Hindawi
Advances in Multimedia
Volume 2022, Article ID 1686298, 12 pages
https://doi.org/10.1155/2022/1686298

*Research Article*

# Image Quality Predictor with Highly Efficient Fully Convolutional Neural Network

**Cao Yu-Dong, Liao Xin-Lin, and Liu Hai-Yan**

*School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China*

Correspondence should be addressed to Cao Yu-Dong; caoyd@lnut.edu.cn

Received 27 October 2021; Revised 10 January 2022; Accepted 11 January 2022; Published 23 February 2022

Academic Editor: Marco Roccetti

A highly efficient deep fully convolutional neural network (DFCN) for image quality assessment (IQA) is designed in this paper. The DFCN consists of two branches, one scoring local patches and the other estimating the weights of local patches to enhance quality prediction. Then, the DFCN outputs quality score of the whole image with aggregate weighted average pooling. There are no fully connected layers in the DFCN, resulting in far fewer parameters. In addition, the network model utilizes multiscale images as inputs to enrich the extracted distortion information. Furthermore, the parameters of the model are optimized in two steps to reduce the requirement for computing power and the risk of overfitting. The parameters of the shared layers and the quality module are optimized firstly, and then, the parameters of the weight module are optimized with the designed loss function. The extensive experimental results show that the proposed DFCN outperforms other competing IQA methods and has strong generalization ability.

## 1. Introduction

The quality of digital images is degraded by noise or other factors during acquisition, compression, storage, or transmission. Only after correctly evaluating the quality of image can the image postprocessing be executed effectively. In addition, the performance comparison on digital image processing algorithms requires IQA metrics.

The IQA methods are generally classified into two categories: subjective quality evaluation and objective quality evaluation [1]. Subjective evaluation is performed by the observer. The quality of subjective evaluation is usually tiered in five discrete levels: excellent, good, fair, poor, or bad. Objective evaluation is usually executed by a mathematical model that yields the numerical description of the image quality. The objective IQA methods are commonly divided into three types according to the availability of reference images: full-reference image quality assessment (FR-IQA), no-reference image quality assessment (NR-IQA), and reduced-reference image quality assessment (RR-IQA). FR-IQA evaluates the quality of the distorted image using a source reference image, whereas NR-IQA evaluates the quality of the distorted image without a source reference image (NR-IQA is also called blind image quality assessment, B-IQA). In contrast, RR-IQA evaluates the image quality using limited information from a reference image. IQA research started with FR-IQA, which is usually focused on evaluating the difference between the distorted image and the reference image based on the study of human vision. NR-IQA is not easily utilized due to the lack of the reference images. However, NR-IQA is of importance because it can be used in some real-time applications. NR-IQA has grown into a vigorous research topic in the past ten years.

In the past decade, deep learning technology has developed rapidly in the automation and artificial intelligence fields. Deep learning usually contains three different types of neural network structures: convolutional neural networks (ConvNets), recurrent neural networks (RNNs), and generative adversarial networks (GANs). Deep neural network models incorporate image feature extraction and classification or regression into the unified optimizing framework to implement a real end-to-end training process. In image recognition, deep learning has performed better than traditional algorithms. Specifically, ConvNet-based deep learning has been widely applied to IQA metrics. Deep learning-based IQA methods usually require strong computing power and a large amount of training data. However, IQA datasets are usually small. Some deep learning-based IQA methods enhanced data by segmenting images into patches, but the label noise was produced.

Motivated by Kang and Wang [2] and Ma et al. [3], we proposed a simple and highly efficient deep fully convolutional neural network model; one branch predicts primary scores of local patches, and the other branch enhances the predicted quality by estimating the weights for the local patches. Converting fully connected layers into convolutional layers at the end of the network enables our IQA method to generate conveniently quality scores and weights for the local patches with matrix format.

This study and its features are summarized as follows. First, multiscale images are used as inputs to acquire more detailed distorted information. Second, the quality of the entire image is predicted by weighted scoring local patches of the image. Third, the weights of local patches (derived from the learning stage) have adaptive characteristics. Fourth, substituting convolutional layers for fully connected layers in the network results in fewer parameters, which mitigates overfitting. Fifth, the parameter optimization is executed through two sequential stages using the designed different loss functions, which reduces the requirements of computing power and facilitates training.

Besides, we also included a review of IQA-related works before we proposed the network model in detail with experimental results presented. Our conclusions are presented at the end of this paper.

## 2. Related Works

The conventional IQA approaches try to design elaborate feature descriptors empirically which can efficiently depict the image degradation [4]. The structural similarity index (SSIM) [5] is such a classic FR-IQA algorithm that depends heavily on hand-crafted features. The SSIM imitates the human visual system's evaluation of image quality by calculating the similarity measure of two images in terms of luminance, structure, and contrast features. Since the origin of the SSIM, Lin Zhang et al. [6] have made subsequent improvements to its performance.

Early NR-IQA methods first extracted the statistical features of a distorted image and then predicted the quality score of the image using regression. Kumar and Singh Bawa [7] calculated the regional mutual information in the spatial domain based on the loss information of the distortion and then predicted the quality of the entire image. In addition, Oszust [8] converted the RGB image into the YCbCr color space, extracted the local features of key points, and then used the kernel-based support vector regression to output the image quality score. Bampis et al. [9] extracted detail loss measure features in the spatial domain and then used support vector regression to predict the quality of an image. Mittal et al. [10] presented the natural image quality evaluator (NIQE), which utilized the distance between two multivariate Gaussian models to quantify the score of a distorted image after extracting the quality-aware features. In 2012, Ye et al. [11] proposed COdebook Representation for No-reference Image Assessment (CORNIA), which is one of the first learning-based IQA methods. In CORNIA, spatially normalized patches are clustered using $k$-means based on which soft assignment encoding and maximal pooling are executed for codebook representation. CORNIA features have also been applied to dipIQ [12]. The deep artificial neural network has subsequently arisen in some IQA studies as the mainstream approach from conventional methods.

The deep neural network predicts image quality through end-to-end optimization. For example, Kang et al. [13] executed a deep neural network using one convolutional layer and two fully connected layers as an end-to-end version of CORNIA. Lin and Wang [14] designed a GAN structure that generates simulated reference images for IQA. However, the training process of the GAN model is very complicated and prone to failure. Bosse et al. [15] proposed a deep neural network-based IQA model (WaDIQaM) that split the distorted images into many true patches as inputs, which led to label noise. WaDIQaM regresses the quality scores of the patches using a group of fully connected layers, each of which was weighted by the other fully connected layers to predict the final quality of a distorted image.

In recent years, the transformer has been executed firstly in natural language processing, and it also attracts research interests in the computer vision field. You and Korhonen [16] applied the transformer to image quality assessment. The transformer encoder was used on the top of a feature map extracted by convolution neural networks. The transformer has achieved success in some computer vision tasks, but it has not exceeded convolutional neural networks for IQA.

We designed the DFCN-IQA model without fully connected layers, which used multiscale images as inputs. To the best of our knowledge, the deep fully convolutional neural network has not yet been applied to the IQA tasks.

## 3. DFCN for NR-IQA

The developed DFCN-IQA approach employs a deep fully convolutional neural network with multiscale images as inputs and the adaptive weights for enhancing quality prediction. Multiscale images have much more distorted feature information. An image pyramid is used to represent multiscale images.

*3.1. Image Pyramid for Multiscale Input.* Lindeberg [17] has theoretically proved that the Gaussian function is the only possible scale-space kernel and that the scale space generated by the Gaussian kernel is closely related to visual cognition. Accordingly, the scale space of an image is expressed as

$$L(x, y, \sigma) = \mathbf{I}(x, y) \otimes G(x, y, \sigma), \tag{1}$$

where $\mathbf{I}$ represents the initial input image, $x$ and $y$ are the coordinates of the pixels, and the symbol $\otimes$ represents the convolution operation. Gaussian kernel is defined as

$$G(x, y, \sigma) = Ae^{-\left((x^2+y^2)/2\sigma^2\right)}, \tag{2}$$

where $A$ is a constant coefficient and $\sigma$ is the standard deviation, which controls the shape of the function. In practice, Gaussian convolution is executed with a sliding window, the boundary of which is usually limited to $3\sigma$.

Our image pyramid is composed of four scale images; the first scale image is the original image. After the original image is convoluted with a Gaussian function, it is downsampled sequentially by a factor of two to produce three other coarser scale images.

As shown in Figure 1, both the normal clean image and the distorted noisy image are decomposed into 2 scale images. The $5 \times 5$ patches at the same relative coordinates in all four images are displayed exaggeratedly. There is a similarity between the corresponding patches across the scales in the "clean" image; however, there are significant differences between the corresponding patches across the scales in the "noisy" image. Thus, the distortion changes for different scales in the noisy image. Multiscale distorted images contain rich feature information. We use 4 scale images as inputs in the IQA model, which can reach a good performance in the experiments. It can be seen that there is little difference between the two far coarser scales from Figure 1.

*3.2. Network Architecture.* The structure of the designed IQA network model consists of the quality module g1, the weight module g2, and shared convolutional layers as shown in Figure 2. In the "conv-1" layer, the input original image is convolved with 32 filters, and each filter generates a feature map. Then, pooling is executed on each feature map to reduce the filter responses to a lower dimension. The second scale image is sent to the "conv-2" layer. The second feature map generated from the "conv-2" layer and the first feature map are sent to the "conv-3" layer together. Likewise, the third scale and fourth scale images are sent to the "conv-4" and "conv-6" layers, respectively. The "conv-7" layer produces the seventh feature map. Finally, the input image is mapped as 128-dimensional feature maps. The feature learning process in the shared layers is completed. The last aggregate average pooling operation yields the quality score of the overall image.

On top of the shared layers, the quality module $g_1$, which implements subtask I, generates the primary quality scores for the local patches. The normal region of interest (ROI) in an image can be detected by the visual salience method [19], but the saliency region and the distortion region may not be the same location. Thus, the weight module $g_2$, which implements subtask II, learns the weights for the local patches. Finally, the IQA model enhances the quality prediction through the weighted scoring method. Compared to a neural network containing fully connected layers with same depth, a fully convolutional neural network reduces the number of parameters because of the inductive bias of local connection and parameter sharing, which simplifies the computational complexity to a linear order.

The configuration of the proposed model is shown in Table 1. The stride of the convolution is fixed at 1 to extract the refined feature information of the distorted image. The size of the convolutional kernel, which determines the receptive field of the filter, is $3 \times 3$. Such a small convolutional kernel offers two main advantages. First, it can capture more refined changes in the image. Second, it can increase the depth of the network for the same receptive field. The spatial padding of the convolutional layer is structured such that the spatial resolution is preserved after the convolution operation (i.e., the padding is one pixel around the border for the $3 \times 3$ filter). The spatial pooling operation is performed after the convolutional layers have been processed. The spatial size was reduced by a factor of four via each maxpooling over a $2 \times 2$ window, with a stride of 2.

*3.3. Optimization of the Model.* More parameters and less training data can easily lead to overfitting. A distinct feature of the proposed network model is that there are no fully connected layers, which results in far fewer parameters; however, the number of parameters is still much larger than traditional IQA methods. Thus, optimization of the parameters is executed in two steps with the different designed loss functions. Such an optimization method can reduce the requirements for computing power and the risk of overfitting.

First, we optimize the shared layers and the quality module $g_1$. The primary quality scores matrix $\mathbf{y}$ can be defined as

$$\mathbf{y} = g_1\left(f\left(\mathbf{I}, \mathbf{w}_0\right), \mathbf{w}_1\right), \tag{3}$$

where the parameters of the shared layers are denoted by $\mathbf{w}_0$, the parameters of module $g_1$ are denoted by $\mathbf{w}_1$, $\mathbf{I}$ represents the input image (which is sent as input to the shared layers), and $f(\cdot,\cdot)$, which is the input to module $g_1$, represents the output of shared layers. Module $g_2$ outputs the weight matrix $\mathbf{a}$, which is defined as

$$\mathbf{a} = g_2\left(f\left(\mathbf{I}, \mathbf{w}_0\right), \mathbf{w}_2\right), \tag{4}$$

where the parameters of module $g_2$ are denoted by $w_2$ and $f$ $(\cdot,\cdot)$ is the input of module $g_2$.

The quality score of the overall image is generated through an aggregate weighted average pooling, and it is computed via

$$s = \sum_{i,j} s_{ij} = \lceil \mathbf{a} \times .\mathbf{y} \rceil = \sum_{i,j} a_{ij} \cdot y_{ij}, \tag{5}$$

where the symbol $\times.$ denotes Hadamard product between identically shaped primary quality score matrix and weight matrix and the symbol $\lceil \cdot \rceil$ means to find the sum of all matrix elements. $s_{ij}$ represents the ultimate quality score of a local patch, $a_{ij}$ is an element of $\mathbf{a}$, and $y_{ij}$ is an element of primary quality scores matrix $\mathbf{y}$. The maximal values of subscripts $i$ and $j$ are equal to the size of the final feature maps, which is the same as the size of the fourth scale image.

In the first step, the parameters of the shared layers and the quality module $g_1$ are optimized, and the loss function is designed as

$$\mathscr{L}_1 = \frac{1}{M} \sum_{m=1}^{M} \left\| g_1\left(f\left(\mathbf{I}_m, \mathbf{w}_0\right), \mathbf{w}_1\right) - \mathbf{l}_m \right\|_1, \tag{6}$$

where $\|\cdot\|_1$ represents the norm of matrix and $\mathbf{l}_m$ represents label scores of local patches in an image with a matrix format.

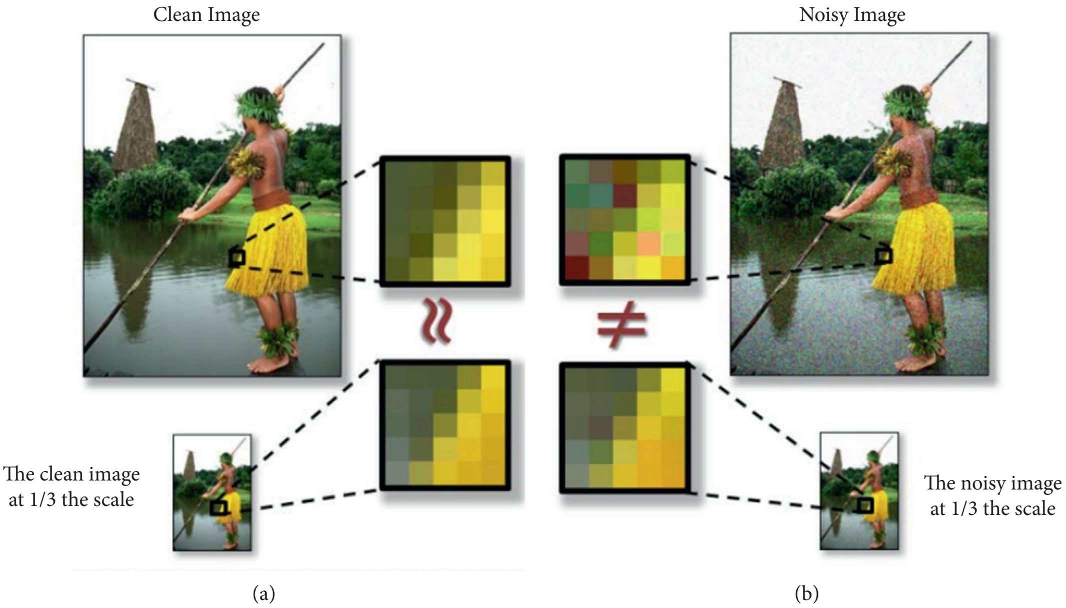
Clean Image
Noisy Image
≈
≠
The clean image at 1/3 the scale
The noisy image at 1/3 the scale
(a)
(b)

FIGURE 1: The 5 × 5 patches at the same relative coordinates in all four images (from [18]).

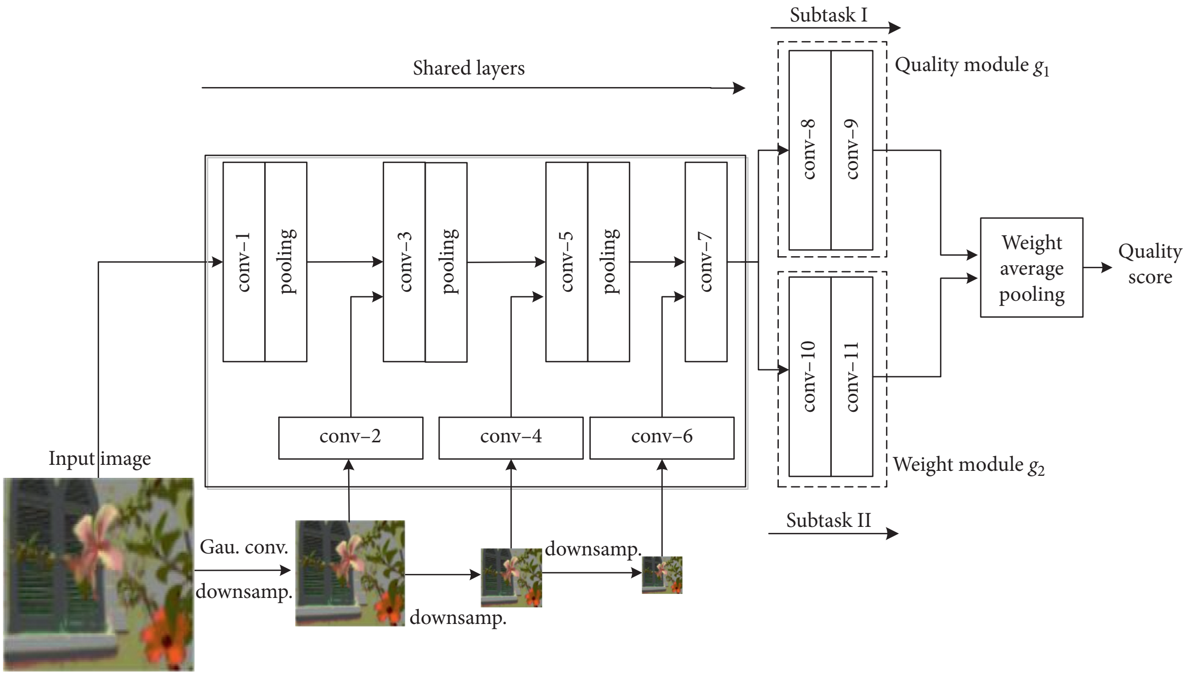

FIGURE 2: Network structure of the IQA model.

We supposed that the local patches are evenly assigned quality labels from the whole annotated images. $\mathbf{I}_m$ represents the $m$th image. Equation (6) describes the average error between the predicted quality score and ground truth score. To mitigate overfitting of the model, the regularization constraint item is introduced as

$$\mathscr{L}_1 = \frac{1}{M}\sum_{m=1}^{M}\left\|g_1\left(f\left(\mathbf{I}_m, \mathbf{w}_0\right), \mathbf{w}_1\right) - \mathbf{l}_m\right\|_1 + \lambda\left\|\mathbf{w}_1\right\|_p, \qquad (7)$$

where $\|\cdot\|_p$ represents $p$ norm. $p$ takes 0, 1, or 2. We set $p = 2$. $\lambda$ is a balance parameter between two items. The assumption that local quality is uniformly assigned over the distorted image causes a tremendous amount of label noise. So we design the other branch, weight module, which learns the weights for the local patches to enhance quality prediction.

In the second step, the parameters of module $g_2$ are optimized with a unified loss function after $w_0$ and $w_1$ are fixed, which is described as

Table 1: Configurations of the proposed deep ConvNet model.

| Module | Layer name | Receptive field size | Channel dim. | Stride | Pad |
|---|---|---|---|---|---|
| Shared layer | conv-1 | 3 × 3 | 32 | 1 | 1 |
| | pool | 2 × 2 | | 2 | 0 |
| | conv-2 | 3 × 3 | 32 | 1 | 1 |
| | conv-3 | 3 × 3 | 64 | 1 | 1 |
| | pool | 2 × 2 | | 2 | 0 |
| | conv-4 | 3 × 3 | 64 | 1 | 1 |
| | conv-5 | 3 × 3 | 128 | 1 | 1 |
| | pool | 2 × 2 | | 2 | 0 |
| | conv-6 | 3 × 3 | 64 | 1 | 1 |
| | conv-7 | 3 × 3 | 128 | 1 | 1 |
| $g_1$ | conv-8 | 3 × 3 | 128 | 1 | 1 |
| | conv-9 | 3 × 3 | 1 | 1 | 1 |
| $g_2$ | conv-10 | 3 × 3 | 128 | 1 | 1 |
| | conv-11 | 3 × 3 | 1 | 1 | 1 |

$$\mathcal{L}_2 = \frac{1}{M}\sum_{m=1}^{M}\left|\lceil g_2\left(f\left(\mathbf{I}_m,\mathbf{w}_0\right),\mathbf{w}_2\right)\times .g_1\left(f\left(\mathbf{I}_m,\mathbf{w}_0\right),\mathbf{w}_1\right)\rceil - l_m\right|, \qquad (8)$$

where $l_m$ represents the ground score of the distorted image (which satisfies $l_m = \lceil \mathbf{l}_m \rceil$). The symbols $\lceil \cdot \rceil$ and $\times .$ have the same meanings as in equation (5). The symbol $|\cdot|$ represents absolute value. The norm regularization term constraint is introduced to mitigate overfitting, and then the unified loss function is rewritten as

$$\mathcal{L}_2 = \frac{1}{M}\sum_{m=1}^{M}\left|\lceil g_2\left(\cdot,\mathbf{w}_2\right)\times .g_1\left(\cdot,\mathbf{w}_1\right)\rceil - y_m\right| + \lambda\left\|\mathbf{w}_2\right\|_p. \qquad (9)$$

Data enhancement, norm regularization, and dropout technology are employed during training to reduce the risk of overfitting. The training process is outlined as follows:

**Input:** distorted image **I**

**Output:** predicted quality score

Step 1. Given the original image, obtain four different scale images using an image pyramid method

Step 2. Input the four scale images into the network, and optimize the parameters of the shared layers and module $g_1$ using equation (7)

Step 3. Optimize the parameters of module $g_2$ according to equation (9)

Our network model can predict the quality score from coarse-to-fine grains with the dual-branch structure after training. Four scale images are generated with the image pyramid. The primary quality scores and the weights for the local patches are obtained through trained modules $g_1$ and $g_2$, respectively. Then, the more accurate quality scores of the local patches are computed via the Hadamard product. Finally, the quality score of the whole image is computed by summing all the elements of this matrix.

## 4. Experimental Results and Analysis

In this section, we first describe the IQA datasets and the evaluation metrics. We then conduct ablation experiments to identify the contributing factors in DFCN-IQA model. Finally, we compare the DFCN to classic and state-of-the-art IQA metrics. These results are computed through source code released by authors or come from existing papers.

*4.1. Datasets and Their Enhancement Processing.* LIVE [20] and TID2013 [21], adopted in our experiments, are popular IQA datasets. The KonIQ-10k dataset is not chosen because it is not used by most of the compared IQA algorithms. The main differences between the datasets are the numbers of reference images and distorted images, the types and levels of distortions, and the scoring standards. LIVE contains 29 source reference images, five distortion types, and 779 distorted images. The difference mean opinion score (DMOS) was used for subjective scoring in LIVE [22]. TID2013 dataset contains 25 source reference images, 24 distortions, and 3000 distorted images. The mean opinion score (MOS) was used for subjective scoring in TID2013. A reference image and its associated distorted images are shown in Figure 3.

It is challenging to train a deep learning model on LIVE and TID because the number of distorted images is scarce. To prevent overfitting, data enhancement is performed by rotating and mirroring distorted images. We then divide the training and test sets according to the reference images to ensure content independence regarding synthetic datasets LIVE and TID2013. We take 80% random images from each dataset to construct the training set. Then, we leave the remaining 20% for the testing. Different from the literature [15, 23], the training images were not normalized to reserve the distortion information such as original size, contrast, and luminance change.

*4.2. Evaluation Criteria.* The Spearman rank-order correlation coefficient (SROCC) and the Pearson linear correlation coefficient (PLCC) are standard evaluation criteria used by the Video Quality Experts Group (VQEG) [24]. The SROCC measures prediction monotonicity and the PLCC quantifies prediction accuracy. Both are correlation metric criteria, and their values close to 1 indicate good performance. The SROCC is defined as

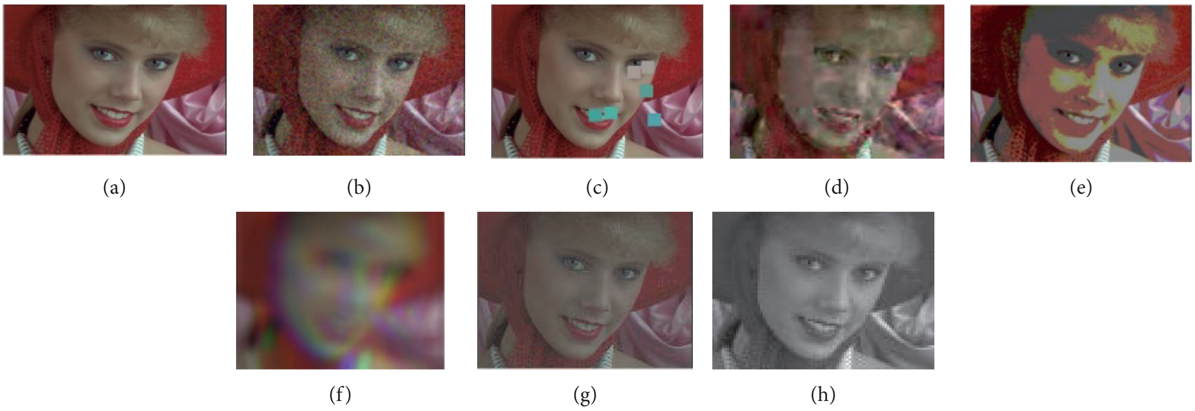

Figure 3: Examples of different types of distortions. (a) Reference image. (b) Spatially correlated noise. (c) Local block-wise distortions. (d) Comfort noise. (e) Quantization noise. (f) Chromatic aberrations. (g) Contrast change. (h) Change of color saturation.

$$P_{\mathrm{SROCC}} = 1 - \frac{6\sum_{k=1}^{N}\left(r_{x_k} - r_{y_k}\right)^2}{N\left(N^2 - 1\right)}, \tag{10}$$

where $N$ refers to the size of the sample data, $x_k$ represents the subjective evaluating score, $y_k$ represents the predicted scores of the model, and $r_{xk}$ and $r_{yk}$ describe the rankings of $x_k$ and $y_k$, respectively, in their own sample data. The PLCC is defined as

$$P_{\mathrm{PLCC}} = \frac{1}{N-1}\sum_{k=1}^{N}\left(\frac{x_k - \overline{x}}{\delta_x}\right)\left(\frac{y_k - \overline{y}}{\delta_y}\right), \tag{11}$$

where $\overline{x}$ and $\overline{y}$ are the average values of the two sets of data and $\delta_x$ and $\delta_y$ are their corresponding standard deviations.

*4.3. Ablation Study.* A series of ablation experiments were conducted to identify the influences of the core factors. The DFCN model enhances the quality prediction of the local patches by estimating the weights. To verify the effect of the weight module $g_2$, comparative experiments were conducted with multiple distortion types in TID2013: additive Gaussian noise (#1), spatially correlated noise (#3), quantization noise (#7), JPEG compression (#11), local block-wise distortions of different intensity (#15), change of color saturation (#18), and image color quantization with dither (#22). If the weight module $g_2$ does not work, it means that all the elements in matrix $\mathbf{a}$ are 1 in equation (5). It can be seen from Figure 4 that the weight module improves the performance of the model.

The performance of the SROCC was compared between a single input and multiscale inputs. As shown in Figure 5, the multiscale images as inputs enhance the quality prediction performance. Although the size of the distorted image is not fixed, we find that the four scale inputs can get excellent experimental performance.

Gaussian convolution is indispensable in generating the image pyramid. We investigated how different $\sigma$ values affect the quality prediction performance, as illustrated in Figure 6. The value of $\sigma$ was set to 1.65.

*4.4. Evaluation on Datasets.* Our approach outperforms state-of-the-art NR-IQA methods, even some FR-IQA methods, and competes with other deep learning-based IQA methods.

*4.4.1. Performance Comparison on LIVE.* We compare the proposed DFCN with 16 IQAs (8 traditional and 8 deep learning-based) on the LIVE dataset. Experimental results are displayed in Tables 2 and 3, respectively. The best performance for each column is in bold. The distortion types include JPEG2000 compression distortion (JP2K), JPEG compression distortion (JPEG), white noise distortion (WN), and Gaussian blur distortion (BLUR). “ALL4” in the last column is a comprehensive value for the four distortions to the left.

The 8 traditional competing methods are SSIM [5], CORNIA [11], BIQI (Blind Image Quality Index) [25], SOM (semantic obviousness metric) [26], DIIVINE (distortion identification-based image verity and integrity evaluation) [27], SNP-NIQE (Structure, Naturalness, and Perception quality-driven Natural Image Quality Evaluator) [28], BRISQUE (Blind/Referenceless Image Spatial QUality Evaluator) [29], and FSI (Free-energy principle and Sparse representation-based Index) [30].

The 8 state-of-art deep learning-based methods are CNN (convolutional neural network for image quality assessment) [13], WaDIQaM-NR [15], SGDNet (Saliency-Guided Deep neural network) [31], DLIQA (Deep Learning-based Image Quality Assessment) [32], BIECON (Blind Image Evaluator based on a CONvolutional neural Network) [33], MS-C (Multiple Scale Concat) [34], UNIQUE (Unified No-reference Image Quality and Uncertainty Evaluator based on a deep neural network) [35], and IQT (Image Quality assessment with Transformers) [36].

In the compared algorithms, CORNIA [11] is a classic learning-based method, and it was subsequently refined to SOM [26]. DLIQA [32] classified a distorted image into five levels and then estimated the quality score. DIIVINE [27] is an improved version of BIQI [25] with more advanced natural scene statistics. SNP-NIQE [28] was extended from NIQE, which extracted features of the structure, naturalness,

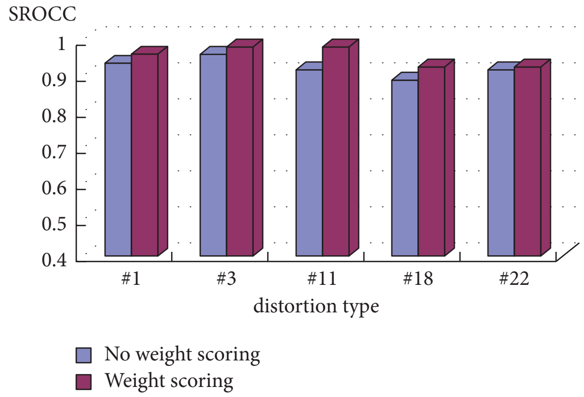

Figure 4: Influence of weight module on performance.

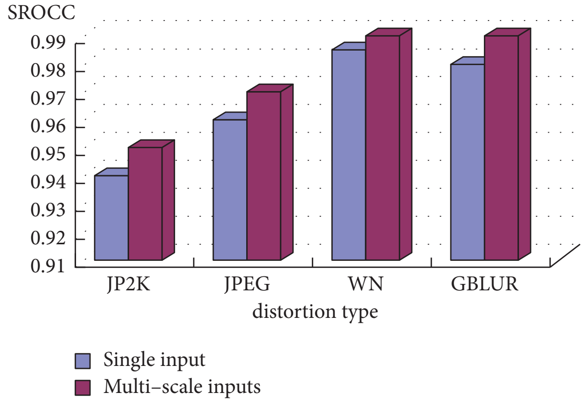

Figure 5: Comparison between single input and multiscale inputs.

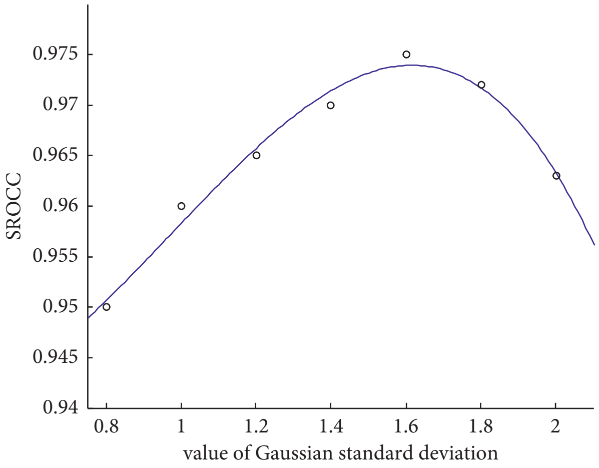

Figure 6: Effect of different $\sigma$ (in the Gaussian function) on SROCC performance with the LIVE dataset.

and perception attributes. BRISQUE [29] extracted natural scene statistics features from locally normalized luminance coefficients. UNIQUE [35] used complex ResNet-34 as its backbone, and its parameters were initialized with the weights pretrained on ImageNet [37]. IQT [36] applied successfully a transformer architecture to a perceptual full-reference IQA task. Our DFCN-IQA demonstrates a state-of-the-art performance with LIVE compared to UNIQUE and IQT, despite a slight weak performance on JP2K. The excellent performance is mainly due to enhanced dual-task learning.

As shown in Tables 2 and 3, the overall performance of our approach is superior to that of the competing methods. In particular, it performed the best for the JPEG, WN, and

Table 2: Comparison of SROCC on LIVE.

| Models | JP2K | JPEG | WN | BLUR | ALL4 |
|---|---|---|---|---|---|
| SSIM | 0.95 | 0.94 | 0.97 | 0.91 | 0.95 |
| CORNIA | 0.91 | 0.95 | 0.95 | 0.90 | 0.92 |
| DIIVINE | 0.84 | 0.82 | 0.88 | 0.88 | 0.84 |
| SOM | 0.94 | 0.95 | 0.98 | 0.97 | 0.96 |
| BIQI | 0.78 | 0.88 | 0.92 | 0.83 | 0.91 |
| SNP-NIQE | 0.92 | 0.97 | 0.97 | 0.96 | 0.96 |
| BRISQUE | 0.91 | 0.91 | 0.95 | 0.94 | 0.93 |
| CNN | 0.95 | 0.97 | 0.97 | 0.96 | 0.96 |
| WaDIQaM | 0.94 | 0.96 | 0.97 | 0.94 | 0.95 |
| BIECON | 0.95 | 0.97 | 0.98 | 0.96 | 0.97 |
| DLIQA | 0.93 | 0.92 | 0.97 | 0.95 | 0.95 |
| MS-C | **0.97** | 0.97 | 0.98 | 0.98 | 0.96 |
| FSI | 0.90 | 0.96 | 0.92 | 0.96 | 0.88 |
| SGDNet | 0.96 | 0.96 | 0.97 | 0.98 | 0.98 |
| UNIQUE | 0.96 | 0.97 | 0.98 | 0.96 | 0.98 |
| IQT | 0.96 | 0.97 | 0.97 | 0.97 | 0.98 |
| DFCN | 0.95 | **0.97** | **0.99** | **0.99** | **0.98** |

The top result is highlighted in boldface in each column.

Table 3: Comparison of PLCC on LIVE.

| Models | JP2K | JPEG | WN | BLUR | ALL4 |
|---|---|---|---|---|---|
| SSIM | 0.95 | 0.98 | 0.98 | 0.92 | 0.96 |
| CORNIA | 0.91 | 0.95 | 0.95 | 0.90 | 0.92 |
| DIIVINE | 0.90 | 0.82 | 0.90 | 0.91 | 0.86 |
| SOM | 0.95 | 0.99 | 0.99 | 0.97 | 0.96 |
| BIQI | 0.78 | 0.93 | 0.93 | 0.83 | 0.92 |
| SNP-NIQE | 0.94 | 0.98 | 0.97 | 0.97 | 0.96 |
| BRISQUE | 0.93 | 0.95 | 0.95 | 0.93 | 0.94 |
| CNN | 0.95 | 0.98 | 0.98 | 0.95 | 0.96 |
| WaDIQaM | 0.96 | 0.94 | 0.94 | 0.94 | 0.96 |
| BIECON | 0.96 | 0.94 | 0.94 | 0.94 | 0.95 |
| DLIQA | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 |
| MS-C | 0.96 | 0.96 | 0.96 | 0.97 | 0.96 |
| FSI | 0.91 | 0.93 | 0.93 | 0.97 | 0.89 |
| SGDNet | 0.96 | 0.96 | 0.96 | 0.97 | 0.97 |
| UNIQUE | 0.96 | 0.98 | 0.97 | 0.98 | 0.97 |
| IQT | 0.97 | 0.98 | 0.98 | 0.97 | 0.97 |
| DFCN | 0.95 | **0.99** | **0.99** | **0.98** | **0.97** |

GBLUR distortions. The network architecture of WaDIQaM-NR [15] is composed of 10 convolutional layers and five pooling layers. MS-C was proposed in 2020, and its basic modes are composed of parallel convolutional networks with fully connected layers. The experimental results demonstrate that our DFCN-IQA outperforms almost all other IQA methods for the four distortion types in LIVE. The comprehensive SROCC and PLCC values ("ALL4") for the proposed DFCN-IQA method are superior to other mainstream methods on the LIVE.

*4.4.2. Performance Comparison on TID2013.* The level and number of distortions in TID2013 far exceed those in LIVE. Accordingly, the IQA method was further tested on TID2013. The experimental results are shown in Figure 7. The compared methods include CaHDC [4], SSIM [5], WaDIQaM-NR [15], SGDNet [31], BIQI [25], BIECON [33], HOSA (high order statistics aggregation) [38], MS-C [34], and FSI [30]. HOSA [38] utilizes improved CORNIA feature sets; its SROCC value is about 11% lower than that of DFCN-IQA. Here, CaHDC [4], WaDIQaM-NR [15], BIECON [33], MS-C [34], and our DFCN-IQA are deep learning-based IQA methods. The DFCN-IQA outperforms other IQA methods, including the four deep learning-based methods just mentioned, on TID2013. CaHDC [4] and SNP-NIQE [28] (published in 2020) are the latest IQA methods, which are an end-to-end blind image quality predictor with the cascaded deep neural network. CaHDC jointly optimized the multilevel feature extraction, hierarchical degradation concatenation, and quality score prediction; its SROCC performance is also slightly lower than that of our approach.

To further test the performance of our approach, the DFCN was compared to IL-NIQE [39], HOSA [38], SNP-NIQE [28], RankIQA [40], Pseudo [41], VRPON [42],

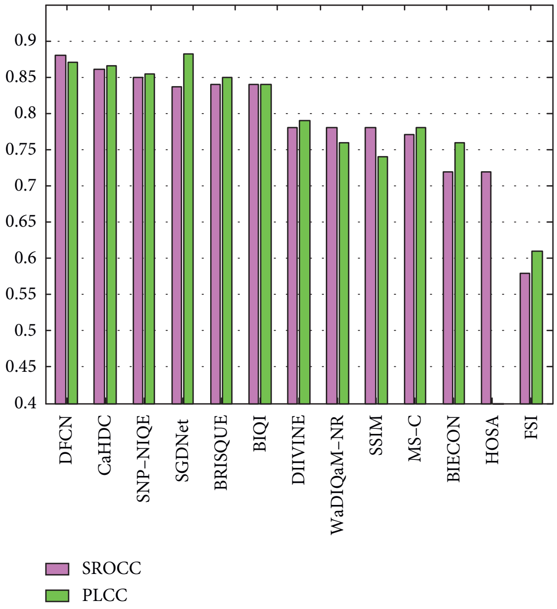

Figure 7: Comparison of performances on TID2013.

MSDD [43], NSSI [44], and DIQA (deep image quality assessor) [45] for each individual distortion in TID2013.

As illustrated in Table 4, our DFCN is superior to the other methods. RankIQA [40] trains networks using ranking quality data and then trains deeper networks through transfer learning; its performance is 7% worse than that of our approach. DIQA [45] combined additionally two handcrafted features with a deep convolutional neural network model. In particular, VRPON [42], MSDD [43], NSSI [44], and DIQA [45] are the latest methods, but their performance is worse than that of our approach. In short, our approach exhibited the higher accuracy on most distortions of TID2013, and its quality prediction was highly consistent with human subjective evaluation scores.

The proposed DFCN significantly outperforms the other methods in Table 4. In particular, it exhibited a relatively high performance for masked noise (#4), mean shift noise (#16), and change of color saturation (#18). This indicates that our approach can effectively address visual linear distortion. The relatively high performances for impulse noise (#6), quantization noise (#7), and local block-wise distortions (#15) show that our approach can also manage visual nonlinear distortion. It has only a slight performance bias for four distortion types.

To summarize, the validation results indicate that the predicted quality of our approach is consistent with human subjective evaluation. According to the current literature, there is no IQA index that is superior to others in terms of all the individual distortion types in TID2013, but our experimental results demonstrate that the proposed DFCN outperforms other IQA metrics on the two benchmark databases: LIVE and TID2013.

*4.4.3. Generalizability Evaluation.* The most commonly encountered distortion types in TID2013 are shared with LIVE, which are JP2K, JPEG, WN, and BLUR. The SROCC and PLCC values of the cross-dataset evaluation with comparable results are displayed in Tables 5 and 6. All models are trained on LIVE and tested on TID2013.

The competing methods were selected to cover a diversity of design philosophies, containing five classic ones (CORNIA [11], IL-NIQE [39], DIIVINE [27], HOSA [38], and BRISQUE [29]) and three state-of-the-art deep learning-based ones (MEON [3], dipIQ [12], and WaDIQaM [15]). DFCN-IQA approach exhibits more powerful generalization ability compared to the deep neural network-based methods (i.e., MEON [3] and WaDIQaM-NR [15]). Unlike early deep neural networks, MEON [3] uses generalized divisive normalization (GDN) as the activation function, which is followed by four convolutional layers and two fully connected layers. The validation results indicate that the simple and highly efficient DFCN-IQA is superior to comparable methods, as it features excellent accuracy as well as strong generalization ability.

Similar to other deep learning-based IQA methods, our approach has common features such as purely data-driven processes and end-to-end optimization. Moreover, we believe that the reasons for the superior performance of DFCN-IQA are as follows: first, the whole images, instead of the segments, are utilized as the inputs to reduce the label noise; second, the dual-branch learning framework enhances the quality score prediction regularized by the weights of the local patches; third, the staged optimization enables the network to decrease the risk of overfitting, resulting in a more robust model.

Table 4: Comparison of SROCC on 24 distortions of TID2013.

| Models | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 | #11 | #12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HOSA | 0.85 | 0.62 | 0.78 | 0.36 | 0.90 | 0.77 | 0.81 | 0.89 | 0.87 | 0.89 | 0.93 | 0.74 |
| RankIQA | 0.67 | 0.62 | 0.82 | 0.36 | 0.76 | 0.73 | 0.78 | 0.80 | 0.76 | 0.86 | 0.87 | 0.70 |
| IL-NIQE | 0.87 | 0.81 | 0.92 | 0.51 | 0.86 | 0.75 | 0.87 | 0.81 | 0.75 | 0.83 | 0.85 | 0.28 |
| SNP-NIQE | 0.88 | 0.73 | 0.65 | 0.74 | 0.87 | 0.80 | 0.86 | 0.86 | 0.61 | 0.88 | 0.88 | 0.29 |
| FSI | 0.71 | 0.72 | 0.70 | 0.72 | 0.77 | 0.70 | 0.26 | **0.95** | 0.83 | 0.86 | 0.90 | 0.36 |
| Pseudo | 0.92 | 0.86 | 0.53 | 0.75 | 0.92 | 0.46 | 0.49 | 0.86 | 0.42 | **0.91** | 0.87 | 0.79 |
| BIQI | 0.34 | 0.20 | 0.70 | 0.18 | 0.61 | 0.02 | 0.67 | 0.89 | 0.79 | 0.78 | 0.88 | 0.55 |
| VRPON | 0.83 | 0.73 | 0.90 | 0.56 | 0.88 | 0.91 | 0.88 | 0.92 | 0.83 | 0.89 | 0.92 | 0.71 |
| MSDD | 0.65 | 0.48 | 0.78 | 0.37 | 0.78 | 0.68 | 0.80 | 0.90 | 0.82 | 0.84 | 0.92 | 0.60 |
| NSSI | 0.92 | 0.86 | 0.93 | 0.83 | 0.94 | 0.89 | 0.83 | 0.92 | 0.84 | 0.90 | 0.92 | 0.68 |
| DIQA | 0.91 | 0.75 | 0.87 | 0.73 | 0.94 | 0.84 | 0.86 | 0.92 | 0.79 | 0.89 | 0.91 | 0.86 |
| DFCN | **0.96** | **0.96** | **0.98** | **0.91** | **0.96** | **0.96** | **0.95** | 0.93 | **0.88** | 0.89 | **0.93** | **0.93** |
| Models | #13 | #14 | #15 | #16 | #17 | #18 | #19 | #20 | #21 | #22 | #23 | #24 |
| HOSA | 0.70 | 0.19 | 0.32 | 0.23 | 0.29 | 0.11 | 0.78 | 0.53 | 0.83 | 0.85 | 0.80 | 0.90 |
| RankIQA | 0.81 | 0.51 | 0.62 | 0.26 | 0.61 | 0.66 | 0.61 | 0.64 | 0.80 | 0.77 | 0.62 | 0.85 |
| IL-NIQE | 0.52 | 0.08 | 0.13 | 0.18 | 0.01 | 0.16 | 0.69 | 0.36 | 0.83 | 0.76 | 0.68 | 0.86 |
| SNP-NIQE | 0.60 | 0.02 | 0.03 | 0.10 | 0.26 | 0.11 | 0.74 | 0.21 | 0.84 | 0.79 | 0.64 | 0.53 |
| FSI | 0.63 | 0.44 | 0.56 | 0.62 | 0.57 | 0.26 | 0.64 | 0.53 | 0.36 | 0.76 | 0.75 | 0.88 |
| Pseudo | 0.49 | 0.01 | 0.23 | 0.11 | 0.18 | 0.38 | 0.86 | 0.07 | 0.60 | 0.68 | 0.73 | 0.79 |
| BIQI | 0.55 | 0.16 | 0.10 | 0.01 | 0.42 | 0.06 | 0.26 | 0.61 | 0.55 | 0.59 | 0.76 | 0.90 |
| VRPON | 0.80 | **0.60** | 0.52 | 0.36 | 0.47 | 0.69 | 0.84 | 0.54 | 0.83 | 0.80 | 0.79 | 0.86 |
| MSDD | 0.64 | 0.21 | 0.15 | 0.21 | 0.42 | 0.12 | 0.38 | 0.62 | 0.60 | 0.68 | 0.78 | 0.90 |
| NSSI | 0.68 | 0.18 | 0.65 | 0.09 | 0.76 | 0.45 | 0.89 | 0.42 | 0.76 | 0.86 | **0.89** | 0.91 |
| DIQA | 0.81 | 0.66 | 0.41 | 0.30 | 0.69 | −0.15 | 0.90 | 0.66 | 0.93 | 0.94 | 0.75 | 0.91 |
| DFCN | **0.82** | 0.33 | **0.76** | **0.76** | **0.90** | **0.93** | **0.96** | **0.78** | **0.96** | **0.93** | 0.82 | **0.93** |

Table 5: Cross-dataset evaluation on SROCC with comparable results from [3].

| Models | JP2K | JPEG | WN | BLUR | ALL4 |
|---|---|---|---|---|---|
| IL-NIQE | 0.91 | 0.87 | 0.89 | 0.82 | 0.88 |
| DIIVINE | 0.86 | 0.68 | 0.88 | 0.86 | 0.80 |
| BRISQUE | 0.91 | 0.89 | 0.89 | 0.89 | 0.88 |
| CORNIA | 0.91 | 0.92 | 0.80 | 0.93 | 0.89 |
| HOSA | 0.93 | 0.92 | 0.84 | 0.92 | 0.90 |
| dipIQ | 0.93 | 0.93 | 0.90 | 0.92 | 0.88 |
| WaDIQaM | 0.95 | 0.92 | 0.94 | 0.91 | 0.89 |
| MEON | 0.91 | 0.92 | 0.91 | 0.89 | 0.91 |
| DFCN | **0.96** | **0.93** | **0.96** | **0.94** | **0.92** |

Table 6: Cross-dataset evaluations on PLCC with comparable results from [3].

| Models | JP2K | JPEG | WN | BLUR | ALL4 |
|---|---|---|---|---|---|
| IL-NIQE | 0.93 | 0.94 | 0.90 | 0.82 | 0.89 |
| DIIVINE | 0.90 | 0.70 | 0.88 | 0.86 | 0.79 |
| BRISQUE | 0.92 | 0.95 | 0.88 | 0.88 | 0.90 |
| CORNIA | 0.93 | 0.96 | 0.78 | 0.94 | 0.90 |
| HOSA | 0.95 | 0.95 | 0.84 | 0.92 | 0.92 |
| dipIQ | 0.95 | **0.97** | 0.91 | 0.93 | 0.89 |
| WaDIQaM | 0.95 | 0.92 | 0.94 | 0.91 | 0.89 |
| MEON | **0.96** | 0.96 | 0.94 | 0.90 | 0.91 |
| DFCN | 0.95 | 0.94 | **0.96** | **0.94** | **0.93** |

## 5. Conclusions

This study established a simple and highly efficient blind image quality predictor that exhibited the superior performance and generalization capability compared to other existing state-of-the-art IQA methods. The multiscale image inputs enrich the extracted distortion feature. A weighted scoring method enhances the mapping from the distorted image to the quality score.

The experimental results verify that our model is effective for both linear and nonlinear distortions, and the test results are highly consistent with human subjective evaluation. However, there are still challenges for deep learning-based IQA with purely data-driven processes. The deep learning model is easy to overfit due to its high complexity. To prevent data overfitting and reduce the parameters, we implemented a series of measures, such as enhancing the data, introducing constraints to the loss functions, and adopting full convolution. Our DFCN can estimate image quality from coarse-to-fine grains, using two-stage method which makes the training easy.

Deep neural networks with far more parameters than training data can also be reliably trained [46], but it is a long way from the current artificial neural network to produce human-like intelligence [47]. The big IQA datasets will contain more types of distortions in the future, which makes it difficult for the existing IQA metrics without retraining to achieve good performance on the specific distortion types. The types of distorted images in the real world are more complicated. Deep learning model needs big data. However, how to use big data is a research topic worth exploring [48]. We did not propose a full-fledged solution but believe that deep learning-based IQA methods will still have lots of potential in the future.

## Data Availability

The IQA data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors would like to thank Ms. Ma Yao for her preliminary work that contributed to this study and Professor Sun Fuming for his helpful discussions and insightful comments. This work was partly supported by the Natural Science Foundation of Liaoning Province under Grant 2019ZD0702.

## References

[1] S. Athar and Z. Wang, "A comprehensive performance evaluation of image quality assessment algorithms," *IEEE Access*, vol. 7, Article ID 140030, 2019.

[2] K. Kang and X. Wang, "Fully convolutional neural networks for crowd segmentation," *Computer Science*, vol. 49, no. 1, pp. 25–30, 2014.

[3] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2018.

[4] J. Wu, J. Ma, F. Liang, W. Dong, G. Shi, and W. Lin, "End-to-end blind image quality prediction with cascaded deep neural network," *IEEE Transactions on Image Processing*, vol. 29, pp. 7414–7426, 2020.

[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. Lin Zhang, Z. Lei Zhang, M. Xuanqin 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.

[7] V. Kumar and V. Singh Bawa, "No reference image quality assessment metric based on regional mutual information among images," 2019, https://arxiv.org/abs/1901.05811.

[8] M. Oszust, "Local feature descriptor and derivative Filters for blind image quality assessment," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 322–326, 2019.

[9] C. G. Bampis, Z. Li, and A. C. Bovik, "Enhancing temporal quality measurements in a globally deployed streaming video quality predictor," in *Proceedings of the Twenty Fifth IEEE International Conference on Image Processing*, pp. 614–618, Athens, Greece, October 2018.

[10] 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, 2013.

[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1098–1105, Providence, RI, USA, June 2012.

[12] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[13] L. Kang, P. Ye, and Y. Li, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1733–1740, IEEE Computer Society, Columbus, OH, USA, June 2014.

[14] K. Y. Lin and G. Wang, "Hallucinated-IQA: no-reference image quality assessment via adversarial learning," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 732–741, IEEE, Salt Lake City, UT, USA, June 2018.

[15] 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 : A Publication of the IEEE Signal Processing Society*, vol. 27, no. 1, pp. 206–219, 2018.

[16] J. You and J. Korhonen, "Transformer for image quality assessment," 2020, https://arxiv.org/abs/2101.01097.

[17] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 117–156, 1998.

[18] M. Zontak, I. Mosseri, and M. Irani, "Separating signal from noise using patch recurrence across scales," in *Proceedings of*

*the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1195–1202, IEEE, Portland, OR, USA, June 2013.

[19] L. Itti and C. Koch, "Computational Modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[20] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," 2005, http://live.ece.utexas.edu/research/quality.

[21] N. Ponomarenko, L. Jin, O. Ieremeiev et al., "Image database TID2013: peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[22] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, USA, May 2015.

[24] Vqeg, *Final report from the video quality experts group on the validation of objective models of video quality assessment*, ITU Telecommunication Standardization Sector., Geneva, Swiss, 2000.

[25] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[26] P. Zhang, W. Zhou, L. Wu, and H. Li, "SOM: Semantic obviousness metric for image quality assessment," in *Proceedings of the 2015 IEEE Conference Computer Vision and Pattern Recognition*, pp. 2394–2402, IEEE Computer Society, Boston, MA, USA, June 2015.

[27] 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, 2011.

[28] Y. Liu, K. Gu, Y. Zhang et al., "Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 929–943, 2020.

[29] 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, 2012.

[30] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and Sparse representation," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 379–391, 2018.

[31] S. Yang, Q. Jiang, W. Lin, and Y. Wang, "SGDNet: an end-to-end saliency-guided deep neural network for No-reference image quality assessment," in *Proceedings of the the Twenty Seventh ACM International Conference on Multimedia*, pp. 1383–1391, Nice, France, October 2019.

[32] H. Weilong Hou, G. Xinbo Gao, T. Dacheng Tao, and fnm Xuelong Li, "Blind image quality assessment via deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1275–1286, 2015.

[33] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, 2017.

[34] Y. Ma, X. Cai, and F. Sun, "Towards No-reference image quality assessment based on multi-scale convolutional neural network," *Computer Modeling in Engineering and Sciences*, vol. 123, no. 1, pp. 201–216, 2020.

[35] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.

[36] M. Cheon, S. J. Yoon, B. Kang, and J. Lee, "Perceptual image quality assessment with transformers," 2021, https://arxiv.org/abs/2104.14730.

[37] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "ImageNet: a large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE Computer Society, Miami, FL, USA, June 2009.

[38] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4444–4457, 2016.

[39] 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, 2015.

[40] X. Liu, J. van de Weijer, and A. Bagdanov, "RankIQA:, Learning from rankings for no-Reference image quality assessment," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1040–1049, IEEE Computer Society, Venice, Italy, October 2017.

[41] 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, 2018.

[42] L. He, Y. Zhong, W. Lu, and X. Gao, "A visual residual perception optimized network for blind image quality assessment," *IEEE Access*, vol. 7, pp. 176087–176098, 2019.

[43] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.

[44] I. F. Nizami, M. U. Rehman, M. Majid, and S. M. Anwar, "Natural scene statistics model independent no-reference image quality assessment using patch based discrete cosine transform," *Multimedia Tools and Applications*, vol. 79, no. 2, pp. 1–20, 2020.

[45] J. Kim, A.-D. Nguyen, and S. Lee, "Deep CNN-based blind image quality predictor," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 11–24, 2019.

[46] N. O. Hodas and P. Stinis, "Doing the impossible: why neural networks can be trained at all," *Frontiers in Psychology*, vol. 9, pp. 1185–1187, 2018.

[47] A. Oleinik, "What are neural networks not good at? On artificial creativity," *Big Data & Society*, vol. 6, no. 1, pp. 25–30, 2019.

[48] M. Roccetti, G. Delnevo, L. Casini, and G. Cappiello, "Is bigger always better? A controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures," *Journal of Big Data*, vol. 6, no. 70, pp. 1–23, 2019.