

No-reference image quality assessment based on visual explanation images and deep transfer learning

Basma Ahmed¹, Osama A. Omer², Vivek Kumar Singh³, Amal Rashed¹, Mohamed Abdel-Nasser²

¹Faculty of Computers and Information, South Valley University, Qena, Egypt

²Department of Electrical Engineering, Faculty of Engineering, Aswan University, Aswan, Egypt

³Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

Article Info

Article history:

Received Jun 19, 2024

Revised Aug 23, 2024

Accepted Aug 31, 2024

Keywords:

Deep learning

Grad-CAM

Image quality assessment

Pseudo-reference

Similarity measures

Visual interpretations

ABSTRACT

Quantifying image quality in the absence of a reference image continues to be a challenge despite the introduction of numerous no-reference image quality assessments (NR-IQA) in recent years. Unlike most existing NR-IQA methods, this paper proposes an efficient NR-IQA method based on deep visual interpretations. Specifically, the main components of the proposed method are: i) generating a pseudo-reference image (PRI) for the input distorted images, ii) employing a pretrained convolutional network to extract feature maps from the distorted image and the corresponding PRI, iii) producing visual explanation images (VEIs) by using the feature maps of the distorted image and the corresponding PRI, iv) measuring the similarity between the two VEIs using an image similarity metric, and v) employing a non-linear mapping function for quality score alignment. In our experiments, we evaluated the efficacy of the proposed method across various forms of distortion using four benchmark datasets (LIVE, SIQAD, CSIQ, and TID2013). The proposed approach demonstrates parity with the latest methods, as evidenced by comparisons with both hand-crafted NR-IQA and deep learning-based approaches.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mohamed Abdel-Nasser

Department of Electrical Engineering, Faculty of Engineering, Aswan University

Aswan, Egypt

Email: mohamed.abdelnasser@aswu.edu.eg

1. INTRODUCTION

Indeed, 2-dimensional (2D) multimedia technologies with their applications, notably communications, education, and industry, are a key element for human visualization. However, digital images show low perceived visual quality for some capturing settings, notably noise sensitivity, aperture, lighting, exposure, and lens limitations. These settings can create distracting image artifacts that negatively impact the related human visual quality. The implementation of an automatic image quality assessment (IQA) algorithm to forecast the digital images' quality can facilitate the management of substandard images or the correction of their quality while they are being captured. Recently, numerous IQA algorithms have been suggested. Subjective and objective IQA algorithms can be distinguished by the degree to which human eyes evaluate the rankings. Based on the existence or lack thereof of a reference image, objective assessment can be essentially categorized into three groups: i) full reference IQA (FR-IQA), ii) no reference or reference-free IQA (NR-IQA), and iii) reduced reference IQA (RR-IQA) [1]–[4].

Indeed, regardless of the absence or presence of a reference image during the quality assessment procedure, the fundamental concern of IQA is to quantify the “distance” between the reference and distorted images. It is worth noting that the term “distance” refers to the index that is being assessed (e.g. mean square

error and signal-to-noise ratio). It should be noted that NR-IQA algorithms operate under the assumption that assessing image quality does not require a direct comparison between the original and test images. Thus, in situations where the original image is unavailable, they may be utilized [1]. Despite the existence of numerous NR-IQA approaches, only a limited number of studies have focused on an additional severe quality evaluation, namely the degree of degradation in an image's quality. An NR-IQA method was proposed by Li *et al.* [5] to handle fuzzy images. This method used the original blurred image as a pseudo-reference image (PRI) and imposed additional blur degradation on it. The final evaluation score was calculated as a fuzzy index between the two images. One of the main limitations of this approach is that it solely accounted for a solitary hazy distortion. Additionally, Min *et al.* [6] proposed three distortion-specific metrics based on PRI to estimate blockiness, sharpness, and noisiness, respectively. In order to generate the PRI, a degradation was applied to the distorted image for every form of distortion. Next, the PRI was used as a new benchmark to determine the "distance" between the PRI and the distorted image, and each estimating model used this distance as the quality score prediction. Zhan and Zhang [7] proposed an effective NR-IQA method for assessing the quality of JPEG images based on computing the luminance and blockiness in 8×8 blocks in a single JPEG image. One of the main limitations of this method is that it has been customized for JPEG images and it may give limited performance with other kinds of images. Oszust [8], introduced a novel NR-IQA approach by characterizing the distorted image by the distribution of the orientations of local gradients in regions of different sizes. With the aim of improving the performance of NR-IQA methods, researchers have recently employed deep learning as the core technology to address the challenges of quality assessment. Most deep learning-based methods have employed deep convolutional neural networks (CNNs) for feature extraction from input images to assess their quality [9]–[12]. Ravela *et al.* [13], initially determined the nature of the image distortion. Then, for each distinct form of distortion, they predicted the perceived degradation of image quality utilizing CNN. Pan *et al.* [14] introduced a distortion-aware CNN model called DACNN for blind IQA. The DACNN method demonstrated efficacy in handling genuine image distortions as well as those induced artificially. Zhou *et al.* [15], introduced a novel NR-IQA approach utilizing an attentional feature fusion mechanism. In this approach, thickness, local, and global information were extracted from the images, correspondingly, using three CNNs networks, namely VGG19, VGG16, and ResNet-50. Then, these features were fused to facilitate the perception of various types of distortions through the incorporation of attention mechanisms. Lin and Wang [16] used generative adversarial networks (GANs) for NR-IQA. The generative network was employed to create a reference image that was free of distortion for input image that was distorted. The features derived from the distorted image were subsequently combined with the information extracted from the hallucinated reference image in order to forecast the quality of the perceptual image. Other studies in literature have employed pre-trained CNNs like AlexNet or VGG16 as feature extractors. Bianco *et al.* [2] demonstrated the efficacy of a fine-tuning pre-trained CNN model to extract feature vectors from random regions of an input image. Gao *et al.* [17], on the other hand, utilized global minimum and maximum pooling to extract features that were not dependent on resolution from multiple layers of an AlexNet model. Like [2], the perceptual quality was ascertained through the arithmetic mean of the subscores obtained from the layer-wise feature vectors and a trained support vector regressor. Also, Varga [18] proposed a NR-IQA method based on attaching global average pooling layers to Inception modules of CNN network pre-trained on ImageNet dataset.

In general, most related works exhibit some limitations, such as a lack of comprehensive evaluation, dependence on specific assumptions, and limited interpretability in deep learning-based methods. In contrast to all current NR-IQA methods, we present an effective NR-IQA method in this paper utilizing deep visual interpretations. In particular, we propose extracting deep visual explanations from the degraded image and the corresponding PRI and then measuring the distance between the visual interpretations of the input image and its PRI using a full reference image quality assessment metric (e.g., signal-to-noise ratio). The similarity score that is obtained serves as a metric to assess the quality of the input images. The Grad-CAM [19] method is utilized to extract deep visual interpretations. In this paper, we assume that robust methods, like Grad-CAM, for interpreting deep learning models will produce the same visual interpretation for two identical images, and the similarity between the two visual interpretations of two identical images will be maximum. Therefore, measuring the similarity between the visual interpretations of the degraded image and its PRI can be a suitable way of assessing NR image quality. Up to the authors' knowledge, very few works have been proposed in the literature that consider visual interpretations for NR-IQA. Primary results have been presented in [20]; however, in this paper, we present a detailed study of various distortion types and many datasets. A comparative study with both traditional and deep learning NR-IQA methods is provided.

Section 2 explains the main steps of the proposed NR-IQA algorithm. Section 3 provides the results of our method on different benchmark datasets considering different distortion types, and also it presents a detailed ablation study and comparison with existing NR-IQA methods. Section 4 concludes our study and provides some suggestions for future research on NR-IQA.

2. METHOD

Figure 1 depicts the workflow of the suggested method. The key steps of the proposed approach are:

- Simulating a PRI,
- extracting deep learning features through deep transfer learning,
- generating visual explanation image (VEI) using Grad-CAM,
- calculating the similarity between the VEI of the distorted input image and the VEI of PRI using a similarity measure like the peak signal-to-noise-ratio (PSNR),
- employing a nonlinear mapping function to transform quality scores into a final NR-IQA score, quantifying the overall quality of the distorted input image.

In the subsequent subsections, we explain the key steps of the proposed approach in detail.

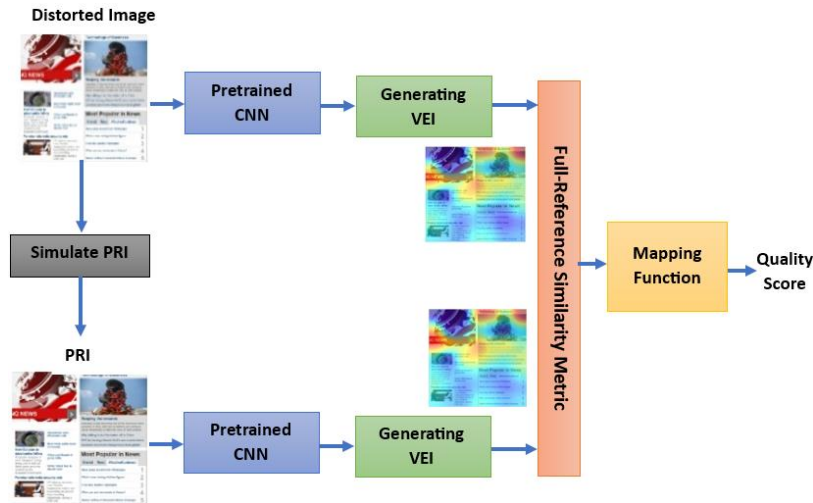


Figure 1. The workflow of the proposed NR-IQA approach

2.1. Step 1: simulating PRI

Min *et al.* [6], employed the same degradation process that generates the distorted image to further degrade it is used to produce a PRI image. In the case of varying distortion types, a degradation model for each type is defined. In this study, Gaussian blur (GB), white noise (WN) distortion, JPEG and JPEG2000 compression, are considered. We briefly introduce each degradation model:

- Gaussian blur degradation model [21]: following the fact that excessive blur introduces spurious structures that can be used as a reference image for the blurred image. To determine the PRI, a 3×3 averaging filter is used. Given a “Y” distorted image, the PRI is determined as:

$$PRI = f * Y = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} * Y \tag{1}$$

where "f" is the blurring filter, we configured it as an averaging filter, and * is the convolution operator.

- White noise degradation model: in this case, PRI is formed by introducing Gaussian noise to the distorted image Y [6]:

$$PRI = Y + N(0, \sigma_w) \tag{2}$$

where white noise $N(0, \sigma_w)$ with a mean of zero and a σ_w standard deviation.

- JPEG and JPEG2000 compression [22]: here, PRI is obtained by compressing the distorted image. Given the distorted image Y, PRI is produced as:

$$PRI = JPEG(Y, QT) \tag{3}$$

In this expression, JPEG refers to the JPEG encoder, while QT refers to the quantization table used which is fixed and represents very low compression quality.

2.2. Step 2: extracting feature maps utilizing pretrained CNN models

Transfer learning has demonstrated its ability to achieve cutting-edge performance across a range of applications, particularly when confronted with limited datasets [23]. The employment of pre-trained CNNs for feature extraction allows leveraging the powerful representation capabilities of a pre-trained CNN without needing to train an entire new network from scratch. Within the scope of this study, we will employ transfer learning, leveraging both pre-trained and fine-tuned networks. Specifically, we employ a set of robust pre-trained CNNs (pretrained on ImageNet dataset), such as DarkNet19, InceptionResNetV2 and VGG16 [24] to extract feature maps from the input distorted image and the corresponding PRI. Each feature map captures different levels of abstraction, from basic edges and textures in the earlier layers to complex object parts and higher-level concepts in the deeper layers. Table 1 presents a brief description of the pretrained CNN models used in this study.

Table 1. Description of the pretrained CNN models

Pretrained model	Input image size	Depth	No. of parameters (million)
Darknet19	256×256	19	20.8
InceptionResNetV2	299×299	164	55.9
VGG16	224×224	16	138

The pretrained CNN model used in this study can be downloaded from <https://www.mathworks.com/>. It should be noted that NR-IQA algorithms that rely on transfer learning have often incorporated data augmentation to enhance the diversity of training datasets, resulting in improved performance [25]. In this paper, multiple image transformations have been used, which include rotation, zoom, horizontal/vertical flips, and horizontal/vertical shifts [23], [26]. It is worth noting that the performance of the linear functions is improved by increasing the quantity of distorted input images used in this work by increasing and diversifying the size of the databases input to the pre-trained models. Although these models are trained on very large databases, they may be inconsistent with some tasks, so we increase the input data to improve their performance [27].

2.3. Step 3: generating VEI

A VEI is a heat map that identifies the important regions in the input image. It depicts the most important areas of the image that contribute to the decision to classify the model for a particular class. Various methods have been proposed in last years to explain CNNs, such as Grad-CAM, guided Backprop, class-activation mapping (CAM), saliency map, and occlusion map. In this paper, we employ the Grad-CAM generated from pre-trained CNNs to generate VEIs of the distorted input image and the PRI. Grad-CAM is a widely used visual explanation method and it provides accurate interpretation when compared with other methods [28].

Grad-CAM is a visualization technique used to obtain a class-discriminative localization map (i.e., VEI), denoted as $L_{\text{Grad-CAM}}^c \in \mathbb{R}^{u \times v}$ in generic CNN-based architectures. The Grad-CAM method has several steps starting with the computation of the gradient of the predicted class y^c with respect to feature maps A of a convolutional layer, represented as $\frac{\partial y^c}{\partial A_{ij}^k}$. To get weights, these gradients are globally average-pooled, denoted as α_k^c , which can be computed as [28]:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (4)$$

where the weight α_k^c represents a partial linearization of the deep network downstream from A , and these weights capture the importance or relevance of each feature map channel k for the target class c . It should be noted that y^c could be any differentiable activation.

The Grad-CAM heat-map is generated as a weighted combination of feature maps using the obtained weights. Specifically, it is computed as the non-linear rectified linear unit (ReLU) activation function applied to the sum of α_k^c multiplied by A^k across all channels, as:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}(\sum_k \alpha_k^c A^k) \quad (5)$$

The resulting Grad-CAM heat-map represents a visualization of the important regions in the feature maps for the target class. This heat-map is used as a VEI. Notably, the VEIs of the distorted image and its corresponding PRI are the same size.

2.4. Step 4: similarity calculation

In this paper, various similarity measures are assessed including, gradient magnitude similarity deviation (GMSD) [29], peak signal to noise ratio (PSNR) [30], NCC [30], feature similarity (FSIM) [31], normalised absolute error (NAE) [30], visual saliency-based index (VSI) [32] and structural similarity index (SSIM) [33]. The ablation study (section 3.5) demonstrated that the PSNR yields to the best NR-IQA results, and thus it is employed in the proposed method.

2.5. Step 5: mapping function for score alignment

Oszust [8], employ a five-parameter logistic function as a nonlinear mapping function to assign quality scores:

$$q = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(q - \beta_3))} \right) + \beta_4 q + \beta_5 \tag{6}$$

The parameters $\{\beta_j | j = 1, 2, \dots, 5\}$ are those five parameters derived through curve fitting. The original and mapped quality scores, denoted as \hat{q} and q , respectively.

2.6. Evaluation metrics

In this study, three performance metrics were used to evaluate the performance of NR-IQA methods: i) pearson linear correlation coefficient (PLCC), ii) Spearman's rank ordered correlation coefficient (SROCC) [34], and iii) root-mean-square error (RMSE) [6].

3. RESULTS AND DISCUSSION

3.1. Benchmark databases

Four standard IQA databases, including LIVE [35], CSIQ [36], SIQAD [37], and TID2013 [38], are used in this study to evaluate the performance of the proposed method. Moreover, four common distortion types: GB, WN, JPEG, and JP2K, are considered in our experiments. Detailed information about the databases is summarized in Table 2.

Table 2. Detailed information on the benchmark databases

Benchmark databases	No. of reference images	No. of GB distorted images	No. of WN distorted images	No. of JPEG distorted images	No. of JP2K distorted images	Score type
LIVE [35]	29	145	145	169	175	DMOS[0,100]
CSIQ [36]	30	150	150	150	150	DMOS[0,1]
SIQAD [37]	20	140	140	140	140	DMOS [0,100]
TID2013 [38]	25	125	125	125	125	MOS[0,9]

3.2. Implementation details and parameter setting

A few parameters are required to be set in this study. For the Gaussian blur, a PRI is obtained by employing a 3×3 average filter; for white noise, a blue PRI is generated by combining Gaussian noise with an average variance of 0 and 0.5 variance; and for JPEG and JP2K, the JPEG encoder in MATLAB is utilized. The "imwrite" function is utilized in particular. Setting the "Quality" parameter of "imwrite" to the absolute value 0 corresponds to the quantization table QT. It represents the most stringent compression that can be achieved by the encoder. Corner detection is accomplished using the MATLAB application of the minimum eigenvalue method [39]. The function in question is the "corner" function. In practice, there is no restriction on the utmost number of detected corners, as the value is set to an extremely high value. The value of 0:001 is specified for the "QualityLevel" parameter, which determines the minimum corner quality. In order to enhance the detection of corners for the purpose of describing structures, the quality parameter is initialized to an extremely low value. All experiments have been carried out on a computer with a 2.4 GHz Intel Core i5-2430M CPU and 8 GB RAM.

3.3. Analysing the performance of the proposed method

This section assesses the performance of the suggested technique taking into account four typical distortion kinds (GB, WN, JPEG, and JP2K) and four different benchmark IQA datasets: LIVE, SIQAD, CSIQ, and TID2013. The proposed method employs a pretrained InceptionResNetV2 model and PSNR as a similarity measure. As presented in Table 3, the proposed method demonstrates robust performance across diverse benchmark datasets and distortion types, as evidenced by PLCC and SROCC values. Our method

achieves a remarkable PLCC of 0.987 in the LIVE dataset for JPEG distortion and high SROCC values. A strong correlation observed in the LIVE dataset across all distortion types, emphasizing the method's effectiveness in capturing image quality variations.

The method exhibits reliable predictions in the CSIQ dataset, excelling notably in JPEG distortion. However, the TID2013 dataset reveals varied performance, with high correlation in WN and JPEG distortions but comparatively lower values in JP2K (PLCC of 0.878). Besides, the low RMSE values across datasets further underscore the accuracy of the proposed method's predictions. Despite minor fluctuations in PLCC, SROCC and RMSE values, the overall strong correlation and low error metrics reveals the method's effectiveness in providing reliable image quality assessments across diverse scenarios. Figure 2 presents examples of the input distorted images and the corresponding PRI with the GB, WN, JPEG and JP2K distortions considering different levels of distortions. As one can see, the VEIs of the distorted images and PRI are quite similar in the case of GB, JPEG and JP2K, and the same important regions are highlighted in the VEIs of the distorted images and PRI. In turn, some similarities appear in the VEIs of the distorted images and PRI, because of the WN in the Grad-CAM. Figure 3 shows different examples of quality scores for the proposed method. As one can see, less noisy images have higher predicted scores. The predicted scores provide good predictions for GB, WN, JPEG, and JP2K distortions.

Table 3. Assessing the performance of the proposed method in terms of the PLCC, SROCC and RMSE on three benchmark databases on four distortion types

Metric	Benchmark dataset	Distortion			
		GB	WN	JPEG	JP2K
PLCC	LIVE	0.972	0.891	0.987	0.939
	CSIQ	0.942	0.845	0.956	0.934
	TID2013	0.936	0.956	0.878	0.826
SROCC	LIVE	0.919	0.992	0.974	0.955
	CSIQ	0.942	0.954	0.963	0.943
	TID2013	0.960	0.939	0.974	0.948
RMSE	LIVE	8.062	6.849	6.037	7.946
	CSIQ	0.110	0.497	0.087	0.097
	TID2013	0.473	0.298	0.400	0.596

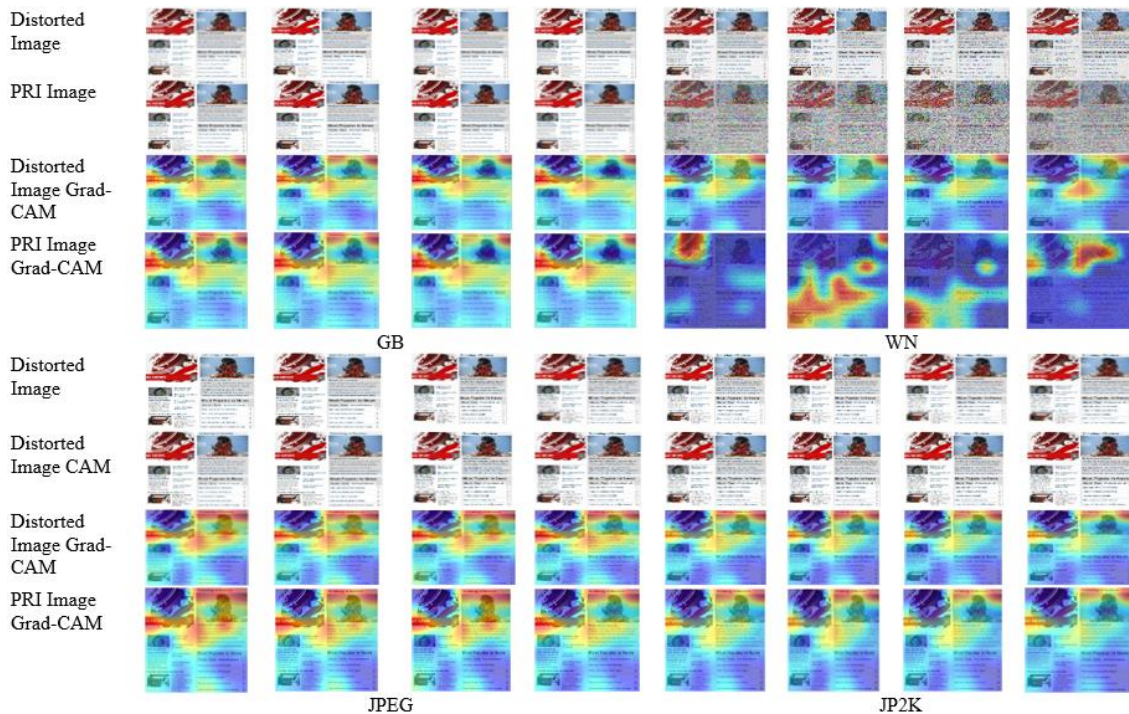


Figure 2. Examples of the input distorted images and the corresponding PRI with the GB,WN.JPEG and JP2K distortions considering different levels of distortions

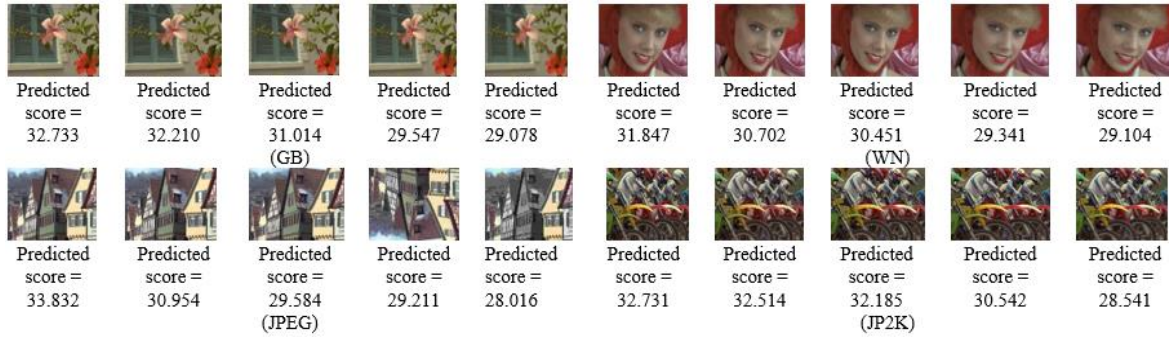


Figure 3. Examples of the quality scores of the proposed method with different distortion types

3.4. Ablation study

First, we study the performance of the proposed method with different deep pre-trained feature extractor. Table 4 demonstrates the efficacy of the proposed method with different deep pre-trained feature extractors across various benchmark datasets and distortion types. Noteworthy is the consistently superior performance of the proposed method utilizing InceptionResNetV2, evident in high PLCC values. For instance, in the LIVE dataset, the proposed method achieves a remarkable PLCC of 0.972 for JPEG distortion, surpassing both VGG16 (0.963) and DarkNet19 (0.969). Similarly, in the CSIQ dataset, the proposed method attains a PLCC of 0.956 for JPEG distortion, outperforming VGG16 (0.931) and DarkNet19 (0.928). These numerical examples underscore the method's robustness and highlight the impactful role of the InceptionResNetV2 feature extractor in consistently delivering accurate image quality assessments across diverse datasets and distortion types.

Table 4. Analyzing the performance of the proposed method with different deep pre-trained feature extractors

Benchmark database	Feature extractor	Distortion			
		GB	WN	JPEG	JP2K
LIVE	VGG16	0.963	0.988	0.978	0.976
	DarkNet19	0.969	0.980	0.983	0.971
	Proposed (InceptionResNetV2)	0.972	0.891	0.987	0.939
CSIQ	VGG16	0.931	0.750	0.887	0.846
	DarkNet19	0.928	0.821	0.939	0.908
	Proposed (InceptionResNetV2)	0.942	0.845	0.956	0.934
TID2013	VGG16	0.925	0.931	0.963	0.928
	DarkNet19	0.918	0.944	0.960	0.844
	Proposed (InceptionResNetV2)	0.936	0.956	0.878	0.826

Second, we study the impact of the selection of the similarity measure on the performance of the proposed method. Besides PSNR, we assess the impacts of other similarity measures: GMSD [29], PSNR [30], NCC [30], FSIM [31], NAE [30], VSI [32] and SSIM [33]. Table 5 tabulates the PLCC values of the proposed method with different similarity measures with four common distortion types, i.e., GB, WN, JPEG and JP2K distortion to guide the score alignment. The average of the previous distortion for each dataset is calculated. It is noted that the PSNR leads to the best results, and thus it is used in this proposed method.

Table 5. Performance of the proposed method with different similarity measures

Similarity Measure	PLCC			
	TID2013	CSIQ	LIVE	Average
PSNR [30]	0.8990	0.9193	0.9323	0.9168
GMSD [29]	0.7417	0.8378	0.7538	0.7911
NCC [30]	0.9190	0.8854	0.9425	0.9156
FSIM [31]	0.9755	0.8855	0.89355	0.9181
NAE [30]	0.8456	0.9054	0.9301	0.8937
VSI [32]	0.7654	0.7106	0.8454	0.7738
SSIM [33]	0.9478	0.8458	0.8275	0.8737

3.5. Comparing the proposed method with traditional and deep learning-based NR-IQA methods

Table 6 provides a comprehensive comparison between the proposed method with 8 traditional NR-IQA methods, namely IL-NIQE [40], QAC [41], NIQE [42], LPSI [43], DIIVINE [44], BLIINDS-II [45], BRISQUE [46], WaDIQaM-NR [11], as well as TSPR [22] deep learning-based NR-IQA method. In the LIVE dataset, the proposed method consistently outperforms existing methods such as QAC, IL-NIQE, NIQE, LPSI, DIIVINE, BLIINDS-II, BRISQUE, WaDIQaM-NR, and TSPR, achieving a remarkable PLCC of 0.972 for JPEG distortion. Similarly, in the CSIQ dataset, the proposed method excels against various methods, including IL-NIQE, NIQE, LPSI, DIIVINE, BLIINDS-II, BRISQUE, WaDIQaM-NR, and TSPR, with a PLCC of 0.942 for JPEG distortion. In the TID2013 dataset, the proposed method continues to demonstrate strong performance, surpassing traditional methods and revealing competitive results against deep learning-based methods. These results highlight the effectiveness of the proposed method in providing accurate and reliable no-reference image quality assessments across a diverse range of distortions and datasets, positioning it favorably among existing state-of-the-art methods.

Table 6. Comparison between the proposed method and traditional and deep learning-based NR-IQA methods

Benchmark database	Method	Distortion			
		GB	WN	JPEG	JP2K
LIVE	QAC [41]	0.911	0.928	0.944	0.866
	IL-NIQE [40]	0.933	0.987	0.959	0.905
	NIQE [42]	0.945	0.976	0.952	0.926
	LPSI [43]	0.915	0.965	0.975	0.936
	DIIVINE [44]	0.921	0.984	0.910	0.913
	BLIINDS-II [45]	0.938	0.980	0.968	0.935
	BRISQUE [46]	0.951	0.985	0.973	0.923
	WaDIQaM-NR [11]	0.961	0.989	0.953	0.962
	TSPR [22]	0.968	0.995	0.980	0.973
	Proposed	0.972	0.891	0.987	0.939
CSIQ	QAC [41]	0.857	0.878	0.938	0.895
	IL-NIQE [40]	0.894	0.864	0.955	0.926
	NIQE [42]	0.925	0.811	0.935	0.926
	LPSI [43]	0.930	0.687	0.969	0.918
	DIIVINE [44]	0.899	0.888	0.824	0.896
	BLIINDS-II [45]	0.893	0.774	0.938	0.915
	BRISQUE [46]	0.928	0.938	0.946	0.897
	WaDIQaM-NR [11]	0.966	0.918	0.943	0.870
	TSPR [22]	0.903	0.952	0.945	0.876
	Proposed	0.942	0.845	0.956	0.934
TID2013	QAC [41]	0.848	0.797	0.869	0.809
	IL-NIQE [40]	0.848	0.884	0.900	0.890
	NIQE [42]	0.819	0.827	0.893	0.907
	LPSI [43]	0.836	0.775	0.954	0.916
	DIIVINE [44]	0.848	0.860	0.664	0.901
	BLIINDS-II [45]	0.849	0.648	0.877	0.920
	BRISQUE [46]	0.848	0.851	0.900	0.918
	WaDIQaM-NR [11]	0.821	0.832	0.976	0.938
	TSPR [22]	0.902	0.938	0.949	0.964
	Proposed	0.936	0.956	0.878	0.826

3.6. Computational complexity of the proposed method

Table 7 gives information about the computational complexity of the suggested approach compared to the IQA models that are currently in use. The reported average running times (in seconds per image) are based on images randomly picked up from the TID2013 dataset on a computer with a 2.4 GHz Intel Core i5-2430M CPU and 8 GB RAM, with an image resolution of 512×384. The results indicate that the proposed method has a running time of 5.1 seconds per image. While not the fastest among the models listed, it achieves competitive results, especially when compared to other established IQA models. It should be noted that report here the run time of an unoptimized version of our code. Optimizing the code of the proposed method will significantly reduce the run time.

Table 7. Computational complexity of the proposed method

Method	Time (s)
NIQE [42]	3.3352
BRISQUE [46]	1.358
DIIVINE [44]	15.120
Proposed	5.123

3.7. Limitations of the proposed method

The current study, while showcasing the proposed method's high performance across various distortion types and datasets, acknowledges several limitations. Firstly, the study focuses on constructing a single restoration model for PRI by separately addressing the four individual distortion types. This approach may limit the model's ability to handle the complexities of mixed multi-distortions. Addressing this limitation could enhance the method's versatility in real-world scenarios where images often suffer from multiple distortions simultaneously. Additionally, the study points out that the quality of data generated by the PRI is closely tied to the type of distortion. This limitation raises questions about the generalizability of the proposed method across a broader range of distortion types. Future research could explore extending the method to handle a more diverse set of distortions and assess its performance in scenarios involving mixed distortions. Furthermore, the study notes that the reported run time of 5.1 seconds is based on an unoptimized version of the code. Acknowledging this, the authors suggest that optimization could significantly reduce the run time, enhancing the computational efficiency of the proposed method. The discussion emphasizes that the method, despite the current run time, strikes a balance between computational efficiency and precision, making it well-suited for applications prioritizing higher accuracy in image quality assessment over faster processing times.

4. CONCLUSION AND FUTURE WORK

This paper has presented a novel NR-IQA method based on deep visual explanations. Specifically, the main steps of our method are generating a PRI from the input degraded image, extracting deep learning features through deep transfer learning, generating VEI from the input degraded image and its PRI using Grad-CAM, and calculating the distance between the VEI of the distorted input image and the VEI of PRI.

The efficacy proposed method has been assessed on various benchmark datasets (LIVE, CSIQ, TID2013) and distortion types (White Noise, Gaussian blur, JPEG, and JPEG2000), finding that it achieves competitive performance. Different pre-trained CNN networks have been employed to extract VEIs like VGG16, InceptionResNetV2 and Darknet19, finding that InceptionResNetV2 feature extractor yielded the best results. The comparisons with several IQA models, including QAC, IL-NIQE, NIQE, LPSI, DIIVINE, BLIINDS-II, BRISQUE, WaDIQaM-NR, and TSPR, underscores the method's superiority.

One of the main implications of the proposed method is that it enhances interpretability of deep learning-based NR-IQA by creating heat maps that highlight significant areas contributing to the quality score utilizing Grad-CAM. Also, our method approach allows for a more informed model design and ensures a more effective and trustworthy NR-IQA solution, overcoming challenges of limited interpretability and assumptions prevalent in existing methodologies. Additionally, our method demonstrates broader applicability across various distortion types and datasets, providing a comprehensive study to validate its effectiveness. The proposed IQA method has practical applications in various fields such as: digital imaging and Photography, improving the quality control of medical imaging, and optimizing image compression algorithms to maintain high visual quality. Future research directions involve extending the proposed method to handle a broader range of distortion types and optimizing the computational efficiency of the proposed method. This expansion would contribute to a more comprehensive evaluation of the proposed method's applicability in real-world, dynamic settings. Recent advancements in deep learning have significantly improved image quality assessment methods. Our findings provide conclusive evidence that the proposed NR-IQA method, based on deep visual explanations, effectively enhances interpretability and accuracy across various distortion types and datasets, outperforming traditional models.

REFERENCES




- [1] M. Liu, J. Huang, D. Zeng, X. Ding, and J. Paisley, "A multiscale approach to deep blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 32, pp. 1656–1667, 2023, doi: 10.1109/TIP.2023.3245991.
- [2] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018, doi: 10.1007/s11760-017-1166-8.
- [3] W.-H. Kim, C. Hahm, A. Bajjal, N. Im, I. Cho, and J. Koo, "Pixel-by-pixel mean opinion score (pMOS) for no-reference image quality assessment," *arXiv preprint arXiv:2206.06541*, 2022, doi: 10.48550/arXiv.2206.06541.
- [4] S. Athar and Z. Wang, "Degraded reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 32, pp. 822–837, 2023, doi: 10.1109/TIP.2023.3234498.
- [5] C. Li, W. Yuan, A. C. Bovik, and X. Wu, "No-reference blur index using blur comparisons," *Electronics Letters*, vol. 47, no. 17, p. 962, 2011, doi: 10.1049/el.2011.0921.
- [6] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo-reference image," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2049–2062, Aug. 2018, doi: 10.1109/TMM.2017.2788206.
- [7] Y. Zhan and R. Zhang, "No-reference JPEG image quality assessment based on blockiness and luminance change," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 760–764, Jun. 2017, doi: 10.1109/LSP.2017.2688371.
- [8] M. Oszust, "No-reference image quality assessment with local gradient orientations," *Symmetry*, vol. 11, no. 1, p. 95, Jan. 2019, doi: 10.3390/sym11010095.

- [9] Z. Liang, W. Lu, Y. Zheng, W. He, and J. Yang, "The context effect for blind image quality assessment," *Neurocomputing*, vol. 521, pp. 172–180, Feb. 2023, doi: 10.1016/j.neucom.2022.11.026.
- [10] C. Fan, Y. Zhang, L. Feng, and Q. Jiang, "No reference image quality assessment based on multi-expert convolutional neural networks," *IEEE Access*, vol. 6, pp. 8934–8943, 2018, doi: 10.1109/ACCESS.2018.2802498.
- [11] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, Jan. 2018, doi: 10.1109/TIP.2017.2760518.
- [12] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2200–2211, May 2019, doi: 10.1109/TIP.2018.2883741.
- [13] R. S. Ravela, M. V. Shirvaikar, and C. Grecos, "No-reference image quality assessment based on deep convolutional neural networks," in *Real-Time Image Processing and Deep Learning 2019*, May 2019, p. 1099604, doi: 10.1117/12.2518438.
- [14] Z. Pan *et al.*, "DACNN: blind image quality assessment via a distortion-aware convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7518–7531, Nov. 2022, doi: 10.1109/TCSVT.2022.3188991.
- [15] M. Zhou *et al.*, "Attentional feature fusion for end-to-end blind image quality assessment," *IEEE Transactions on Broadcasting*, vol. 69, no. 1, pp. 144–152, Mar. 2023, doi: 10.1109/TBC.2022.3204235.
- [16] K. Y. Lin and G. Wang, "Hallucinated-IQA: no-reference image quality assessment via adversarial learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 732–741, doi: 10.1109/CVPR.2018.00083.
- [17] F. Gao, J. Yu, S. Zhu, Q. Huang, and Q. Tian, "Blind image quality prediction by exploiting multi-level deep representations," *Pattern Recognition*, vol. 81, pp. 432–442, Sep. 2018, doi: 10.1016/j.patcog.2018.04.016.
- [18] D. Varga, "Multi-pooled inception features for no-reference image quality assessment," *Applied Sciences*, vol. 10, no. 6, p. 2186, Mar. 2020, doi: 10.3390/app10062186.
- [19] S. Wang and Y. Zhang, "Grad-CAM: understanding AI models," *Computers, Materials & Continua*, vol. 76, no. 2, pp. 1321–1324, 2023, doi: 10.32604/cmc.2023.041419.
- [20] B. Ahmed, O. A. Omer, A. Rashed, D. Puig, and M. Abdel-Nasser, "Referenceless image quality assessment utilizing deep transfer-learned features," *Frontiers in Artificial Intelligence and Applications*, vol. 356, pp. 243–248, Oct. 2022, doi: 10.3233/FAIA220345.
- [21] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 508–517, Jun. 2018, doi: 10.1109/TBC.2018.2816783.
- [22] J. Hu, X. Wang, F. Shao, and Q. Jiang, "TSPR: deep network-based blind image quality assessment using two-side pseudo reference images," *Digital Signal Processing*, vol. 106, p. 102849, Nov. 2020, doi: 10.1016/j.dsp.2020.102849.
- [23] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, 2020, doi: 10.1007/s00034-019-01246-3.
- [24] F. Chen, H. Fu, H. Yu, and Y. Chu, "No-reference image quality assessment based on a multitask image restoration network," *Applied Sciences*, vol. 13, no. 11, p. 6802, Jun. 2023, doi: 10.3390/app13116802.
- [25] Y. Peskine, M.-C. Boucher, and F. Cheriet, "An interpretable data-driven score for the assessment of fundus images quality," in *Image Analysis and Recognition: 17th International Conference, ICIAR 2020, Póvoa de Varzim, Portugal*, 2020, pp. 325–331, doi: 10.1007/978-3-030-50516-5_28.
- [26] S. Sarkar *et al.*, "RL-CAM: visual explanations for convolutional networks using reinforcement learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2023, vol. 2023-June, pp. 3861–3869, doi: 10.1109/CVPRW59228.2023.00400.
- [27] J. Sheng, J. Fan, P. Ye, and J. Cao, "JNDMix: Jnd-based data augmentation for no-reference image quality assessment," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10096234.
- [28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
- [29] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, Feb. 2014, doi: 10.1109/TIP.2013.2293423.
- [30] W. A. Mustafa, H. Yazid, A. Salihah, A. Nasir, and M. View, "A review of image quality assessment (IQA): SNR, GCF, AD, NAE, PSNR, ME," *Journal of Advanced Research in Computing and Applications*, vol. 7, no. 1, pp. 1–7, 2017.
- [31] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011, doi: 10.1109/TIP.2011.2109730.
- [32] L. Zhang, Y. Shen, and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014, doi: 10.1109/TIP.2014.2346028.
- [33] J. Peng *et al.*, "Implementation of the structural SIMilarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection," *Medical Physics*, vol. 47, no. 4, pp. 1907–1919, 2020, doi: 10.1002/mp.14010.
- [34] D. Varga, "No-reference image quality assessment using the statistics of global and local image features," *Electronics*, vol. 12, no. 7, p. 1615, Mar. 2023, doi: 10.3390/electronics12071615.
- [35] H. R. Sheikh, "LIVE image quality assessment database release 2," 2005. [Online] <http://live.ece.utexas.edu/research/quality>.
- [36] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of electronic imaging*, vol. 19, no. 1, pp. 011006–011006, 2010.
- [37] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015, doi: 10.1109/TIP.2015.2465145.
- [38] N. Ponomarenko *et al.*, "Image database TID2013: peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, Jan. 2015, doi: 10.1016/j.image.2014.10.009.
- [39] Jianbo Shi and Tomasi, "Good features to track," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, 1994, pp. 593–600, doi: 10.1109/CVPR.1994.323794.
- [40] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015, doi: 10.1109/TIP.2015.2426416.
- [41] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013, pp. 995–1002, doi: 10.1109/CVPR.2013.133.
- [42] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2012, doi: 10.1109/LSP.2012.2227726.
- [43] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 339–343, doi: 10.1109/ICIP.2015.7350816.




- [44] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011, doi: 10.1109/TIP.2011.2147325.
- [45] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: a natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012, doi: 10.1109/TIP.2012.2191563.
- [46] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012, doi: 10.1109/TIP.2012.2214050.

BIOGRAPHIES OF AUTHORS






Basma Ahmed    received her M.Sc. degree in Computer Science from the Faculty of Science, South Valley University, in 2018. Currently, she is an Assistant Lecturer in the Department of Computer Science, Faculty of Computers and Information, South Valley University. Her research interests include image processing, information security, feature extraction, deep learning, and artificial intelligence. She can be contacted at email: basma_hafez@sci.svu.edu.eg.






Osama A. Omer    (Member, IEEE) received the B.Sc. and M.Sc. degrees from South Valley University, in 2000 and 2004, respectively, and the Ph.D. degree from the Tokyo University of Agriculture and Technology, in 2009. He has spent six months as a Postdoctoral Researcher with the Medical Engineering Department, Luebeck University, Germany. He has also spent three months as a Postdoctoral Researcher at Kyushu University, Japan. The last but not least, he has spent six months as a Research and Development Scientist Engineer at the NOKIA Research and Development Center, Tokyo/Japan, in 2008. He is currently a Full Professor at the Faculty of Engineering, Aswan University. His research interests include wireless communications, deep learning, and image/signal processing. He can be contacted at email: omer.osama@aswu.edu.eg.






Vivek Kumar Singh    received his Ph.D. degree in Computer Engineering in 2019 (CumLaude) from Universitat Rovira i Virgili, Spain. Currently, he is a Group Leader at Barts Cancer Institute (BCI) Queen Mary University of London, UK. He participated in many R&D projects funded by the European Commission and local funding agencies in USA, Spain, France and UK. He serves as a Guest Editor for many Special Issues in Indexed journals. He has published more than 40 scientific papers in international journals and conferences. His research interests include image processing, medical image analysis, machine learning, deep learning, and signal processing. He can be contacted at email: vivek.singh@qmul.ac.uk.



Amal Rashed    received her M.Sc. degree in Computer Science in 2008 and a PhD in 2016. Dr. Amal is an Assistant Professor in the Computer Science Department, Faculty of Computers and Information, South Valley University. Her research interests include image processing, machine learning, artificial intelligence, and information hiding. She can be contacted at email: amal.rashed2016@sci.svu.edu.eg.



Mohamed Abdel-Nasser    (Member, IEEE Senior Member) received his Ph.D. degree in Computer Engineering in July 2016 (CumLaude) from Universitat Rovira i Virgili, Spain. Currently, he is an Associate Professor of Electronics and Communications at Aswan University, Egypt. From 2020–2022, he was a Director of Research at the Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili. In 2022, he received the State Encouragement Award from the Egyptian Academy of Science and Technology for his excellent research track. He has received the 2017 Marc Esteva Vivanco prize for the best Ph.D. dissertation on A.I. He was the Principle Investigator of the Ensenyo project (AI-Powered online education platform funded by E.U.). He participated in many R&D projects funded by the European Commission and local funding agencies. He serves as a Guest Editor for many Special Issues in Indexed journals. He has published more than 100 scientific papers in International Journals and Conferences. His research is focused on image processing, medical image analysis, smart grid analysis using AI, and wireless communications. He can be contacted at email: mohamed.abdelnasser@aswu.edu.eg.