Article

A Visual Saliency-Based Neural Network Architecture for No-Reference Image Quality Assessment

Jihyoung Ryu [ID]

Electronics and Telecommunications Research Institute (ETRI), Gwangju 61012, Korea; jihyoung@etri.re.kr

Abstract: Deep learning has recently been used to study blind image quality assessment (BIQA) in great detail. Yet, the scarcity of high-quality algorithms prevents from developing them further and being used in a real-time scenario. Patch-based techniques have been used to forecast the quality of an image, but they typically award the picture quality score to an individual patch of the image. As a result, there would be a lot of misleading scores coming from patches. Some regions of the image are important and can contribute highly toward the right prediction of its quality. To prevent outlier regions, we suggest a technique with a visual saliency module which allows the only important region to bypass to the neural network and allows the network to only learn the important information required to predict the quality. The neural network architecture used in this study is Inception-ResNet-v2. We assess the proposed strategy using a benchmark database (KADID-10k) to show its efficacy. The outcome demonstrates better performance compared with certain popular no-reference IQA (NR-IQA) and full-reference IQA (FR-IQA) approaches. This technique is intended to be utilized to estimate the quality of an image being acquired in real time from drone imagery.

Keywords: image quality assessment (IQA); visual saliency; Inception-ResNet-v2; saliency map; no-reference IQA (NR-IQA)

1. Introduction

With the rapid rise of digital technology, more people utilize applications related to multimedia content, which includes images and video-related content [1]. Unlike in the past, now people have also included another item in their multimedia consumption, which includes 3D images and video content. People are inclined toward 3D content more as it focuses on providing a near-reality experience and perception. An end user demands multimedia content, which could be either images, videos or 3D content. However, the crucial part is that the transmitted content should be of a high quality [2]. The development continued, as users expect high-quality images. The process starts with image acquisition, where an image is simply captured and converted into a digital format using reliable cameras. Furthermore, preprocessing is performed on the images, and they are transmitted over the channel [3]. Finally, image restoration is performed, after which an end user can have an experience with the multimedia content [4]. The problem is that each of these subsystems from image acquisition to displaying the image can induce certain types of distortions in the image. Therefore, the end resultant image will be a distorted version. On the other hand, the end user requires higher-quality images and multimedia content. A lot of research has been conducted to develop solutions to assess the quality of images in a way that should be quite accurate, and this assessment needs to be automatic [5]. Nonetheless, image quality assessment (IQA) is essential to ensure that the image is not affected by any type of distortion before the image reaches the end user. IQA can play a vital role in improving the quality of an image and in image restoration as well.

Further down the line, IQA can be divided up into subjective quality scores and objective quality scores. In subjective image quality assessment (SIQA), the quality of an image is assessed from a human visual perspective. As the human eye can perceive and

assess the quality in the best way possible, this gives SIQA an edge to be more accurate and reliable. However, this process can be more time-exhaustive and more expensive, and this process is not automatic, so it cannot be applied in practical real-time situations as, in this digital age, everything is moving toward automation [1]. On the contrary, objective image quality assessment (OIQA) focuses on the development of mathematical models assessing the quality of images, where the image quality will be quite accurate and automatic, which will allow this to be suitable for real-time applications [6]. Due to this reason, more research has been carried out on objective IQA. Objective IQA is further split into three types. The first one is full-reference IQA (FR-IQA), in which a pristine image is available, termed as a reference image. This reference image is used in the image quality assessment process to determine the quality scores of distorted images [7]. The second type is reduced-reference IQA (RR-IQA), which contains only partial information about the pristine image (also termed a reference image) [8]. The last type is no-reference IQA (NR-IQA), in which there is no reference image available which is a pristine version of a distorted image. NR-IQA is more practical and applies to real-time situations as well [9].

As the NR-IQA approach is more adaptive and suitable for real-time applications, it is hence widely applicable in multimedia content applications. This reason makes NR-IQA a hot topic for research in the research community. Furthermore, NR-IQA is subdivided into two types, which include NR-IQA methods that are designed for a specific type of distortion (i.e., Gaussian blur). The other type includes NR-IQA methods which are designed for general purposes and do not focus on a specific type of distortion and noise in an image. Rather, they focus on image quality assessment on a general scale [6]. As was briefly described above that, FR-IQA and RR-IQA are only able to assess the image quality by using reference images. Therefore, both of these approaches depend on the reference images for image quality assessment. In NR-IQA, there is no dependence on the reference image in the process of assessing the image quality, and due to this, NR-IQA is also called blind image quality assessment (BIQA). A brief comparative review of the FR-IQA and NR-IQA methods from the literature shows that in FR-IQA, the conventional methods are the structural similarity index (SSIM) and feature similarity index (FSIM) [5]. However, when NR-IQA is considered, this quality assessment approach has a vital part which is termed feature extraction. Feature extraction (FE) in NR-IQA is split into a few types, which are handcrafted FE (HC-FE), using deep neural networks (DNNs) and using convolutional neural networks (CNNs). In the HC-FE feature extraction approach, the features are mainly extracted by using either of the two methods, which are the human visual system (HVS) or by utilizing natural scene statistics (NSS) [1,10,11]. The feature extraction part is considered a core of IQA, as it defines the robustness and accuracy of the image quality assessment method.

The human visual system (HVS) extracts the structural information from an image, while in the case of NSS, its basis is that the scenes have statistical characteristics, and these characteristics are affected by the type and level of distortion present in an image. The HVS and NSS are not robust enough to accurately predict the image quality of a distorted image [1]. In this situation, CNNs and DNNs are better options for IQA. CNNs and DNNs are widely used in various other fields [12–16], and similarly, they can be utilized in IQA to predict the quality of distorted images. Many CNN-based techniques are proposed in the literature [17,18]. A CNN is used for feature extraction, and further down the line, regression is used to build a model which can predict the image quality [1]. IQA algorithms that are based on CNNs are based on a special network structure (SNS). The depth of the networks determines how much detailed information can be captured, which is further utilized to predict the quality of distorted images [19]. As the depth of the network increases, more image information can be recorded to be used for quality assessment. The CNN-based IQA approach has higher accuracy for predicting the quality of distorted images. Deep neural networks for IQA are divided into two categories, which are the DNN feature extraction approach and the DNN end-to-end approach for IQA. The DNN feature extraction approach for IQA uses deep neural networks to extract features

from distorted images and then utilizes support vector regression to obtain quality scores for distorted images [5]. On the other hand, the DNN end-to-end IQA approach directly obtains and predicts the quality scores of distorted images just by using an input image whose quality is to be predicted or by using an input image patch [20,21]. Image quality assessment has a drawback, as IQA data are in a very limited amount, and this DNN end-to-end approach is a data-extensive method. This puts a limitation on this approach, but various solutions are provided in the literature for this situation. The biggest solution is to use the pretrained models and fine-tune them to achieve better performance.

In this work, we have proposed a framework with a visual saliency module that helps to extract only those regions in the image which are important to predicting the quality of the image. We have used the static fine-grained visual saliency technique as images have no motion involved, and therefore, we intend to use a technique that requires less computational resources. The proposed framework is designed to be used in real time to evaluate the images acquired from the drone while engaged in surveillance. However, it is important that the framework has to be fine-tuned on drone images once the dataset is prepared. The performance of the network is expected to be similar, as the images to be acquired from the drone will have similar image-acquiring sensors to that used to acquire images for the dataset used in this study.

The remainder of the manuscript is outlined as follows. Section 2 discusses the details regarding the KADID-10K dataset, Section 3 presents the proposed framework, the results and discussion on them are given in Section 4, and Section 5 has the conclusion.

2. KADID-10K Dataset

The KADID-10k dataset [22] was used to fine-tune the model and assess its capability. There are total of 10,100 distorted images in the KADID-10k dataset. These distorted images are generated from 81 pristine images. There are 25 different types of distortions introduced in these pristine images. Figure 1 shows the sample distorted images from the KADID-10k dataset. These distortion types contain different categories, such as blurs, compression, color-related and noise-related. Further details regarding the KADID-10k dataset can be found in [22]. These many categories of distortions make the KADID-10k dataset most suitable for training the model to be used in real-time scenarios.
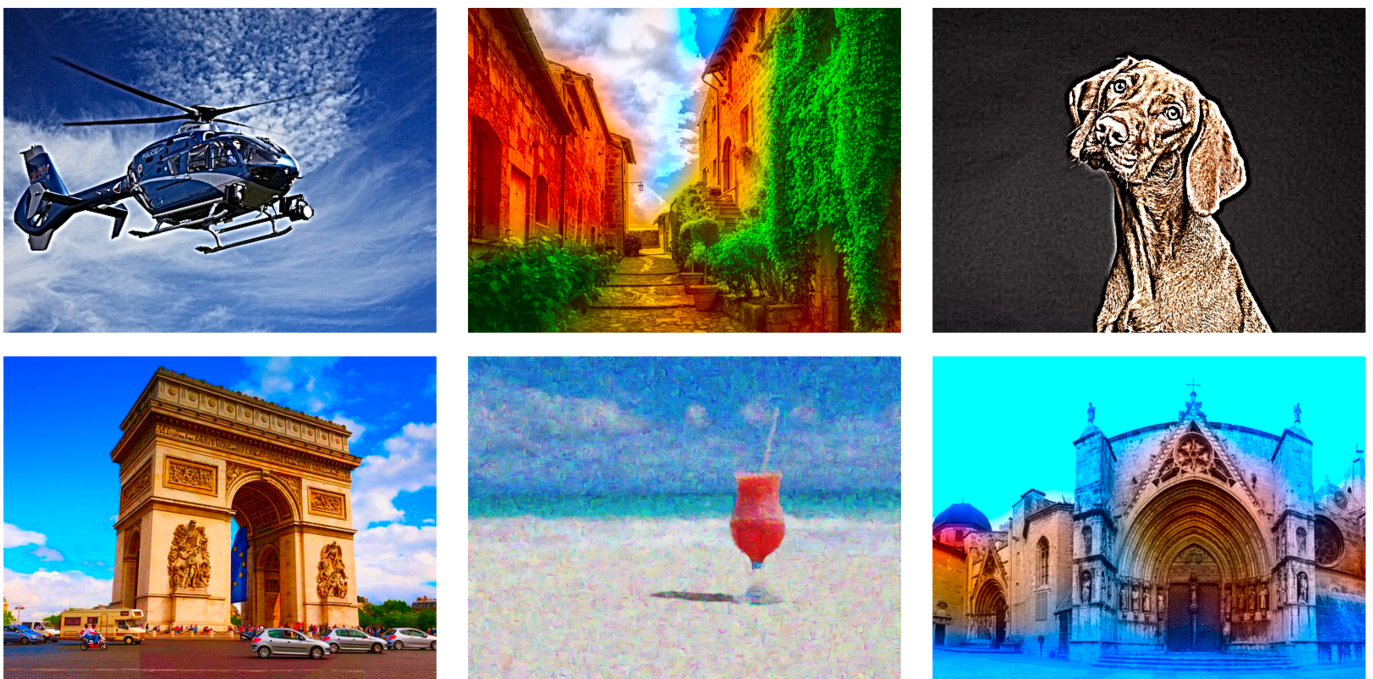

Figure 1. Sample distorted images from KADID-10K dataset.

3. Proposed Framework

Figure 2 shows the proposed framework for the prediction of the quality of the image. The proposed architecture includes a visual saliency module that extracts the important region of the image and allows the neural network model to pay attention to the important region only. This allows removing the maximum outliers from the learned feature space. The distorted images are affected by noise, and the noise mainly affects the high-frequency regions. The high-frequency regions are extracted well by the static visual saliency module. The visual saliency map is thresholded to find the binary mask, which is multiplied by the input distorted image in order to obtain the masked distorted image, as illustrated in Figure 2. The extracted masked and distorted image is passed by Inception-ResNet-V2 in order to predict the quality of the image.
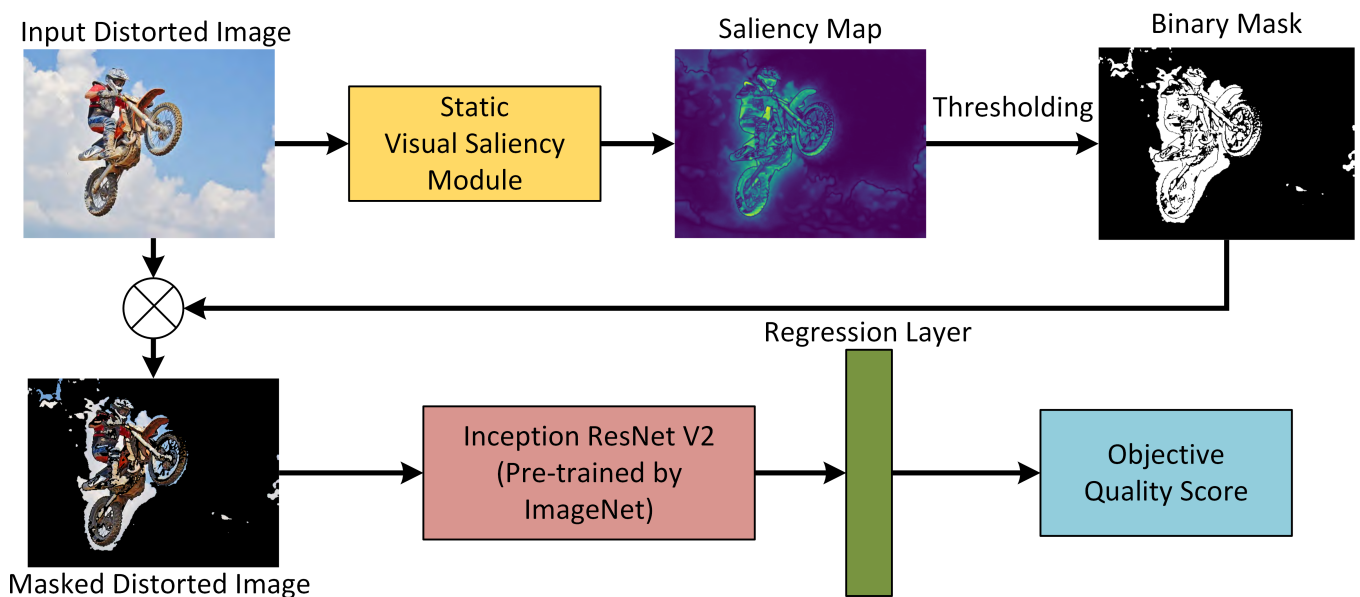

Figure 2. Visual saliency-based proposed framework for no-reference image quality assessment.

There are various existing models using saliency detection for IQA. MMMNet, SGDNet and VSI are a few among these. In order to increase the generalization of the IQA task through multi-task training, MMMNet adds an auxiliary saliency task [23]. SGDNet [24] is another architecture which uses a saliency prediction subnetwork for quality prediction of an image. The VSI model [25] has used multiple visual saliency techniques to compute the quality of an image. However, VSI does not explore the neural entwork features to perform a quality assessment. Regardless of these networks, a network is required which performs saliency mapping as an individual task regardless of the type of distortion. That is where the proposed network can be used to solve such an issue.

Below are the details regarding the static visual saliency module and Inception-ResNet-V2 used in the proposed framework.

3.1. Static Visual Saliency Module

Images and multimedia content have become a vital part of everyone's lives due to the widespread rise of digital technology. An image, just like a signal, contains useful information which is captured using a camera. Image processing utilizes various algorithms in order to extract information from an image, and computer vision is used to derive meaningful information from images. Images have multiple areas of interest which exist in them. Various image processing applications focus on extracting such regions of interest from images. Saliency can be simply defined as the special attention and focus of human vision on various meaningful regions in an image or a scene in real time. This can be

termed as an eye-catching region in an image or a scene, and there can be multiple regions of interest in an image [26].

We can also define it as visual attention and focus that humans give to some specific regions in an image or a scene. The same region is a high-frequency region that is mostly affected by noise and can be effectively used for quality assessment. These specific regions are mostly areas of the utmost interest in an image. Saliency is something that is embedded in the human biological system. In image processing, various applications involve the detection of saliency. Static saliency refers to an application which involves images only but not videos. In static saliency, algorithms are designed to capture and extract the region of specific interest from an image by utilizing various features of that particular image. Such features of an image can be the intensity level of an image, a particular pattern in an image or the contrast ratio of an image. Static saliency has remained part of various algorithms developed for image processing [27]. One of the widely used static saliency algorithm is the fine-grained algorithm that is part of our proposed framework.

The fine-grained approach was first proposed by Montabone et al. [28]. This approach utilizes the concept of the retina of the human eye. The retina of the eye has on-center and off-center ganglion cells. The on-center cells have a particular purpose; they are used in situations when there is a bright area that is surrounded by a darker region. The off-center cells are completely opposite to them, as they are utilized when there is a darker region that is surrounded by a bright region. Therefore, saliency in this algorithm, which finds out the region of interest in an image, is calculated by finding out the difference between the on-center and off-center regions [29]. In a fine-grained algorithm, the right view and left view of the image are projected, on which depth estimation is applied using stereo analysis. After this, the connected components are analyzed, and different blobs are generated. The filteration process is carried out in order to select the blobs using the size and shape features.

3.2. Inception-ResNet-V2

A convolutional neural architecture-based architecture, Inception-ResNet-v2, expands on the Inception network family of models while incorporating skip connections. ResNet and Inception, which provide outstanding results at a comparatively low operating complexity, have been instrumental in the biggest improvements in machine vision performance in the past few years. In Inception-ResNet-V2, the filter-expanding layer (1×1 convolution with no activation layer) is utilized to scale up the dimensions of the filters prior to addition to equal the depth of the input, which is placed after the inception block. Further batch normalization has only been applied on the standard layers in the situation of Inception ResNet and not applied on top of the summations.

In this work, pretrained Inception-ResNet-V2 is used, which is trained on the ImageNet dataset, having more than 1 million images. There are total of 164 layers in this network. Image quality assessment is a regression problem where the output layer of Inception-ResNet-V2 is replaced with a linear layer having one neuron.

4. Results and Discussion

4.1. Implementation Details

The database was randomly divided into training, testing, and validation datasets, and 60% of the data were used for training, while the remaining data were equally divided into testing and validation sets. Moreover, it was made sure that there would be no overlapping between the datasets. In total, 10 repetitions were carried out to train and evaluate the performance of the proposed framework. The pretrained Inception-ResNet-V2 was used, and each time, the model was fine-tuned using the training dataset.

4.2. Figure of Merits

To evaluate the performance of the proposed framework and carry out its comparison with the existing techniques, we used three metrics that were commonly used in the

literature. These metrics were the Spearman rank order correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC) and Kendall rank order correlation coefficients (KROCC). These metrics are mathematically expressed as follows:

$$SROCC = 1 - \frac{6 \sum_{i-1}^{N}(v_i - p_i)^2}{N(N^2 - 1)} \tag{1}$$

where N is the total number of images and v_i and p_i are the ranked actual and predicted quality scores, respectively:

$$PLCC = \frac{\sum_{i=1}^{N}(x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \hat{x})}\sqrt{\sum_{i=1}^{N}(y_i - \hat{y})}} \tag{2}$$

where N is the total number of samples (Images), x and y are sample points of the actual and predicted scores, respectively, and $\hat{x}$ and $\hat{y}$ are the means of the two different samples:

$$KROCC = \frac{P - Q}{\sqrt{(n_{[0]} - n_1)(n_0 - n_2)}} \tag{3}$$

where

$$n_0 = \frac{n(n - 1)}{2} \tag{4}$$

$$n_1 = \sum_j \frac{q_j(q_j - 1)}{2} \tag{5}$$

$$n_2 = \sum_k \frac{r_k(r_k - 1)}{2} \tag{6}$$

where P represents the number of concordant pairs, Q represents the number of discordant pairs, q_j shows the tied samples from the jth group, which ties with the first quantity, and r_k shows the tied samples from the kth group, which ties with the second quantity.
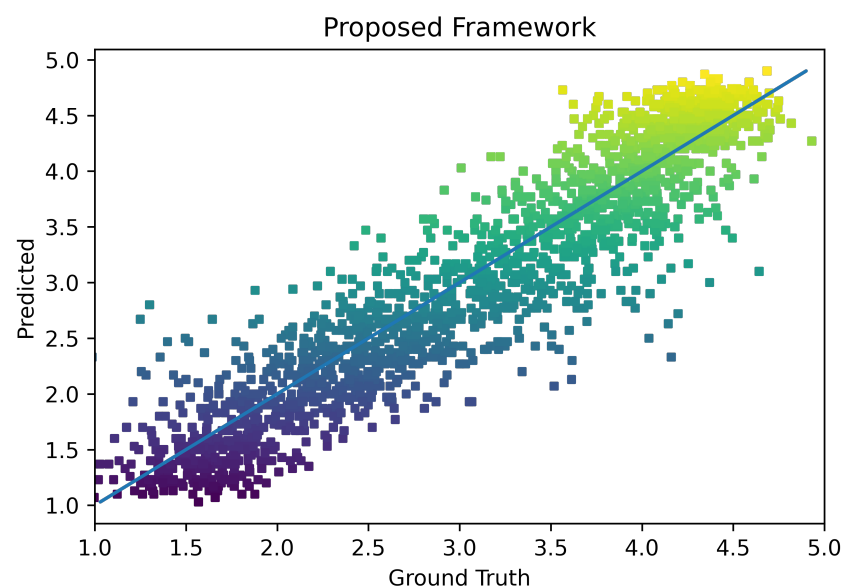
4.3. Performance Comparison

Table 1 shows the comparison of the proposed framework with the existing state-of-the-art techniques. As shown in the table, the proposed framework achieved the best performance. The proposed algorithm achieved 0.834 SROCC, 0.867 PLCC and 0.680 KROCC. The proposed framework showed an improvement of 0.103, 0.133 and 0.138 for the SROCC, PLCC and KROCC, respectively. The high performance of the framework depicts that using a visual saliency module for no-reference image quality assessment is beneficial. It helps the neural network model to learn the useful information required to predict the quality of the image. The InceptionResNetV2-based model exhibited good results when compared with other neural network models. Therefore, we utilized InceptionResNetV2 as our baseline model with the visual saliency module.

Figure 3 shows a scatter plot between the predicted quality score and the actual quality score. As can be seen, there are minimal outliers in the scatter plot. The scatter plot was very near to the ideal inclined line. This demonstrates that the visual saliency was able to find the important regions, and predicting based on those regions made the prediction more accurate, as unwanted regions leading to wrong predictions were ignored. It is important to highlight here the limitation of the proposed framework. The proposed framework intends to fail in the situation where there is no object (having no high-frequency component) in the image, as the visual saliency model will not extract the important region.

Table 1. Performance comparison of proposed framework with existing NR-IQA techniques.

Technique	SROCC	PLCC	KROCC	
BRISQUE [30]	0.519	0.554	0.368	
BIQI [31]	0.431	0.460	0.229	
CORNIA [32]	0.541	0.580	0.384	
SSEQ [33]	0.424	0.463	0.295	
DIVINE [34]	0.489	0.532	0.341	
BLINDS-II [35]	0.527	0.559	0.375	
HOSA [36]	0.609	0.653	0.438	
CNN [19]	0.603		0.619	-
BosICIP [37]	0.630		0.628	-
LPIPS [38]	0.721	0.713	-	
InceptionResNetV2 [22]	0.731	0.734	0.546	
Proposed Framework	**0.834**	**0.867**	**0.680**	

Figure 3. Scatter plot of the proposed framework.

5. Conclusions

Digital images have emerged as the primary means of disseminating information in the contemporary age of technology. Human social interactions provide a lot of picture data. The social lives of individuals are made much more convenient by clear and greater-quality photos. This is, however, constrained by the state of information technology today. The collection of picture information is directly impacted by the introduction of various disruptions during the operations of image synthesis, communication, preservation and other inter-processing. Studies in the area of objective IQA are a way to solve this issue, because it has practical implications for measuring this attenuation at the algorithm level. The least restrictive requirements were found in the NR-IQA method in particular. Deep neural networks' recent growth has substantially improved technology, making it easier to complete the NR-IQA assignment. In this paper, we proposed a framework based on deep neural networks and visual saliency. Visual saliency helps to extract high-frequency regions in the image which are the most affected regions due to the introduction of noise. The proposed framework was evaluated using the KADID-10K dataset. The achieved re-

sults of the model demonstrate the superiority of the proposed framework. Visual saliency plays a significant role in improving NR-IQA performance. In the future, we intend to use this framework to evaluate the image quality acquired and transferred from drones.

References

1. Shen, L.; Chen, X.; Pan, Z.; Fan, K.; Li, F.; Lei, J. No-reference stereoscopic image quality assessment based on global and local content characteristics. *Neurocomputing* **2021**, *424*, 132–142. [CrossRef]
2. Bovik, A.C. Automatic prediction of perceptual image and video quality. *Proc. IEEE* **2013**, *101*, 2008–2024.
3. Chandra, M.; Agarwal, D.; Bansal, A. Image transmission through wireless channel: A review. In Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 4–6 July 2016; pp. 1–4.
4. He, J.; Dong, C.; Qiao, Y. Modulating image restoration with continual levels via adaptive feature modification layers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11056–11064.
5. Lu, Y.; Li, W.; Ning, X.; Dong, X.; Zhang, L.; Sun, L.; Cheng, C. Blind image quality assessment based on the multiscale and dual-domains features fusion. In *Concurrency and Computation: Practice and Experience*; Wiley: Hoboken, NJ, USA, 2021; p. e6177.
6. Varga, D. No-Reference Image Quality Assessment with Convolutional Neural Networks and Decision Fusion. *Appl. Sci.* **2021**, *12*, 101. [CrossRef]
7. Md, S.K.; Appina, B.; Channappayya, S.S. Full-reference stereo image quality assessment using natural stereo scene statistics. *IEEE Signal Process. Lett.* **2015**, *22*, 1985–1989.
8. Ma, L.; Wang, X.; Liu, Q.; Ngan, K.N. Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment. *Neurocomputing* **2016**, *215*, 21–31. [CrossRef]
9. Liu, Y.; Tang, C.; Zheng, Z.; Lin, L. No-reference stereoscopic image quality evaluator with segmented monocular features and perceptual binocular features. *Neurocomputing* **2020**, *405*, 126–137. [CrossRef]
10. Nizami, I.F.; Majid, M.; Anwar, S.M. Natural scene statistics model independent no-reference image quality assessment using patch based discrete cosine transform. *Multimed. Tools Appl.* **2020**, *79*, 26285–26304. [CrossRef]
11. Nizami, I.F.; Majid, M.; Anwar, S.M.; Nasim, A.; Khurshid, K. No-reference image quality assessment using bag-of-features with feature selection. *Multimed. Tools Appl.* **2020**, *79*, 7811–7836. [CrossRef]
12. Rehman, M.U.; Tayara, H.; Zou, Q.; Chong, K.T. i6mA-Caps: A CapsuleNet-based framework for identifying DNA N6-methyladenine sites. *Bioinformatics* **2022**, *8*, 3885–3891. [CrossRef]
13. Rehman, M.U.; Akhtar, S.; Zakwan, M.; Mahmood, M.H. Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology images. *Biomed. Signal Process. Control* **2022**, *71*, 103212. [CrossRef]
14. Rehman, M.U.; Cho, S.; Kim, J.; Chong, K.T. Brainseg-net: Brain tumor mr image segmentation via enhanced encoder–decoder network. *Diagnostics* **2021**, *11*, 169. [CrossRef] [PubMed]
15. Rehman, M.U.; Tayara, H.; Chong, K.T. DCNN-4mC: Densely connected neural network based N4-methylcytosine site prediction in multiple species. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 6009–6019. [CrossRef] [PubMed]
16. Rehman, M.U.; Cho, S.; Kim, J.H.; Chong, K.T. Bu-net: Brain tumor segmentation using modified u-net architecture. *Electronics* **2020**, *9*, 2203. [CrossRef]
17. ur Rehman, M.; Nizami, I.F.; Majid, M. DeepRPN-BIQA: Deep architectures with region proposal network for natural-scene and screen-content blind image quality assessment. *Displays* **2022**, *71*, 102101. [CrossRef]
18. Nizami, I.F.; Waqar, A.; Majid, M. Impact of visual saliency on multi-distorted blind image quality assessment using deep neural architecture. *Multimed. Tools Appl.* **2022**, *81*, 25283–25300. [CrossRef]
19. 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.
20. Kim, J.; Lee, S. Fully deep blind image quality predictor. *IEEE J. Sel. Top. Signal Process.* **2016**, *11*, 206–220. [CrossRef]
21. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3667–3676.

22. 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; pp. 1–3.
23. Li, F.; Zhang, Y.; Cosman, P.C. MMMNet: An end-to-end multi-task deep convolution neural network with multi-scale and multi-hierarchy fusion for blind image quality assessment. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 4798–4811. [CrossRef]
24. Yang, S.; Jiang, Q.; Lin, W.; Wang, Y. SGDNet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1383–1391.
25. Zhang, L.; Shen, Y.; Li, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* **2014**, *23*, 4270–4281. [CrossRef]
26. Nguyen, T.V.; Xu, M.; Gao, G.; Kankanhalli, M.; Tian, Q.; Yan, S. Static saliency vs. dynamic saliency: A comparative study. In Proceedings of the 21st ACM International Conference on Multimedia, Barcelona, Spain, 21–25 October 2013; pp. 987–996.
27. Zhou, W.; Bai, R.; Wei, H. Saliency Detection with Features From Compressed HEVC. *IEEE Access* **2018**, *6*, 62528–62537. [CrossRef]
28. Montabone, S.; Soto, A. Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image Vis. Comput.* **2010**, *28*, 391–402.
29. Sun, X.; Yang, X.; Wang, S.; Liu, M. Content-aware rate control scheme for HEVC based on static and dynamic saliency detection. *Neurocomputing* **2020**, *411*, 393–405. [CrossRef]
30. 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. [CrossRef] [PubMed]
31. Moorthy, A.K.; Bovik, A.C. A two-step framework for constructing blind image quality indices. *IEEE Signal Process. Lett.* **2010**, *17*, 513–516. [CrossRef]
32. 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.
33. 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]
34. 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]
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. 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. [CrossRef]
37. Bosse, S.; Maniry, D.; Wiegand, T.; Samek, W. A deep neural network for image quality assessment. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3773–3777.
38. 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.