQA Databases and Evaluation Metrics

Proposed method is verified on various databases LIVE [20], LIVE Wild [21], CSIQ [22], LIVEMD [23], IVC [24] and MDID [25].

7.2. Evalution Metrics

Two standard assessment metrics are utilized by video quality experts group (VQEG) [26]. Spearman rank order coefficient (SROCC) & Pearson liner correlation coefficient (PLCC) are utilized to analyse proposed method on synthetic, authentic and waterloo 3D phase-II databases.

SROCC: A non-parametric metric as described below

$$SROCC = 1 - \frac{6\sum_i p_i^2}{m(m.^2 - 1)} \tag{14}$$

m symbolizes a number of samples and p_i signifies distinction between ith image's ranking within the subjective and objective evaluations, respectively.

PLCC: Another non-parametric metric as described below.

$$PLCC = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(p_i - \overline{p})}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2(p_i - \overline{p})^2}}, \tag{15}$$

where p_i symbolizes subjective score (MOS or DMOS) of ith test image, and $\bar{p}$ and $\bar{Q}$ denote average values of predicted and subjective scores respectively.

8. Expermental Results

The following section discusses the proposed method's performance on individual databases and individual distortions and cross database validation.

8.1. Performance on Individual Databases

Proposed method is estimated against various state-of-art metrics like dipIQA [27], MEON [28], LPIPS [29], MetaIQA [30], CONTRIQUE [31], GraphIQA [32], Deepsim [33], UNIQUE [34]. SROCC and PLCC on different databases is shown in Tables 1–10. Metrics

that are performing well are highlighted. From Tables 1, 5, 9 and 10 it can be seen that for LIVE and IVC database proposed method performed better for SROCC and performance is good for CSIQ, LIVE Wild, LIVE-MD and IVC for PLCC. Further the scatter plots curve is displayed in Figure 2, which shows high correlation among the scores predicted by proposed method and subjective scores.

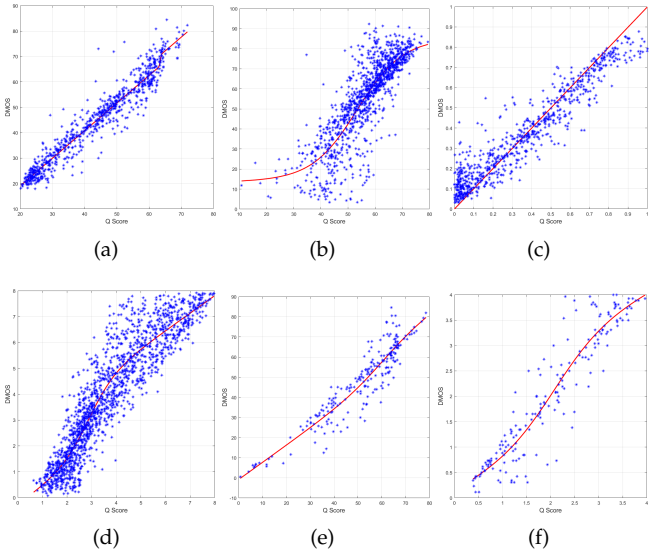

Figure 2. Scatter plots of quality scores, predicted by NiN vs. Subjective scores(MOS/DMOS) of (**a**) LIVE (**b**) LIVE Wild (**c**) CSIQ (**d**) MDID-2016 (**e**) LIVE-MD and (**f**) IVC databases.

Table 1. SROCC on LIVE Database.

Method	**LIVE [20]**
CONTRIQUE [31]	0.961
Deepsim [33]	0.968
dipIQA [27]	0.958
GraphIQA [32]	**0.980**
LPIPS [29]	0.934
MEON [28]	0.954
UNIQUE [34]	**0.968**
Proposed	0.964

Table 2. SROCC on CSIQ Database.

Method	**CSIQ [22]**
CONTRIQUE [31]	**0.958**
Deepsim [33]	0.919
dipIQA [27]	0.949
GraphIQA [32]	**0.959**
LPIPS [29]	0.896
MEON [28]	0.944
UNIQUE [34]	0.927
Proposed	0.951

Table 3. SROCC on LIVE Wild Database.

Method	LIVE Wild
GraphIQA [32]	**0.862**
MEON [28]	0.693
MetaIQA [30]	0.835
UNIQUE [34]	**0.890**
Proposed	0.841

Table 4. SROCC on LIVE-MD Database.

Method	LIVE-MD
GraphIQA [32]	**0.940**
MEON [28]	**0.935**
Proposed	0.929

Table 5. PLCC on LIVE Database.

Method	LIVE
CONTRIQUE [31]	0.960
Deepsim [33]	**0.974**
dipIQA [27]	0.941
LPIPS [29]	0.932
MEON [28]	0.943
UNIQUE [34]	**0.969**
Proposed	0.963

Table 6. PLCC on CSIQ Database.

Method	CSIQ
CONTRIQUE [31]	**0.947**
Deepsim [33]	0.919
dipIQA [27]	0.930
LPIPS [29]	0.876
MEON [28]	0.932
UNIQUE [34]	0.902
Proposed	**0.935**

Table 7. PLCC on LIVE Wild Database.

Method	LIVE Wild
MEON [28]	0.688
MetaIQA [30]	0.802
UNIQUE [34]	**0.854**
Proposed	**0.844**

Table 8. PLCC on LIVE-MD Database.

Method	LIVE-MD
MEON [28]	**0.920**
Proposed	**0.916**

Table 9. SROCC on IVC Database.

Method	IVC
LPIPS [29]	**0.809**
Proposed	**0.940**

Table 10. PLCC on IVC Database.

Method	IVC
LPIPS [29]	**0.822**
Proposed	**0.930**

8.2. Performance on Individual Distortions

Robustness of proposed model is assesses by conducting analysis over individual distortions of LIVE and CSIQ databases with state-of-art metrics like BLINDS-II [6], BLISS [35], BRISQUE [8], CNN [9], CORNIA [36], DB-CNN [37], dipIQ [27], FRIQUEE [38], HOSA [39], M3 [40], MEON [28], TWO-STREAM [41]. With the help of Tables 11–14, proposed method showed improved performance in most of the individual distortions.

Table 11. SROCC for individual distortions in the LIVE database.

Method	FF	JPEG	WN	JP2K	Gblur
BLIINDS-II [6]	0.927	0.942	0.978	0.951	0.944
BLISS [35]	0.914	0.929	0.903	0.927	0.952
BRISQUE [8]	0.828	0.965	0.982	0.929	0.964
CNN [9]	0.908	**0.977**	0.978	0.952	0.962
CORNIA [36]	0.921	0.947	0.958	0.924	0.951
DB-CNN [37]	0.930	**0.972**	0.980	0.955	0.935
dipIQ [27]	**N/A.**	0.969	0.975	0.956	0.940
FRIQUEE [38]	0.884	0.947	0.983	0.919	0.937
HOSA [39]	**0.954**	0.954	0.975	0.935	0.954
M3 [40]	0.902	0.966	**0.986**	0.930	0.935
MEON [28]	0.926	0.951	0.972	0.914	0.944
SOM [42]	0.937	0.952	0.984	0.947	**0.976**
TWO-STREAM [41]	0.911	0.950	0.979	**0.966**	0.963
Proposed	**0.942**	0.957	**0.985**	**0.962**	**0.967**

Table 12. PLCC for individual distortions in the LIVE database.

Method	FF	JPEG	WN	JP2K	Gblur
BLIINDS-II [6]	0.944	0.959	0.985	0.956	0.948
BLISS [35]	0.916	0.930	0.948	0.929	0.969
BRISQUE [8]	0.894	0.971	0.989	0.940	0.965
CNN [9]	0.933	**0.981**	0.984	0.953	0.953
CORNIA [36]	0.943	0.962	0.974	0.944	0.961
DB-CNN [37]	0.961	**0.986**	0.988	**0.967**	0.956
dipIQ [27]	**N/A.**	0.980	0.983	**0.969**	0.948
FRIQUEE [38]	0.936	0.955	0.991	**0.935**	0.949
HOSA [39]	**0.967**	0.967	0.983	0.949	0.967
M3 [40]	0.920	0.977	**0.992**	0.945	0.947
MEON [28]	0.936	0.968	0.982	0.923	0.929
SOM [42]	0.954	0.961	**0.991**	0.952	**0.974**
TWO-STREAM [41]	0.949	0.963	0.995	0.966	0.950
Proposed	**0.949**	0.974	0.982	0.966	**0.970**

Table 13. SROCC for individual distortions in the CSIQ database.

Method	WN	Gblur	JP2K	JPEG	CC
BRISQUE [8]	0.723	0.820	0.840	0.806	0.804
CORNIA [36]	0.664	0.860	0.831	0.513	0.462
DB-CNN [37]	**0.948**	**0.947**	**0.940**	**0.953**	**0.870**
dipIQ [27]	0.904	**0.932**	**0.944**	**0.936**	N/A.
FRIQUEE [38]	0.748	0.870	0.846	0.869	0.838
HOSA [39]	0.604	0.841	0.818	0.733	0.716
M3 [40]	0.741	0.868	0.911	0.740	0.770
MEON [28]	**0.951**	0.918	**0.948**	**0.948**	N/A.
Proposed	0.908	0.927	0.945	0.939	0.935

Table 14. PLCC for individual distortions in the CSIQ database.

Method	WN	Gblur	JP2K	JPEG	CC
BRISQUE [8]	0.742	0.891	0.887	0.828	0.835
CORNIA [36]	0.687	0.904	0.883	0.563	0.543
DB-CNN [37]	**0.956**	**0.969**	**0.982**	0.971	**0.895**
dipIQ [27]	0.927	**0.958**	0.959	**0.975**	N/A.
FRIQUEE [38]	0.778	0.905	0.883	0.885	0.864
HOSA [39]	0.656	0.912	0.899	0.759	0.744
M3 [40]	0.728	0.917	0.928	0.768	0.787
MEON [28]	**0.958**	0.946	**0.979**	0.925	N/A.
Proposed	0.914	0.940	0.964	0.987	0.938

9. Conclusions

Proposed method demonstrates superior performance in handling both synthetic and authentic distortions within Image Quality Assessment (IQA) tasks. Specifically, it has achieved remarkable results on the LIVE and IVC databases, with Spearman's Rank Order Correlation Coefficient (SROCC) values of 0.964 and 0.940, respectively, indicating a strong correlation among predicted and ground truth quality scores. Furthermore, proposed method excelled in evaluating image quality across various databases, including CSIQ, LIVE Wild, LIVE-MD, and IVC, achieving Pearson Linear Correlation Coefficient (PLCC) values of 0.935, 0.844, 0.916, and 0.930, respectively. These results highlight the robustness and effectiveness of the proposed approach in accurately assessing image quality across different types of distortions and diverse datasets.

Author Contributions: Conceptualization, J.s.V.; project administration, supervision, Validation, S.K.M.; writing—original draft, J.s.V.; writing—review and editing, S.K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

IQA	Image Quality Assessment
GB	Gaussian Blur
NR-IQA	No-reference image quality assessment
FR-IQA	Full-Reference image quality assessment
MSCN	mean subtracted and contrast normalized
NSS	Natural Scene Statistics
DCT	Discrete Cosine Transform
CNN	Convolutional neural networks
CBAM	Convolutional Block Attention Module
LIP	local importance based pooling

References

1. Bovik, A.C. *Handbook of Image and Video Processing*; Academic Press: London, UK, 2010.
2. Ballé, J.; Laparra, V.; Simoncelli, E.P. End-to-end optimized image compression. *arXiv* **2016**, arXiv:1611.01704.
3. Duanmu, Z.; Ma, K.; Wang, Z. Quality-of-experience of adaptive video streaming: Exploring the space of adaptations. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1752–1760.
4. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212.
5. 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.
6. 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.
7. Zhang, Y.; Moorthy, A.K.; Chandler, D.M.; Bovik, A.C. C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes. *Signal Process. Image Commun.* **2014**, *29*, 725–747.
8. 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.
9. 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.
10. Kim, J.; Lee, S. Fully deep blind image quality predictor. *IEEE J. Sel. Top. Signal Process.* **2016**, *11*, 206–220.
11. Lindeberg, T. Feature detection with automatic scale selection. *Int. J. Comput. Vis.* **1998**, *30*, 79–116.
12. Mikolajczyk, K.; Schmid, C. Scale & affine invariant interest point detectors. *Int. J. Comput. Vis.* **2004**, *60*, 63–86.

13. Lindeberg, T. Image matching using generalized scale-space interest points. *J. Math. Imaging Vis.* **2015**, *52*, 3–36.
14. Schiele, B.; Crowley, J.L. Recognition without correspondence using multidimensional receptive field histograms. *Int. J. Comput. Vis.* **2000**, *36*, 31–50.
15. Sivic.; Zisserman. Video Google: A text retrieval approach to object matching in videos. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Washington, DC, USA, 13–16 October 2003; pp. 1470–1477.
16. Grauman, K.; Darrell, T. The pyramid match kernel: Discriminative classification with sets of image features. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, China, 17–21 October 2005; Volume 2, pp. 1458–1465.
17. Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 2169–2178.
18. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
19. Gao, Z.; Wang, L.; Wu, G. Lip: Local importance-based pooling. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3355–3364.
20. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451.
21. Ghadiyaram, D.; Bovik, A.C. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. Image Process.* **2015**, *25*, 372–387.
22. Larson, E.C.; Chandler, D.M. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* **2010**, *19*, 011006–011006.
23. Jayaraman, D.; Mittal, A.; Moorthy, A.K.; Bovik, A.C. Objective quality assessment of multiply distorted images. 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; pp. 1693–1697.
24. Le Callet, P.; Autrusseau, F. Subjective quality assessment IRCCyN/IVC database. 2005.
25. Sun, W.; Zhou, F.; Liao, Q. MDID: A multiply distorted image database for image quality assessment. *Pattern Recognit.* **2017**, *61*, 153–168.
26. Antkowiak, J.; Baina, T.J.; Baroncini, F.V.; Chateau, N.; FranceTelecom, F.; Pessoa, A.C.F.; Colonnese, F.S.; Contin, I.L.; Caviedes, J.; Philips, F. Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000. 2000.
27. Ma, K.; Liu, W.; Liu, T.; Wang, Z.; Tao, D. dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Trans. Image Process.* **2017**, *26*, 3951–3964.
28. Ma, K.; Liu, W.; Zhang, K.; Duanmu, Z.; Wang, Z.; Zuo, W. End-to-end blind image quality assessment using deep neural networks. *IEEE Trans. Image Process.* **2017**, *27*, 1202–1213.
29. 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, 18–22 June 2018; pp. 586–595.
30. Zhu, H.; Li, L.; Wu, J.; Dong, W.; Shi, G. MetaIQA: Deep meta-learning for no-reference image quality assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 14143–14152.
31. Madhusudana, P.C.; Birkbeck, N.; Wang, Y.; Adsumilli, B.; Bovik, A.C. Image quality assessment using contrastive learning. *IEEE Trans. Image Process.* **2022**, *31*, 4149–4161.
32. Sun, S.; Yu, T.; Xu, J.; Zhou, W.; Chen, Z. GraphIQA: Learning distortion graph representations for blind image quality assessment. *IEEE Trans. Multimed.* **2022**, *25*, 2912–2925.
33. Gao, F.; Wang, Y.; Li, P.; Tan, M.; Yu, J.; Zhu, Y. Deepsim: Deep similarity for image quality assessment. *Neurocomputing* **2017**, *257*, 104–114.
34. Zhang, W.; Ma, K.; Zhai, G.; Yang, X. Uncertainty-aware blind image quality assessment in the laboratory and wild. *IEEE Trans. Image Process.* **2021**, *30*, 3474–3486.
35. Ye, P.; Kumar, J.; Doermann, D. Beyond human opinion scores: Blind image quality assessment based on synthetic scores. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4241–4248.
36. Ye, P.; Kumar, J.; Kang, L.; Doermann, D. Unsupervised feature learning framework for no-reference image quality assessment. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1098–1105.
37. Zhang, W.; Ma, K.; Yan, J.; Deng, D.; Wang, Z. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *30*, 36–47.
38. Ghadiyaram, D.; Bovik, A.C. Perceptual quality prediction on authentically distorted images using a bag of features approach. *J. Vis.* **2017**, *17*, 32–32.
39. Xu, J.; Ye, P.; Li, Q.; Du, H.; Liu, Y.; Doermann, D. Blind image quality assessment based on high order statistics aggregation. *IEEE Trans. Image Process.* **2016**, *25*, 4444–4457.

40. 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.
41. Yan, Q.; Gong, D.; Zhang, Y. Two-stream convolutional networks for blind image quality assessment. *IEEE Trans. Image Process.* **2018**, *28*, 2200–2211.
42. Zhang, P.; Zhou, W.; Wu, L.; Li, H. SOM: Semantic obviousness metric for image quality assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2394–2402.