Abstract

In image quality enhancement processing, it is the most important to predict how humans perceive processed images since human observers are the ultimate receivers of the images. Thus, objective image quality assessment (IQA) methods based on human visual sensitivity from psychophysical experiments have been extensively studied. Thanks to the powerfulness of deep convolutional neural networks (CNN), many CNN based IQA models have been studied. However, previous CNN-based IQA models have not fully utilized the characteristics of human visual systems (HVS) for IQA problems by simply entrusting everything to CNN where the CNN-based models are often trained as a regressor to predict the scores of subjective quality assessment obtained from IQA datasets. In this paper, we propose a novel HVS-inspired deep IQA network, called Deep HVS-IQA Net, where the human psychophysical characteristics such as visual saliency and just noticeable difference (JND) are incorporated at the front-end of the Deep HVS-IQA Net. To our best knowledge, our work is the first HVS-inspired trainable IQA network that considers both the visual saliency and JND characteristics of HVS. Furthermore, we propose a rank loss to train our Deep HVS-IQA Net effectively so that perceptually important features can be extracted for image quality prediction. The rank loss can penalize the Deep HVS-IQA Net when the order of its predicted quality scores is different from that of the ground truth scores. We evaluate the proposed Deep HVS-IQA Net on large IQA datasets where it outperforms all the recent state-of-the-art IQA methods.

1. Introduction

Recently, perceptual image qualities in smartphones and TV displays have become a very important factor that determines the superiority among their competitive products. Therefore, many companies are studying various image enhancement methods to consolidate the competitiveness of their products. However, since human observers are the ultimate consumers of the images, the developed image quality enhancement methods often entail subjective quality verifications, which are cumbersome and time-consuming. Therefore, an objective image quality assessment (IQA) is indispensable.

The most commonly used metrics for measuring image quality include simple peak signal-to-noise ratio (PSNR) and mean square error (MSE). However, it is well known that PSNR and MSE are not highly correlated with the perceived quality of human visual systems (HVS). In order to design an accurate IQA model, it is essential to reflect human visual perception characteristics for image quality.

Based on these observations, many computational model-based IQA methods have been proposed by psychophysical experiments [15, 16, 17, 18, 19, 20]. However, such models tend to have high computational complexity and relatively low prediction accuracy for various distortion types.

Recently, deep convolutional neural networks (CNN) have shown overwhelming performance in most of the image classification and segmentation problems. Based on these successes, learning-based IQA models are intensively proposed [1, 2, 21, 22, 23, 24, 28]. In the beginning, such
learning-based IQA models directly tried to predict a quality score for a single input patch [28]. Recently, the CNN-based IQA models are trained to learn the visual sensitivity maps by weighted pooling for prediction of image quality scores [1, 2, 23]. However, such visual sensitivity maps are simply computed by using the data distributions of IQA datasets without taking into account the characteristics of HVS. Thus, when the HVS-based models (e.g., just noticeable difference (JND) [6, 8], saliency [7]) are incorporated as input to the CNN-based IQA models, it is expected that more precise prediction of image quality scores can be performed, rather than entrusting all works to CNN. Fig. 1 shows the quality measures for two distorted images. As shown in Fig. 1, the popular objective FR-IQA metrics such as PSNR and SSIM are not coincident with the MOS values obtained by subjective IQA. Even the latest CNN-based IQA method by Bosse et al. [2] also failed to predict correctly since it was trained only by using the data distributions of IQA datasets, not considering the visual quality perception characteristics of HVS.

Inspired by these observations, our proposed IQA network is designed by incorporating the HVS’s perception characteristics such as visual saliency and JND. Our contributions can be summarized as follows.

1. We present the first approach on applying HVS-based psychophysical models to the deep learning IQA problem, where visual saliency and JND properties are effectively incorporated as input of the sub-network according to their respective characteristics.

2. The visual saliency map is further used as a guideline for predicting the patch weight map in order to afford a stable training of end-to-end optimization for the proposed Deep HVS-IQA Net.

3. We newly propose a joint optimization framework that considers both prediction of the quality scores and their rank orders for the deep learning-based IQA.

4. We newly apply the channel attention mechanism to train a CNN-based IQA model, which helps to improve quality prediction performance.

2. Related Work

2.1. Human visual systems (HVS)

When HVS perceives a scene, objects or regions in the scene are not perceived with the same importance. The characteristics of HVS for image quality perception can be divided into four categories: contrast sensitivity function (CSF), luminance masking (LM), contrast masking (CM), and foveated masking (FM). The spatial CSF indicates that the sensitivity of HVS depends on spatial frequency values [4]. It shows that HVS operates like a band-pass filter by degrading the sensitivity to relatively low and high-frequency signals. The LM effect indicates that the sensitivity of HVS is influenced by background luminance. It has been proved by thorough experiments that the HVS is more sensitive at mid-luminance regions than relatively dark or bright regions [5]. The CM effect means that the sensitivity of HVS depends on the texture complexity of background. In the background regions with complicated texture, the HVS becomes more insensitive than in the homogeneous regions [5]. The FM effect implies that the sensitivity of HVS is affected by retinal eccentricity from the attention point of the eyes. The distortions that are far away from the focus regions are not easily noticeable [6].

A JND modeling, which refers to the minimum visibility threshold of HVS, takes into account these human visual characteristics effectively [6, 8]. Therefore, the JND model can be used as an important feature in the study of perceptual quality (e.g., perceptual video coding [8], image quality assessment [26]).

A saliency, which indicates the eye-focused region in an image, is also modeled by considering features of HVS [27]. It is a great help in finding the perceptually semantic objects in the entire image. Since the distortions of the semantic objects have greater impacts, the saliency map can be used for blur analysis [30] and image quality assessment [16].

2.2. Deep learning based IQA

After a great success of deep learning in many computer vision problems, there have been a few attempts to adopt deep learning for IQA problems. Kang et al. [28] first applied a CNN to the no-reference (NR) IQA without using any handcraft features. Due to the patch-based training schemes of CNN, there should be a ground truth score of image quality for each patch. However, most of the conventional IQA datasets contain only one visual quality score for each image, which represents the whole image quality. Due to this difficulty, Kang’s method considered all patches of an image to have the same quality scores as that of the whole image. In order to overcome this limitation, Kim et al. [23] proposed a full-reference (FR) IQA method in which the network predicts the sensitivity map for weighted pooling, by feeding the error map of an input image into the network. Bosse et al. [2] proposed both FR-IQA and NR-IQA networks that are trained to predict the patch-wise image quality scores and the patch-wise weights separately in end-to-end training. Then a weighted pooling is performed at the final stage of the network to predict a single quality score for the whole image input. Recently, Prashnani et al. [24] proposed a pairwise learning framework for IQA prediction, named PieAPP. They have shown that pairwise learning has a great impact on accurate image quality prediction and their model has achieved the highest prediction accuracy among the recent IQA methods.
3. Proposed Method

3.1. Architecture

Fig. 2 shows the architecture of our proposed Deep HVS-IQA Net. The Deep HVS-IQA Net takes four kinds of input: a distorted image, a reference image, its saliency map, and its JND probability map. Unlike the conventional methods [1, 2, 22, 23, 24, 28] that only use reference and distorted images as input, the Deep HVS-IQA Net utilizes the saliency maps and the JND probability maps as additional important inputs to extract HVS-based perceptual features so that it can more precisely predict the perceptual qualities of distorted images. The Deep HVS-IQA Net shown in Fig. 2 consists of a feature extractor, a patch weight predictor, a patch quality predictor and a weighted average-pooling module.

The feature extractor of the Deep HVS-IQA Net is comprised of a subnet for reference and distorted image inputs (ImgSubnet), a subnet for saliency map input (SalSubnet) and a subnet for JND probability map input (JndSubnet). The ImgSubnet is a Siamese network with 10 convolution layers and 5 channel attention blocks, which constitutes two identical networks sharing their weights in extracting the features from the reference and distorted image inputs. The JndSubnet has the same architecture as the ImgSubnet but its weights are not shared with those of ImgSubnet. The SalSubnet consists of 4 convolution layers where the output feature maps after its first and second max-pooling layers are fused via concatenation into the intermediate feature maps after the first and second max-pooling layers of the ImgSubnet and JndSubnet. From this, ImgSubnet and JndSubnet can be assisted by this saliency context to extract more effective features from visually focused information.

Saliency plays an important role within IQA, since distortions in image regions with larger visual attention tend to affect the visual quality of the entire image with more influence. As aforementioned, to successfully hand over the saliency information to the ImgSubnet and JndSubnet, the saliency context is fused into the intermediate features of the ImgSubnet and JndSubnet as shown in Fig. 2. After each feature concatenation, a $1 \times 1$ convolution is utilized in order to modulate the number of channels in the ImgSubnet and JndSubnet as it was before feature concatenation.

The feature extractor of the Deep HVS-IQA Net finally yields the feature vectors, $F_{Ref}$ and $F_{Dst}$, which are extracted using Siamese network, which is gray box in Fig. 2. The concatenated feature vector is input to the patch weight predictor and the patch quality predictor, and weighted averaging pooling is used to gather all the scores from the patches in one image.
perceptual image score [1], we construct a large feature vector by concatenating it with $F_{NIO}$, which is shown in the green box with operation $\text{Concat}(F_{Bd}, F_{Dn}, F_{BD})$ in Fig. 2. Then the concatenated vector is fed as an input to both patch weight predictor (PWP) and patch score predictor (PSP). The PWP and PSP have the same structure that consists of two fully connected layers of 512 and 1 outputs. The outputs of PWP and PSP can be interpreted as estimated patch weights ($w_i$) and patch quality scores ($v_i$), respectively. This is because the following weighted average-pooling module produces one final predicted quality score for an entire image according to the weighted average pooling with $w_i$ and $v_i$ as shown in Fig. 2.

3.2. Visual Saliency

For FR-IQA, the visual saliency map is computed for each reference image and is fed into the network as shown in Fig. 2. In addition, it is incorporated into a saliency loss term where the predicted patch weights are adjusted in accordance with the calculated visual saliency map. The motivation behind this is that the prediction of patch weights is guided according to the HVS characteristic of visual quality perception so that the patches in visually salient regions are treated more importantly than those of non-salient regions.

In order to calculate the visual saliency maps, we adopt the minimum barrier detection based approach [7]. Fig. 3-(e) shows a saliency map of the original image in Fig. 3-(a). As explained previously and shown in Fig. 2, each patch of an input saliency map is fed into the SalSubnet and the feature output after the first and second max-pooling layers are fused via concatenation with the feature outputs after the first and second max-pooling layers of the ImgSubnet and JndSubnet. By feeding this saliency context to the ImgSubnet and JndSubnet, the Deep HVS-IQA Net can focus on the feature regions with greater visual attention. It should be noted that, since the initial patch is of size $32 \times 32$ and the feature map size becomes $8 \times 8$ after the second max-pooling layer, the effect of incorporating saliency context is reduced for such a small $8 \times 8$-sized feature map. Therefore, the SalSubnet only incorporates the saliency context right after the first two max pooling layers.

Next, we explain how to incorporate the visual saliency maps into the saliency loss. In the Deep HVS-IQA Net, the visual saliency map can suggest a good direction to the PWP. A normalized $l$-th patch weight $\hat{w}_j$ is given by

$$\hat{w}_j = \frac{w_j}{\sum_{m \in \mathcal{I}} w_m} \quad (1)$$

where $w_i$ is the $i$-th estimated patch weight and $N_p$ is the total number of patches randomly extracted from a training image. Similarly, a visual saliency significance, $v_l$, for the $l$-th patch ($p_l$) in a visual saliency map is defined as

$$v_l = \sum_{(i,j) \in p_l} I(i,j) / \sum_{(i,j) \in \mathcal{I}} I(i,j) \quad (2)$$

where $I(i,j)$ is the visual saliency value at location $(i,j)$. $v_l$ in (2) is proportional to the total sum of the visual saliency values within its patch. Based on (1) and (2), the saliency loss term for the $k$-th distorted image is defined as

$$L_{sa}(I_k; \theta) = \frac{1}{N_p} \sum_{i \in \mathcal{I}} |\hat{w}_i - v_i| \quad (3)$$

3.3. Rank Loss

Now consider the situation when two distorted images are presented to human subjects in a subjective quality assessment. Sometimes it is hard for individuals to directly rate quality scores for the two images and the rated scores would even depend heavily on individuals. However, it is relatively easier to tell which image has better quality between the two. In addition, the consistency on the rank order of quality scores between predicted scores and ground truth scores is more important than simply regressing the score values in IQA problems. Fig. 4 illustrates an example of rank order significance for an IQA problem. In Fig. 4, Image A has a higher MOS value than Image B. The IQA...
methods 1 and 2 produce the same MAE value between ground truth MOS and predicted MOS (pMOS) values. Therefore, in a perspective of rank statistics, the IQA method 2 yielded pMOS values in the reverse order compared to the ground truth and the IQA method 1. Therefore, it is essential to take into account the rank order of pMOS values in IQA. Motivated from this, we add the rank loss for network training when the output score rank does not agree with the ground truth score rank.

Given a set of distorted images $I_n, n = 1, ..., N_B$ where $N_B$ is a batch size, as the network input, let $f(I_n; \theta)$ be a final predicted score (pMOS) of the Deep HVS-IQA Net where $f$ is the whole network function and $\theta$ is the network parameters. The ground truth score for a distorted image $I_n$ is denoted by $s_n$. Then the pairwise rank loss between two distorted images, $I_i$ and $I_j$, can be computed as

$$L(I_i, I_j; \theta) = \max(0, -\varepsilon + (s_i - s_j)\times(f(I_i; \theta) - f(I_j; \theta)))$$

where $\varepsilon$ is a small stability term. When the ranks of the predicted scores and ground truth scores agree, $L(I_i, I_j; \theta)$ is always 0, otherwise it is reduced to the absolute difference of their pMOS values as

$$L(I_i, I_j; \theta) = |f(I_i; \theta) - f(I_j; \theta)|$$

Since the rank loss in (4) can be calculated for each pair of distorted images, there are $N_B \times N_B$ combinations for a batch size $N_B$. For $N_B = 4$, we have the total rank loss as the sum of the six terms as

$$L_{\text{rank}} = L(I_1, I_2; \theta) + L(I_1, I_3; \theta) + L(I_1, I_4; \theta) + L(I_2, I_3; \theta) + L(I_2, I_4; \theta) + L(I_3, I_4; \theta)$$

Consequently, the total loss for training the Deep HVS-IQA Net is composed of three terms: mean absolute error (MAE) loss, the saliency loss in (3) and the rank loss in (6). The total loss can be evaluated as a weighted sum of the three loss terms as

$$L_{\text{tot}} = \alpha \times L_{\text{MAE}} + \beta \times L_{\text{Rank}} + \gamma \times L_{\text{Sal}}$$

where $\alpha = 1$, $\beta = 1$, $\gamma = 1$ is set as default in following experiments.

This enables the joint optimization of both prediction of the quality scores and their rank orders, which is differentiated from the previous works that considers the rank order of the distorted images [24, 31, 32].

### 3.4. JND probability

As another aspect of utilizing the HVS characteristics, it is worthwhile to incorporate the distortion sensitivity property of HVS in predicting the perceptual visual quality of distorted images. For this, the distortion detection probability map, called JND probability map, is used to extract effective features. We used the DCT-based JND model [8] to calculate the JND probability maps for distorted images and to feed them into the JndSubnet.

With the input JND probability map, the Deep HVS-IQA Net is trained to learn different HVS’s sensitivities for various distortion types and amounts. Fig. 3-(c) and -(d) show the squared image difference (SID) and the JND probability map between the original image in Fig. 3-(a) and the distorted image corrupted by the additive white Gaussian noise (AWGN) in Fig. 3-(b). Fig. 3-(g) and -(h) show the SID map and the JND probability map between the original image in Fig. 3-(a) and the distorted image corrupted by a Gaussian blur (GB) in Fig. 3-(f). As shown in Fig. 3-(b), AWGN distortions are easily perceived in the homogeneous regions due to the CM effects of HVS [5]. This can be confirmed in Fig. 3-(d) that the higher JND probability values are observed in the homogeneous regions (more bright regions) like the butterfly area. On the other hand, the GB distortion tends to be easily perceived in complex texture regions. As shown in Fig. 3-(h), higher distortion detection probability values are observed in the complex-textured regions such as the butterfly and the flowers (more bright regions). Thus, the JND probability maps can be effectively incorporated to reflect the HVS’s distortion perception characteristics well into the prediction of subjective IQA scores.

### 3.5. Channel Attention

Inspired by its superior performance in image captioning [9] and super-resolution [10], channel attention is adopted in the Deep HVS-IQA Net. Fig. 5 shows the architecture of a channel attention (CA) block. As depicted in Fig. 5, given the feature maps $P = [p_1, ..., p_c]$ as input to a CA block where $C$ is the number of channels of size $H \times W$, each feature map channel is pooled to its average value via global average pooling (GP). So, we have a pooled feature vector
patches extracted from an image depends on the image size. During validation, the entire input image is divided into 32×32-sized non-overlap patches. Therefore, the number of patches extracted from an image depends on the image size.

4. Experimental Results

4.1. Dataset

We assess the effectiveness of the proposed Deep HVS-IQA Net on four different image quality datasets: LIVE [3], CSIQ [11], TID2008 [12] and TID2013 [13]. The LIVE [3] image quality dataset contains 799 distorted images from 29 reference images. Five different types of distortions are included: JP2K compression (JP2K), JPEG compression (JPEG), additive white Gaussian noise (WN), Gaussian blur (BLUR) and Rayleigh fast fading channel distortion (FF). The CSIQ [11] image quality dataset consists of 866 distorted images from 30 reference images. The images are distorted by six different distortion types: JPEG, JP2K, WN, BLUR, additive pink Gaussian noise (PGN) and global contrast decrements (CTD). The TID2008 [12] image quality dataset contains 1,700 distorted images from 25 reference images. 17 different types of distortions are included. The TID2013 [13] image quality dataset is an extended version of the TID2008 dataset. It has the same set of reference images as TID2008 but more various distortions with 24 different distortion types are introduced.

All the image quality ratings for four different datasets were realigned and normalized to have the range [0, 9], where a higher value indicates perceptually better quality. Each dataset was randomly divided into three subsets by their reference images for training, validation, and test. For the LIVE dataset, 17 reference images are used for training, 6 images for validation and 6 images for test, out of 29 reference images in total. The CSIQ dataset is split into 20 training, 5 validation and 5 test images. For TID2008 and TID2013 datasets, 25 reference images are divided into 15 images for training, 5 images for validation and 5 images for test.

4.2. Performance comparison

The Deep HVS-IQA Net is trained for 1,000 epochs where an epoch describes the number of times the network has seen all the samples in the entire training dataset. The model with the lowest validation loss is chosen as the final model. The performances of the proposed method and various IQA algorithms are evaluated in terms of the correlation metrics such as Spearman rank order coefficient (SRCC) and Pearson linear correlation coefficient (PLCC). For both correlation metrics, the values closer to 1 indicate higher performances.

All the experiments were repeated 5 times and the results were averaged for a fair comparison. Each dataset was divided into training, validation and test image sets randomly every time when we newly start the model training. The Deep HVS-IQA Net is compared to total 15 IQA metrics. The ten model-based methods are included: Mean
Absolute Error (MAE), Root Mean Square Error (RMSE), SSIM [25], MS-SSIM [14], GMSD [15], VSI [16], PSNR-HMA [17], FSIMc [18], SFF [19] and SCQI [20]. The five learning-base methods are included: DOG-SSIMc [21], Lukin et al. [22], Kim et al. [23], Bosse et al. [2], and PieAPP [24].

Table 1 compares the SRCC and PLCC performances for the 15 IQA algorithms and our Deep HVS-IQA Net with four different datasets. It should be noted that the SRCC and PLCC values for the 15 IQA methods come from their original papers. The highest SRCC and PLCC values in are shown in boldface. As can be seen in Table 1, the Deep HVS-IQA Net outperforms all the state-of-the-art IQA methods for all IQA datasets consistently. Its superiority is even more emphasized in terms of SRCC by outperforming the other algorithms with a large margin. From this observation, we can infer that our proposed rank loss adequately contributes to more accurate pMOS prediction by considering the relation in quality score orders among the rated distorted images.

4.3. Ablation study

To examine the effectiveness of the conjunction of network components involved in the Deep HVS-IQA Net, we provide a detailed performance analysis on the networks with different combinations of the network components. Table 2 shows the SRCC performances for the different combinations of network components for the LIVE dataset.

Table 2. An ablation study for the combinations of different network components involved in the Deep HVS-IQA Net. Performances are evaluated on LIVE dataset.

Net | Baseline | Rank Loss | Saliency Map | JND Prob. Map | Channel Attention | SRCC
---|---------|-----------|--------------|----------------|--------------------|-----
A | ✓ | | | | | 0.9689
B | ✓ | ✓ | | | | 0.9768
C | ✓ | ✓ | ✓ | | | 0.9787
D | ✓ | ✓ | ✓ | ✓ | | 0.9821
E | ✓ | ✓ | ✓ | ✓ | ✓ | 0.9854

Fig. 6 shows the SRCC learning curves for five combinations of different network components over epochs.

Figure 6. Effects of network components in the Deep HVS-IQA Net. Performances are evaluated on LIVE dataset. The Net E (Deep HVS-IQA Net) achieves the best performance.
Note that the SRCC learning curves in Fig. 6 are only shown for the first 200 epochs for visualization convenience and most of the models converge after 150 epochs.

Note that in Table 2, the Net E with all given network components, corresponding to the Deep HVS-IQA Net, achieves the state-of-the-art performance. As can be seen in Table 2, the rank loss mostly contributes to the performance improvement over the baseline network (Net A), Bosse’s IQA model [2], which was not trained with the rank loss. Especially, when the JND probability map is added, the network quickly converged in only 10 epochs. Since the network did not need to analyze the HVS characteristics from scratch, it could converge early because such an HVS-based model was fed into the network as an input. In addition, performance improvement has been observed every time when each network component is combined. From this observation, we can infer that each of the network components in the Deep HVS-IQA Net effectively contributes to its performance improvement.

4.4. Comparison of patch quality maps and patch weight maps

In order to show the effectiveness of the Deep HVS-IQA Net, we compare the patch quality maps and the patch weight maps from the two distorted images by GB and AWGN between the Deep HVS-IQA Net and the baseline (Net A). Fig. 7 shows the patch quality maps and the patch weight maps for monarch image. The brighter regions indicate higher values in terms of quality and weight from Fig. 7-(c) to -(h), and from Fig. 7-(j) to -(m). For the GB-distorted image in Fig. 7-(b), the distortion in complex regions (butterfly) are mostly visible while the distortion in the smooth regions (background). Therefore, the complex regions should have lower values than the smooth regions in the patch quality maps. The patch quality map (Fig. 7-(d)) produced by the Deep HVS-IQA Net are well matched with what we expected than those (Fig. 7-(c)) by the baseline.

For the AWGN-distorted image in Fig. 7-(i), the distortion in the smooth area is much more noticeable than that of the complex area. Therefore, the smooth region should have lower values in the patch quality maps. Both baseline and Deep HVS-IQA Net have shown the results that agree with this HVS characteristic.

For patch weight maps, we have incorporated the saliency map into the saliency loss term as a guideline for patch weight maps. The patch weight maps (Fig. 7-(f), (m)) by the Deep HVS-IQA Net successfully follow the saliency map guideline (Fig. 7-(h)), which is well agreed with the HVS characteristic while the baseline network totally fails to do so (Fig. 7-(l)). From this, we can infer that the saliency loss term effectively acts as a successful guideline for learning the patch weight maps.

5. Conclusion

In this paper, our proposed Deep HVS-IQA Net successfully reflects the human visual sensitivity and psychophysical characteristics into the prediction of IQA problems by incorporating the JND probability maps and the saliency maps. Moreover, the rank loss in training has further improved the prediction performance by taking into account the rank order of distorted images. With channel attention, the Deep HVS-IQA Net could effectively figure out the features that are more informative by considering the interdependencies along channels. Through the extensive experiments, the Deep HVS-IQA Net has shown to outperform the most recent state-of-the-art IQA methods for various IQA datasets.
References

[1] S. Bosse, D. Maniry, T. Wiegand, and W. Samek. A deep neural network for image quality assessment. In IEEE International Conference on Image Processing (ICIP), pp. 3773-3777, 2016.

[2] S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek. Deep neural networks for no-reference and full-reference image quality assessment. IEEE Trans. Image Processing, vol. 27, no. 1, pp. 206-219, 2018.

[3] H. Sheikh, M. Sabir, and A. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Processing, vol. 15, no. 11, pp. 3440-3451, 2006.

[4] A. J. Ahumada, Jr., and H. A. Peterson. Luminance model based DCT quantization for color image compression. Proc. SPIE, vol. 1666, pp. 365-374, Aug. 1992.

[5] C.-H. Chou and C.-W. Chen. A perceptually optimized 3-D subband codec for video communication over wireless channels. IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 2, pp. 143-156, Apr. 1996.

[6] S.-H. Bae and M. Kim. A DCT-based Total JND Profile for Spatio-Temporal and Foveated Masking Effects. IEEE Trans. Circuits Syst. Video Technol., vol. PP, no. 99, pp. 1-1, 2016.

[7] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Méch. Minimum barrier salient object detection at 80 fps. Proc. IEEE International Conference on Computer Vision (ICCV), 2015.

[8] S. Bae, J. Kim, M. Kim. HEVC-based Perceptually Adaptive Video Coding using a DCT-based Local Distortion Detection Probability Model. IEEE Trans. Image Process., vol. 27, no. 7, pp. 3343-3357, 2016.

[9] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chu. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[10] Y. Zhang, K. Li, L. Wang, B. Zhong, Y. Fu. Image super-resolution using very deep residual channel attention networks. IEEE European Conference on Computer Vision (ECCV), 2018.

[11] E. C. Larson and D. M. Chandler. Most apparent distortion: Full-reference image quality assessment and the role of strategy. J. Electron. Imaging, vol. 19, no. 1, pp. 011006–011006–21, 2010.

[12] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008 - A database for evaluation of full-reference visual quality assessment metrics. Adv. Mod. Radioelectron., vol. 10, no. 4, pp. 30–45, 2009.

[13] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. Signal Processing: Image Communication, vol. 30, pp. 57–77, 2015.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In Proc. IEEE Asilomar Conf. Signals, Syst. Comput., pp. 1398–1402, 2003.

[15] W. Xue, L. Zhang, X. Mou, and A. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Transactions on Image Processing, vol. 23, no. 2, pp.684–695, 2014.

[16] 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, 2014.

[17] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli. Modified image visual quality metrics for contrast change and mean shift accounting. In Proceedings of the International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics, pp.305-311, 2011.

[18] 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, 2011.

[19] H. Chang, H. Yang, Y. Gan, and M. Wang. Sparse feature fidelity for perceptual image quality assessment. IEEE Transactions on Image Processing, vol. 22, no. 10, pp. 4007–4018, 2013.

[20] S. Bae and M. Kim. A novel image quality assessment with globally and locally consistent visual quality perception. IEEE Transactions on Image Processing, vol. 25, no. 5, pp. 2392–2406, 2016.

[21] S. Pei and Chen. Image quality assessment using human visual DOG model fused with random forest. IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3282–3292, 2015.

[22] V. Lukin, N. Ponomarenko, O. Ieremeiev, K. Egiazarian, and J. Astola. Combining full-reference image visual quality metrics by neural network. In SPIE/IS&T Electronic Imaging, 2015.

[23] Jongyoo Kim and Sanghoon Lee. Deep learning of human visual sensitivity in image quality assessment framework. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[24] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen. PieAPP: Perceptual Image-Error Assessment through Pairwise Preference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1808-1817, 2018.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, vol. 13, no. 4, pp.600–612, 2004.

[26] R. Ferzli and L. J. Karam, A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans. Image Process., vol. 18, no. 4, pp. 717–728, 2009.

[27] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 10, pp. 1915-1926, 2012.

[28] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolual neural networks for no-reference image quality assessment. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1733–1740, 2014.

[29] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In International Conference on Learning Representations (ICLR), 2015.
[30] S. Zhang, X. Shen, Z. Lin, R. Měch, J. P. Costeira, and J. M. Moura. Learning to understand image blur. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6586–6595, 2018.

[31] F. Gao, D. Tao, X. Gao, and X. Li. Learning to rank for blind image quality assessment. IEEE Trans. Neural Netw. Learn. Syst., vol. 26, no. 10, pp. 2275-2290, Oct. 2015.

[32] X. Liu, J. V. Weijer, and A. D. Bagdanov. RankIQA: Learning from rankings for no-reference image quality assessment. Proc. IEEE International Conference on Computer Vision (ICCV), 2017.
1. Performance Comparisons with Additional IQA methods and Datasets

In the main paper, we showed the MOS prediction performance of the Deep HVS-IQA Net on commonly used four IQA datasets which include LIVE [1], CSIQ [2], TID2008 [3], and TID2013 [4], and compared it with popular or state-of-the-art IQA methods [5-17]. In this section, we additionally compare the MOS prediction performance of the Deep HVS-IQA Net against 9 additional IQA methods [18-26]. Table 1 shows the performance comparison for total 25 IQA methods on four IQA datasets. As can be seen in Table 1, our Deep HVS-IQA Net outperforms or gives comparable performance to the existing state-of-the-art IQA methods. Additionally, the standard deviations of the SRCC, PLCC, and KRCC on the test sets of these five random splits are at most 0.004, 0.004, and 0.009 for all IQA datasets. It indicates that our random splits are representative of the entire IQA datasets and are not outliers.

| METHOD | LIVE [1] | CSIQ [2] | TID2008 [3] | TID2013 [4] |
|--------|----------|----------|-------------|-------------|
|        | SRCC (std) | PLCC (std) | KRCC (std) | SRCC (std) | PLCC (std) | KRCC (std) | SRCC (std) | PLCC (std) | KRCC (std) | SRCC (std) | PLCC (std) | KRCC (std) |
| NQM [18] | 0.909 | 0.714 | 0.624 | 0.624 | 0.614 | 0.564 | 0.643 | 0.690 | 0.474 |
| IFC [19] | 0.926 | 0.758 | 0.568 | 0.568 | 0.734 | 0.590 | 0.539 | 0.554 | 0.394 |
| VIF [20] | 0.964 | 0.828 | 0.749 | 0.749 | 0.808 | 0.754 | 0.677 | 0.772 | 0.515 |
| VSNR [21] | 0.927 | 0.762 | 0.705 | 0.705 | 0.682 | 0.625 | 0.681 | 0.740 | 0.508 |
| MAD [22] | 0.967 | 0.842 | 0.834 | 0.834 | 0.831 | 0.797 | 0.781 | 0.827 | 0.604 |
| RFSIM [23] | 0.940 | 0.782 | 0.868 | 0.868 | 0.864 | 0.764 | 0.774 | 0.833 | 0.595 |
| GSM [24] | 0.956 | 0.815 | 0.850 | 0.850 | 0.842 | 0.737 | 0.795 | 0.846 | 0.626 |
| SR-SIM [25] | 0.962 | 0.830 | 0.891 | 0.891 | 0.887 | 0.772 | 0.851 | 0.877 | 0.666 |
| MDSI [26] | 0.967 | 0.840 | 0.921 | 0.921 | 0.916 | 0.813 | 0.890 | 0.908 | 0.712 |
| MAE | 0.936 | 0.814 | 0.813 | 0.813 | 0.644 | 0.639 | 0.484 | 0.294 | 0.351 |
| RMSE | 0.931 | 0.812 | 0.783 | 0.783 | 0.752 | 0.617 | 0.453 | 0.358 | 0.327 |
| SSIM [5] | 0.948 | 0.796 | 0.876 | 0.876 | 0.861 | 0.691 | 0.637 | 0.691 | 0.464 |
| MS-SSIM [6] | 0.951 | 0.804 | 0.913 | 0.913 | 0.899 | 0.739 | 0.854 | 0.845 | 0.657 |
| GMSD [10] | 0.960 | 0.856 | 0.957 | 0.957 | 0.954 | 0.812 | 0.804 | 0.859 | 0.634 |
| VSI [11] | 0.952 | 0.806 | 0.942 | 0.942 | 0.928 | 0.786 | 0.898 | 0.876 | 0.712 |
| PSNR-HMA [7] | 0.871 | 0.726 | 0.922 | 0.922 | 0.888 | 0.780 | 0.847 | 0.819 | 0.673 |
| FSIMc [8] | 0.964 | 0.836 | 0.961 | 0.961 | 0.919 | 0.769 | 0.851 | 0.877 | 0.667 |
| SFF [9] | 0.965 | 0.836 | 0.963 | 0.963 | 0.964 | 0.828 | 0.851 | 0.871 | 0.658 |
| SCQI [12] | 0.941 | 0.784 | 0.943 | 0.943 | 0.927 | 0.787 | 0.905 | 0.890 | 0.729 |
| DOG-SSIMc [13] | 0.963 | 0.844 | 0.954 | 0.954 | 0.943 | 0.813 | 0.926 | 0.934 | 0.768 |
| Lukin et al. [14] | - | - | - | - | - | - | 0.930 | - | 0.770 |
| Kim et al. [15] | 0.981 | 0.965 | - | - | - | - | 0.939 | 0.947 | - |
| Bosse et al. [16] | 0.970 | 0.980 | - | - | - | - | 0.940 | 0.946 | 0.780 |
| PreAPP [17] | 0.977 | 0.894 | 0.973 | 0.973 | 0.975 | 0.881 | 0.945 | 0.946 | 0.804 |
| Deep HVS-IQA Net | **0.984** | **0.987** | **0.899** | **0.976** | **0.977** | **0.868** | **0.956** | **0.957** | **0.821** |

Table 1. Performance comparison on four different IQA datasets. For other methods except of ours, the numbers are collected directly from their original paper. The bolds indicate the model with the best performance.
2. Performance Comparisons of patch weight maps and patch quality maps

In the main paper, we presented four examples of patch weight maps and patch quality maps for one reference image from LIVE dataset (monarch). In this Supplemental, we provide a detailed rich performance analysis on patch weight maps and patch quality maps for various images from LIVE and TID2013 dataset.

2.1. Performance Comparisons on LIVE Dataset

We first provide a detailed analysis of the patch weight maps and the patch quality maps for four reference images selected from LIVE Dataset. Three types of distortions are considered for the four reference images, which include Gaussian blur (GB), JP2K compression (JP2K), additive white Gaussian noise (AWGN).

For GB and JP2K, the distortions in complex regions are mostly visible than the distortions in the smooth regions. Therefore, the complex regions should have lower values of patch quality scores than the smooth regions in the patch quality maps. The patch quality maps produced by the Deep HVS-IQA Net are better agreed with the HVS characteristics than those by the baseline. That is, the HVS characteristics for GB and JP2K distortions is more sensitive in complex (textured) regions than smooth (homogeneous) regions.

For AWGN, the distortion in the smooth area is much more noticeable than that of the complex area. Therefore, the smooth region should have lower values in the patch quality maps. The patch quality maps produced by the Deep HVS-IQA Net shows this characteristic which correlates well with the human perception.

For patch weight maps, we have incorporated the saliency map (SM) into the saliency loss term as a guideline. The patch weight maps by the Deep HVS-IQA Net successfully follow the SM-directed guideline, which is well agreed with the HVS characteristic. From this, we can infer that the saliency loss term effectively acts as a successful guideline for learning the patch weight maps.

Fig. 1 shows the patch quality maps and the patch weight maps for the distorted ‘parrots’ images with GB and AWGN, which were produced by the Deep HVS-IQA Net and the baseline. Fig. 1-(g) to -(j) show the patch weight maps for the Deep HVS-IQA Net and the baseline. As can be shown in Fig. 1-(h) and -(j), the patch weight maps estimated by the Deep HVS-IQA Net well represents the saliency points. Fig. 1-(k) to -(n) show the patch quality maps for the Deep HVS-IQA Net and the baseline. As aforementioned, the patch quality maps for JP2K distortion should have lower values on textured areas. Fig. 1-(l) clearly demonstrates this characteristic and shows a better result than the patch quality map evaluated by the baseline (Fig. 1-(k)). The patch quality maps for AWGN distortion should have higher values on complex regions. The patch quality map estimated by the Deep HVS-IQA (Fig. 1-(n) successfully follows this HVS characteristic while those by the baseline (Fig. 1-(m)) fails.

Fig. 2 shows the patch quality maps and the patch weight maps for the distorted ‘caps’ images with JP2K, GB, and AWGN, which were produced by the Deep HVS-IQA Net and the baseline. Fig. 2-(i) to -(k) show the patch weight maps for the baseline and Fig. 2-(l) to -(n) shows the patch weight maps for the Deep HVS-IQA Net. While the patch weight maps by the Deep HVS-IQA Net consistently follows the saliency map guideline (Fig. 2-(b)), patch weight maps by the baseline do not show any consistency among weight maps. For JP2K and GB, the patch quality maps should have higher scores on smooth regions (sky). This can be well perceived on the patch quality maps evaluated by the Deep HVS-IQA Net (Fig. 2-(r), (s)) while those by the baseline fails. For AWGN, the patch quality maps should have lower scores on smooth regions (sky) because AWGN distortion is more visible on the smooth regions. The patch quality map evaluated by the Deep HVS-IQA Net (Fig. 2-(f)) agrees with the expected result.

Fig. 3 shows the patch quality maps and the patch weight maps for the distorted ‘lighthouse’ images with GB and AWGN, which were produced by the Deep HVS-IQA Net and the baseline. Fig. 3-(g) to -(j) show the patch weight maps for the Deep HVS-IQA Net and the baseline. As can be shown in Fig. 3-(h) and -(j), the patch weight maps estimated by the Deep HVS-IQA Net well represents the saliency points where those by the baseline totally fails to find the saliency points (Fig 3-(g), -(i)). Fig. 3-(k) to -(n) show the patch quality maps for the Deep HVS-IQA Net and the baseline. As aforementioned, the patch quality maps for GB distortion should have lower values on textured areas. Fig. 3-(l) clearly demonstrates this characteristic and shows a better result than the patch quality map evaluated by the baseline (Fig. 3-(k)). The patch quality maps for AWGN distortion should have higher values on complex regions. The patch quality map estimated by the Deep HVS-IQA (Fig. 3-(n)) successfully follows this HVS characteristic.

Fig. 4 shows the patch quality maps and the patch weight maps for the distorted ‘lighthouse2’ images with JP2K and GB, which were produced by the Deep HVS-IQA Net and the baseline. Fig. 4-(g) to -(j) show the patch weight maps for the Deep HVS-IQA Net and the baseline. As can be shown in Fig. 4-(h) and -(j), the patch weight maps estimated by the Deep HVS-IQA Net well represents the saliency points (lighthouse and surroundings). Fig. 4-(k) to -(n) show the patch quality maps for the Deep HVS-IQA Net and the baseline. The patch quality maps for JP2K and GB distortion should have lower values on textured areas (rocks). Fig. 4-(l) and -(n) clearly demonstrates this characteristic and successfully follows the HVS characteristic.
Figure 1. Examples of the patch quality map and the patch weight map for images with different types of distortions for the same reference image ((a): ‘parrots’). For (c), the ground truth MOS value is 58.55, the pMOS values are 61.88 by baseline and 55.37 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 32.25, the pMOS values are 29.02 by baseline and 31.35 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 74, the pMOS values are 68.33 by baseline and 72.11 by the Deep HVS-IQA Net.
Figure 2. Examples of the patch quality map and the patch weight map for images with different types of distortions for the same reference image ((a): ‘caps’). For (c), the ground truth MOS value is 47.60, the pMOS values are 50.81 by baseline and 46.11 by the Deep HVS-IQA Net. For (d), the ground truth MOS value is 48.22, the pMOS values are 54.56 by baseline and 47.11 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 65.92, the pMOS values are 69.16 by baseline and 67.42 by the Deep HVS-IQA Net.
Figure 3. Examples of the patch quality map and the patch weight map for images with different types of distortions for the same reference image ((a): 'lighthouse'). For (c), the ground truth MOS value is 82.54, the pMOS values are 91.11 by baseline and 83.27 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 81.83, the pMOS values are 84.39 by baseline and 79.99 by the Deep HVS-IQA Net.
Figure 4. Examples of the patch quality map and the patch weight map for images with different types of distortions for the same reference image ((a): ‘lighthouse2’). For (c), the ground truth MOS value is 53.40, the pMOS values are 58.37 by baseline and 52.73 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 46.71, the pMOS values are 54.47 by baseline and 45.05 by the Deep HVS-IQA Net.
2.2. Performance Comparisons on TID2013 Dataset

We provide a detailed analysis of the patch weight maps and the patch quality maps from TID2013 Dataset. Five types of distortions are used for the analysis: Gaussian blur (GB), JPEG compression (JPEG), JP2K compression (JP2K), additive white Gaussian noise (AWGN), and multiplicative Gaussian noise (MGN).

For GB, JPEG, and JP2K, the distortion in complex regions are more easily perceived than those in the smooth regions. Therefore, the complex regions should have lower values than the smooth regions in the patch quality maps. The patch quality maps produced by the Deep HVS-IQA Net shows exactly what we expected than those by the baseline.

For AWGN and MGN, the distortion in the smooth area is much more noticeable than that of the complex area. Therefore, the smooth region should have lower values in the patch quality maps. The patch quality maps produced by the Deep HVS-IQA Net shows this characteristic which correlates well with the human perception.

For patch weight maps, the patch weights produced by the Deep HVS-IQA Net successfully follow the saliency map guideline, which is well agreed with the HVS characteristic.

Fig. 5 shows the patch quality maps and the patch weight maps for the distorted ‘sailing1’ images with GB, JP2K, AWGN, and MGN, which were produced by the Deep HVS-IQA Net and the baseline. Fig. 5-(k) to -(n) show the patch weight maps for the baseline and Fig. 5-(o) to -(r) shows the patch weight maps for the Deep HVS-IQA Net. While the patch weight maps by the Deep HVS-IQA Net consistently follows the saliency map guideline (Fig. 5-(b)), patch weight maps by the baseline do not show any consistency among weight maps. For JPEG and JP2K, the patch quality maps should have higher scores on smooth regions (sky). This can be well perceived on the patch quality maps evaluated by the Deep HVS-IQA Net (Fig. 5-(x), (w)). For AWGN and MGN, the patch quality maps should have lower scores on smooth regions (sky) because those distortions are more visible on the smooth regions. The patch quality map evaluated by the Deep HVS-IQA Net (Fig. 5-(y), (z)) agrees with the expected result. They have the lower scores on more distorted (more textured) regions.

Fig. 6 shows the patch quality maps and the patch weight maps for the distorted ‘ocean’ images with GB, JP2K, AWGN, and MGN, which were produced by the Deep HVS-IQA Net and the baseline. Fig. 6-(k) to -(r) show the patch weight maps for the Deep HVS-IQA Net and the baseline. As can be shown in Fig. 6-(o) to -(t), the patch weight maps estimated by the Deep HVS-IQA Net well represents the saliency points where those by the baseline do not show any consistency on their results (Fig 6-(k) to -(n)). Fig. 6-(s) to -(z) show the patch quality maps for the Deep HVS-IQA Net and the baseline. As aforementioned, the patch quality maps for JPEG and JP2K distortion should have lower values on textured areas (ocean). Fig. 6-(x) and -(w) clearly demonstrates this characteristic and shows a better result than the patch quality map evaluated by the baseline (Fig. 6-(s) and -(t)). The patch quality maps for AWGN and MGN distortion should have higher values on complex regions. The patch quality map estimated by the Deep HVS-IQA (Fig. 6-(y) and -(z)) successfully follows this HVS characteristic.
Figure 5. Examples of the patch quality map and the patch weight map for images with different types of distortion for a same reference image (i06). For (c), the ground truth MOS value is 19.75, the pMOS values are 20.13 by baseline and 19.91 by the Deep HVS-IQA Net. For (d), the ground truth MOS value is 34.23, the pMOS values are 36.46 by baseline and 33.84 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 55.86, the pMOS values are 53.44 by baseline and 54.18 by the Deep HVS-IQA Net. For (f), the ground truth MOS value is 54.96, the pMOS values are 51.24 by baseline and 53.52 by the Deep HVS-IQA Net.
Figure 6. Examples of the patch quality map and the patch weight map for images with different types of distortion for a same reference image (i16). For (c), the ground truth MOS value is 27.35, the pMOS values are 23.63 by baseline and 26.29 by the Deep HVS-IQA Net. For (d), the ground truth MOS value is 10.56, the pMOS values are 16.30 by baseline and 15.25 by the Deep HVS-IQA Net. For (e), the ground truth MOS value is 55.85, the pMOS values are 53.71 by baseline and 54.30 by the Deep HVS-IQA Net. For (f), the ground truth MOS value is 61.94, the pMOS values are 61.54 by baseline and 62.06 by the Deep HVS-IQA Net.
References

[1] H. Sheikh, M. Sabir, and A. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[2] E. C. Larson and D. M. Chandler. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging*, vol. 19, no. 1, pp. 011006–011006–21, 2010.

[3] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008 - A database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectron.,* vol. 10, no. 4, pp. 30–45, 2009.

[4] N. Ponomarenko, L. Jin, O. Jeremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chefdi, M. Carli, F. Battisti, and C. C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[6] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, pp. 1398–1402, 2003.

[7] N. Ponomarenko, O. Jeremeiev, V. Lukin, K. Egiazarian, and M. Carli. Modified image visual quality metrics for contrast change and mean shift accounting. In *Proceedings of the International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics*, pp. 305–311, 2011.

[8] 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, 2011.

[9] H. Chang, H. Yang, Y. Gan, and M. Wang. Sparse feature fidelity for perceptual image quality assessment. *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 4007–4018, 2013.

[10] W. Xue, L. Zhang, X. Mou, and A. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.

[11] 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. 9, pp. 4270–4281, 2014.

[12] S. Bae and M. Kim. A novel image quality assessment with globally and locally consistent visual quality perception. *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2392–2406, 2016.

[13] S. Pei and Chen. Image quality assessment using human visual DOG model fused with random forest. *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3282–3292, 2015.

[14] V. Lukin, N. Ponomarenko, O. Jeremeiev, K. Egiazarian, and J. Astola. Combining full-reference image visual quality metrics by neural network. In *SPIE/S&T Electronic Imaging*, 2015.

[15] Jongyoo Kim and Sanghoon Lee. Deep learning of human visual sensitivity in image quality assessment framework. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[16] S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[17] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen. PieAPP: Perceptual Image-Error Assessment through Pairwise Preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1808–1817, 2018.

[18] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.

[19] H. R. Sheikh, A. C. Bovik, and G. De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[20] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[21] D. M. Chandler and S. S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[22] 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.

[23] L. Zhang, L. Zhang, and X. Mou. FSIM: A feature based image quality assessment metric using Riesz transforms. In *Proceedings of the IEEE International Conference on Image Processing*, pp. 321–324, 2010.

[24] A. Liu, W. Lin, and M. Narwaria. Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500–1512, 2012.

[25] L. Zhang and H. Li. SR-SIM: A fast and high performance IQA index based on spectral residual. In *Proceedings of the IEEE International Conference on Image Processing*, pp. 1473–1476, 2012.

[26] H. Z. Nafchi, A. Shahkoarei, R. Hedjam, and M. Cheriet. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *IEEE Access*, vol. 4, pp. 5579–5590, 2016.