Blind Dehazed Image Quality Assessment: A Deep CNN-Based Approach

Xiao Lv, Tao Xiang, *Senior Member, IEEE,* Ying Yang, and Hantao Liu

***Abstract*—Research on image dehazing has made the need for a suitable dehazed image quality assessment (DIQA) method even more urgent. The performance of existing DIQA methods heavily relies on handcrafted haze-related features. Since hazy images with uneven haze density distributions will result in uneven quality distributions after dehazing, the manually extracted feature expression is neither accurate nor robust. In this paper, we design a deep CNN-based DIQA method without a handcrafted feature requirement. Specifically, we propose a blind dehazed image quality assessment model (BDQM), which consists of three components: image preprocessing, a haze-related feature extraction network (HFNet), and an improved regression network (IRNet). In HFNet, we design a perceptual information enhancement (PIE) module to learn powerful feature representations and enhance network capability according to channel attention, multiscale convolution and residual concatenation. IRNet aims to aggregate all patch information for the quality prediction of the whole image, where the effect of inhomogeneous distortion from the dehazing procedure is attenuated via a specifically designed patch attention (PA) mechanism. Experimental results on benchmark datasets demonstrate the effectiveness and superiority of the proposed network architecture over state-of-the-art methods.**

***Index Terms*—Dehazed image quality assessment, channel attention, patch attention, multiscale convolution, residual concatenation.**

I. INTRODUCTION

The visibility of natural images captured in hazy weather is degraded by the refraction reaction of air particles to light. This situation greatly impairs the performance of many image processing algorithms and visual-driven applications, such as image segmentation, detection and video surveillance. To eliminate the uncontrollable factors caused by haze in digital image processing, various image dehazing algorithms (DHAs) [1]–[10] have emerged. The performance evaluation of DHAs and the quality assessment of dehazed images (DHIs) not only help to select and optimize DHAs but also monitor the quality of DHIs in real time. Therefore, designing a method to measure the performance of DHAs and evaluate the visual quality of DHIs has become a highly urgent and beneficial endeavor for practical image processing techniques and visually driven systems.

There is a fact that perceptual quality assessment [11]–[13] plays a vital role in the visual communication systems. The last few years have witnessed an explosion of research on dehazed image quality assessment algorithms (DIQAs) [14]–[19], which can be categorized into subjective evaluation methods and objective evaluation methods. Subjective evaluation is the most straightforward and persuasive method since humans are usually the ultimate observers and evaluators of DHIs. However, experimental factors such as a workload, a specific evaluation environment, and a certain number of trained observers hinder the popularization of subjective evaluation methods. Objective evaluation based on advanced computer algorithms overcomes the drawbacks of subjective evaluation, which shows high efficiency and allows a wide range of application scenarios.

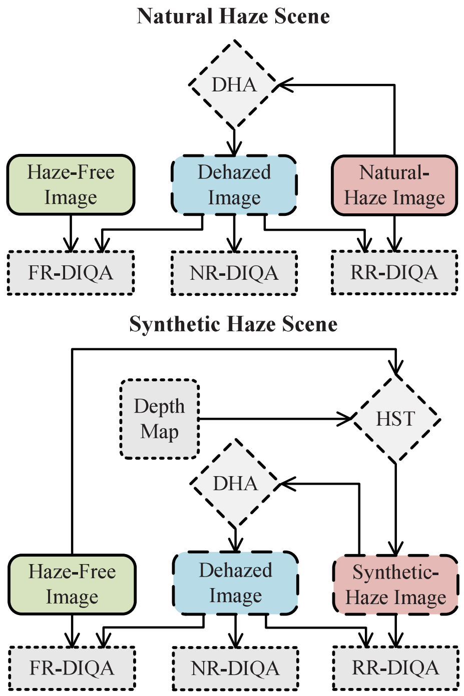


Fig. 1. Three different strategies of DIQA. Solid rectangular boxes indicate images that can be captured from natural scenes. The dashed rectangular boxes indicate images generated by algorithmic processing. FR-DIQA and RR-DIQA use haze-free and hazy images as evaluation references, respectively, while NR-DIQA does not introduce any reference images.

As illustrated in Fig. 1, three strategies can be adopted for objective DIQA by employing different reference images: full-reference dehazed image quality assessment (FR-DIQA) [16], [18], [19], reduced-reference dehazed image quality assessment (RR-DIQA) [20], [21], and nonreference dehazed image quality assessment (NR-DIQA). In traditional IQA tasks [15],

This work was supported by the National Key R&D Program of China under Grant 2022YFB3103500, the National Natural Science Foundation of China under Grants 62072062 and U20A20176.

X lv and T Xiang are with the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: {xiaolv; txiang}@cqu.edu.cn).

Y Yang is with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong, China (e-mail: yyang@math.cuhk.edu.hk).

H Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff, UK (e-mail: liuh35@cardiff.ac.uk).

[22]–[25], FR and RR methods are categorized depending on the degree of ground truth information employed. However, DIQA distinguishes between FR and RR based on the reference image used. The former takes the haze-free image as a reference, while the latter uses the hazy image. FR-DIQA and RR-DIQA methods achieve satisfactory performance by comparing the references and the DHIs. However, they are ill-posed tasks, as haze-free and hazy images are usually unavailable in practical applications. Thereafter, researchers take synthetic-haze images to obtain the reference images required for experiments [16], [21], which are generated by haze synthesis technologies (HSTs) from the corresponding depth maps of haze-free images [26].

In contrast, NR-DIQAs take only DHIs as input for quality evaluation without a reference image, which is more realistic and has received substantial attention in recent years [14], [27]. However, an ill-posed definition becomes the most essential issue for NR-DIQA to ensure good image quality predictions. Numerous efforts have been made to address this problem by establishing robust feature representation models [28]–[31]. Traditional NR-DIQA methods commonly employ hand-designed haze-related feature representations [14], [27], [32], [33], and they lack the diversity and flexibility to capture complex distortion patterns and various image content. Additionally, haze is often randomly distributed and unevenly dispersed in the image. However, existing NR-DIQA methods evaluate visual quality based on the whole image without considering the variation in haze density in different image areas, which results in a performance bottleneck.

To track the abovementioned issues of NR-DIQA, in this paper, we propose an end-to-end blind dehazed image quality assessment model (BDQM) by employing a convolutional neural network (CNN). Due to the powerful feature representation ability of CNN, the extracted deep perceptual features can be effectively combined and used for regression training. The proposed BDQM can automatically learn perceptual-related feature representations and predict visual quality levels without heavy manual work and reference images. Specifically, we propose a perceptual information enhancement (PIE) module based on channel attention, multiscale convolution and residual concatenation. PIE can capture the long-range dependencies between channels and distinguish different types of information for powerful feature representation. In addition, different image regions have different haze densities, which have different effects on image quality assessment. To minimize the effect of different haze densities on the dehazing results, we propose a patch attention (PA) mechanism to capture the dependence of perceptual features between different image patches. This scheme mitigates the performance bottleneck caused by heterogeneous patch information and improves the accuracy of the predicted image quality scores.

The proposed BDQM consists of three parts: image preprocessing, a haze-related feature extraction network (HFNet), and an improved regression network (IRNet). The image preprocessing module contains image sampling and image normalization procedures, which are necessary operations for deep learning models. HFNet takes the sampled patches of the DHI as input and perceptual features that are highly sensitive to the visual quality outputs. In HFNet, a PIE module with channel attention, multiscale convolution and residual concatenation is designed, which explores significant information and extracts multiscale perceptual features for quality prediction. IRNet integrates the features of all patches sampled from one DHI and predicts the final quality score of the entire image through a carefully designed PA scheme. Extensive experiments on five datasets demonstrate that our BDQM has considerable competitive advantages over existing state-of-the-art algorithms.

The main contributions of this paper are as follows:

- To our knowledge, this study is the first to use CNN to explore the NR-DIQA problem. We propose the BDQM to automatically learn perceptual feature representations and predict quality scores of DHIs without manual work or reference images.
- We design a PIE module with channel attention and multiscale convolution to extract haze-related perceptual features for quality prediction. PIE uses feature splicing to capture valuable information and has a powerful network ability for diverse and flexible feature representation.
- We design a PA mechanism to adaptively aggregate perceptual features of all patches sampled from one DHI. The PA mechanism avoids a mismatch between the local quality and the overall image quality due to the uneven distribution of haze density and overcomes the difficulty of separately predicting the quality score of each patch during the training process.
- We conduct extensive experiments to validate the performance of our method on five datasets. Compared to state-of-the-art metrics, our model exhibits superior performance with acceptable computational cost.

The remainder of this paper is organized as follows. Section II reviews related work on haze-related image datasets, DIQA schemes and CNN-based IQAs. Section III details the construction of the proposed BDQM. Section IV presents the experimental results and their analysis. Finally, Section V concludes this paper.

II. RELATED WORK

Generally, DIQA is designed to evaluate the quality of DHIs, which are generated from hazy images by DHAs. In this section, we review the existing haze-related image datasets (including hazy and dehazed image datasets), related state-of-the-art DIQAs and CNN-based IQAs.

A. Haze-Related Image Datasets

1) Hazy Image Datasets: Hazy image datasets are critical to measuring the performance of DHAs. Three types of hazy image datasets are widely used in current research. The first is synthetic-haze image datasets, such as FRIDA [34], FRIDA2 [35], D-HAZY [26], Foggy Cityscapes [36], and RESIDE [37], which are generated from clear images and depth maps processed by computer synthesis techniques. However, their shortcomings, such as low resolution, unrealistic images, and poor simulation of synthetic-haze images, lead to an inaccurate and unsatisfactory measurement of DHAs. The second is

TABLE I
HAZY IMAGE DATASETS AND DEHAZED IMAGE DATASETS

	Haze Type	Dataset	Haze-Free Images	Hazy Images	Dehazed Images	Dehazing Algorithms
Hazy Image Datasets	Synthetic-Haze	FRIDA [34]	18	72	-	-
		FRIDA2 [35]	66	264	-	-
		D-HAZY [26]	1449	1449	-	-
		Foggy Cityscapes [36]	550	550	-	-
		RESIDE (ITS) [37]	1399	13990	-	-
	Artificial-Haze	O-HAZE [38]	45	45	-	-
		I-HAZE [39]	35	35	-	-
	Natural-Haze	BeDDE [18]	23	208	-	-
		Foggy Driving [36]	-	101	-	-
Dehazed Image Datasets	Natural-Haze	DHQ [17]	-	250	1750	7
		exBeDDE [18]	12	167	1670	10
		MRFID [19]	200	800	12800	16
	Synthetic-Haze	SHRQ-Regular [16]	45	45	360	8
		SHRQ-Aerial [16]	30	30	240	8
	Hybrid (Natural & Synthetic)	IVCDD [40]	-	25	200	8

artificial-haze image datasets, such as O-HAZE [38] and I-HAZE [39], which are captured from machine-made hazy scenes both indoors and outdoors. Although artificial-haze image datasets are better than synthetic-haze image datasets in haze simulation and the restoration of real scenes, they still have gaps with natural-haze images. The third is natural-haze image datasets, which are real but difficult to assemble. For example, the BeDDE [18] dataset contains 208 image pairs collected from 23 provincial capital cities, and the Foggy Driving [36] dataset contains 101 color images depicting real-world hazy driving scenarios. However, hazy images serve as the processing objects of DHAs, which lack the corresponding dehazed results and cannot be directly used in DIQA tasks.

2) Dehazed Image Datasets: The crucial task in DIQA research is to establish a public dehazed image dataset. Existing dehazed image datasets can be classified into three categories according to the type of hazy image. One type is the synthetic dehazed image dataset, such as SHRQ-Regular [16] and SHRQ-Aerial [16], which are derived from synthetic-haze images based on reliable HSTs. Another is the dehazed image dataset for natural haze, such as DHQ [17], exBeDDE [18], and MRFID [19], which are generated from natural haze scenes. The last type is the hybrid haze image dataset, which contains dehazed images created from synthetic-haze and natural-haze images, such as IVCDD [40]. Table I provides specific information on the abovementioned datasets.

B. Dehazed Image Quality Assessment Schemes

Three strategies for objective DIQAs have been extensively studied, namely, FR-DIQAs, RR-DIQAs, and NR-DIQAs.

1) FR-DIQAs: FR-DIQAs exhibit superior performance by measuring the difference between DHIs and haze-free images. Liu et al. [19] developed a similarity index (FRFSIM) based on haze-related features. Specifically, dark channel [3] and MSCN features [41] were employed to measure the haze density similarity. Gradient and chrominance features were used to evaluate the variation in artificial distortion. Zhao et al. [18] proposed the visibility index (VI) and realness index (RI) by evaluating the visibility and realness restoration of DHIs independently. VI was designed based on dark channel [3] and gradient features, while RI utilized phase congruence [42] and chrominance features [43]. Min et al. [16] proposed a quality measure by integrating three components: image structure recovery, color rendition, and overenhancement. Obviously, it is reliable to explore haze-related features for visual quality score prediction based on the difference between the reference images and the DHIs. However, FR-DIQA is impractical for real-time applications due to the difficulty of acquiring real haze-free and hazy image pairs.

2) RR-DIQAs: RR-DIQAs take hazy images as references to evaluate the visual quality of DHIs. Song et al. [44] proposed a contrast enhancement index based on the newly proposed haze-line theory. The underlying principle of haze-line theory is that haze lines in a hazy image respect the color clusters in the corresponding haze-free image, and pixels belonging to the same color cluster have similar colors. In addition to the color contrast, Fang et al. [45] considered structural similarity features for RR-DIQA. Specifically, the ascension of contrast degree and structural similarity were measured by comparing the spatial frequency contrast and the edge consistency between the hazy images and the DHIs, respectively. However, it was not sufficient and convincing to focus on only these two types of features for the visual quality evaluation of DHIs. Hsieh et al. [20] proposed an objective assessment of haze removal based on two objective optimizations, the dehazing effect and image distortion. Wang et al. [21] made full use of dark channel features and proposed a pixel-level dehazed image quality assessment method (PDIQA). The advantages of pixel-level quality measures allowed the metric to focus on specific areas. Min et al. [17] proposed an objective dehazing quality index (DHQI) by fusing three groups of features: haze-removing features, structure-preserving features, and overenhancement. However, RR-DIQA emphatically focuses on the degree of haze removal in DHIs compared to hazy images, which is still a gap in haze-free images. It performs slightly worse than FR-DIQA in terms of the quality assessment of DHIs.

3) NR-DIQAs: NR-DIQAs without reference images have received widespread attention in recent years owing to their practicality and convenience. Choi et al. [14] proposed a blind fog aware density evaluator (FADE) based on manually-extracted haze characteristics: low contrast, faint color, and shifted brightness. Shen et al. [27] extracted information, contrast, and luminance to train a SVR model, which can be further used for quality prediction. Zhang et al. [32] learned a multivariate Gaussian (MVG) model of image patches from a collection of pristine natural images, then measured the quality of each image patch using the Bhattacharyya distance [46], and finally obtained the overall quality score by average pooling. Liu et al. [33] learned the pristine MVG model by extracting structure and naturalness from natural images, and visual quality was defined as the distance between the MVG model of DHIs and the learned pristine model.

The performance of the abovementioned NR-DIQAs depends highly on the operations of manually extracted or MVG-generated features. Thus, further efforts should be made to improve diverse feature representations for the quality assessment of DHIs. Haze density variation in different image regions causes uneven distortion after applying the DHAs. However, existing NR-DIQAs do not consider the impact of local distortion on overall image quality, leading to inaccurate quality evaluation results.

C. CNN-Based IQAs

Due to the powerful learning ability of CNNs, numerous CNN-based IQAs [30], [47]–[50] have been gradually developed and have achieved remarkable success by directly mapping the input image to a quality score. Kang et al. [47] were pioneers in applying CNNs for IQAs by designing a shallow network structure consisting of one convolutional layer with max and min pooling, two fully connected layers, and one output node. This network took image patches as input and estimated the overall image quality by pooling the scores of the sampled patches. Bianco et al. [48] estimated image quality by averaging the sum of scores predicted from multiple subregions of the original image. Yan et al. [30] proposed a dual-stream CNN using the image and gradient image to capture different-level information of inputs. Po et al. [50] enhanced CNN-based IQAs by discarding homogenous patches and biasing the final image quality score toward patches with complex structures via weighted average variance. Zhang et al. [51] proposed a CNN-based method named HazDesNet to predict haze density. HazDesNet took hazy images as input and predicted a pixel-level haze density map. The density map was then refined and smoothed, and the average of the refined map was calculated as the global haze density.

Although the application of CNNs to IQA tasks has produced pleasing results, this approach has not been applied to the DIQA task due to the following challenges. First, the difficulty of acquiring haze-free and hazy image pairs makes it difficult to advance CNN-based FR-DIQA and RR-DIQA methods. Second, most models do not effectively use information from the images when finding correlations between images and scores and neglect the correlations between patches. Finally, these methods estimate the subjective scores of the patches as the subjective score of the entire image, ignoring the effect of local quality inhomogeneities.

III. THE PROPOSED BDQM METRIC

In this section, we elaborate on our proposed BDQM. First, we describe the core components of our network: image preprocessing, HFNet, and IRNet. Then, we present the details of the training procedure.

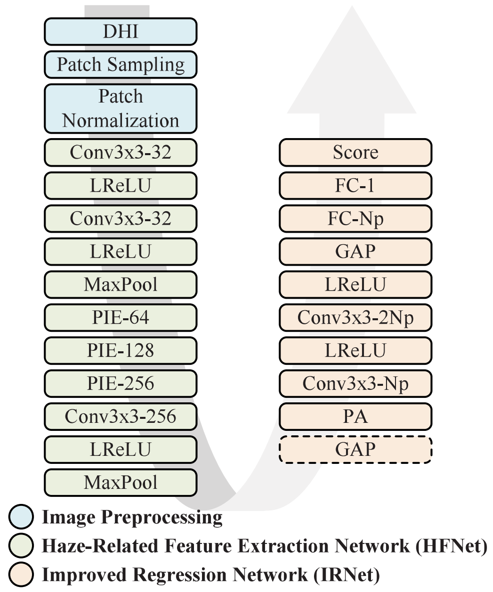


Fig. 2. The framework of the proposed BDQM.

A. Framework

The overall framework of the proposed BDQM is shown in Fig. 2, which contains three parts: image preprocessing, HFNet, and IRNet. Image preprocessing contains patch sampling and patch normalization. HFNet aims to capture efficient perceptual haze-related features for quality regression. In HFNet, we design the PIE module to maximize the haze-related feature enrichment. IRNet aggregates and optimizes all patch features via the PA scheme. The final quality score of the image is derived from all input patches through carefully designed convolution operations.

B. Image Preprocessing

Before training, some preprocessing operations need to be performed on the input images.

1) Patch Sampling: First, we divide the input DHI into nonoverlapping patches of size $m \times m$ [28], [29], [52], [53], which is beneficial for capturing the inhomogeneous quality of the image and ensuring a large number of training samples. Then, we select N_p patches of the input image as training data. Given an input image of size $h \times w$, at most N_m patches of size $m \times m$ can be obtained:

$$N_m = \lfloor \frac{h}{m} \rfloor \times \lfloor \frac{w}{m} \rfloor \tag{1}$$

2) Patch Normalization: Image normalization has a pivotal role in the training procedure of neural networks, which is effective for stability training and network convergence. We perform patch normalization to convert all image patches to the range [0, 1] and use the converted patches as the input of

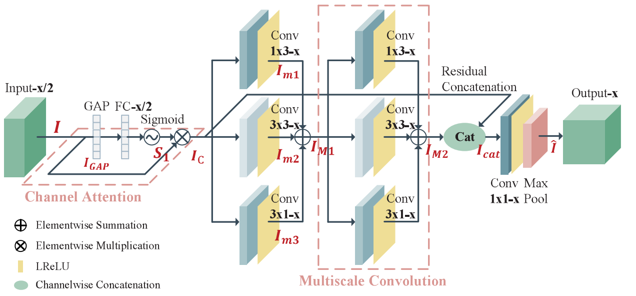
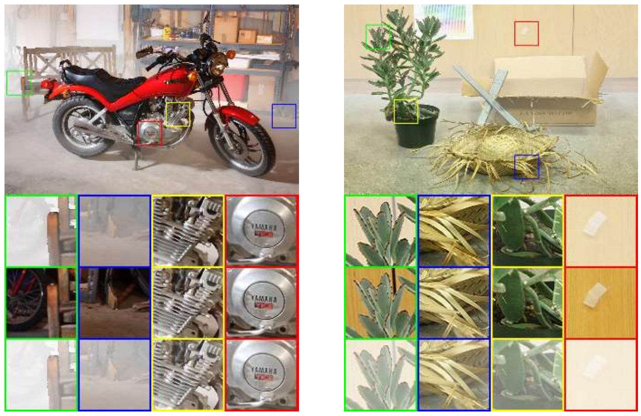


Fig. 3. The architecture of PIE-X.



(a) (b)

Fig. 4. An example shows that patches in different regions of the DHI have different visual qualities. The topmost images in (a) and (b) are the DHIs from the SHRQ-Regular [16] dataset. The three rows of patches below are taken from the corresponding positions of the DHIs, haze-free images and hazy images.

the HFNet. Given an image patch P_i in RGB color space, we perform the normalization operation as follows:

$$\widetilde{P}_i = \frac{P_i}{255} \tag{2}$$

C. Haze-Related Feature Extraction Network (HFNet)

HFNet is a single-stream network for the efficient extraction of haze-related features. First, the simple combination of the convolutional layer and the pooling layer cannot meet the complex feature representation requirements. Inspired by previous work [49], [54], [55], we propose a PIE module with multiscale convolution to effectively increase the feature extraction capability by simulating the complex visual recognition progress of the human visual system (HSV). Second, the local quality of the DHI is unevenly distributed, which has a significant impact on the overall image quality assessment. Therefore, we propose to apply channel attention to enhance useful information and suppress useless information. Finally, to reduce information loss during the convolution procedure, we employ feature splicing to retain more useful perceptual information for the quality prediction of the DHI.

The detailed architecture of HFNet is shown in Fig. 2. First, two 3×3 convolutional layers and one maxpooling layer are applied to extract low-level generic features. Then, the task-specific semantic features are derived by a tandem group of PIE modules. Finally, a feature pooling step is used to aggregate the learned haze-related perceptual features, including one 3×3 convolutional layer and one 2×2 maxpooling layer. We will discuss the impact of different HFNet structures on network performance in Section IV-C1.

PIE Module: We propose a PIE module to replace the simple combination of a single convolutional layer and a pooling layer [30], [47] to enhance the network capability for feature representation.

Fig. 3 shows the detailed structural information of the PIE-$X, X \in \{64, 128, 256\}$. For each PIE-X module, given an input feature I of size $X/2 \times H \times W$, it is first subjected to a simple channel attention mechanism to obtain I_C. Specifically, the one-dimensional channelwise statistics I_{GAP} of length $X/2$ are obtained from I after the global average pooling operation (GAP). I_{GAP} is defined as:

$$I_{GAP} = F_{GAP}(I) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} I(i,j) \tag{3}$$

where F_{GAP} denotes the GAP operation. The fully connected (FC) layer and the sigmoid activation function are applied on I_{GAP} to generate the attention vector S_1. Then, the channel attention feature map I_C of size $X/2 \times H \times W$ is generated with channel attention weights S_1 as follows:

$$I_C = S_1 \otimes I \tag{4}$$

where $\otimes$ denotes the elementwise multiplication operation.

Furthermore, I_{M2} is generated from I_C by two tandem multiscale convolution layers [56]. In the first multiscale convolution layer, I_C generates I_{m1}, I_{m2} and I_{m3} through three convolutional layers with kernel sizes of 1×3, 3×3 and 3×1, respectively. Then, I_{M1} of size $X \times H \times W$ is generated by aggregating I_{m1}, I_{m2} and I_{m3} as follows:

$$I_{M1} = I_{m1} \oplus I_{m2} \oplus I_{m3} \tag{5}$$

where $\oplus$ denotes the elementwise summation operation. Up to this point, the result of the first multiscale convolution layer is output as I_{M1}. Then I_{M1} proceeds to the next multiscale convolution layer and generates I_{M2} of size $X \times H \times W$.

Next, a residual connection is applied to I_C and I_{M2} by the splicing operation. The splicing operation of I_C and I_{M2} is performed in the channel dimension, so the size of the splicing result I_{cat} is $(X+X/2)\times H\times W$. Finally, a 1×1 convolutional layer and a maxpooling layer are responsible for dimension reduction and downsampling. Then, the output of the PIE-X module is $\widehat{I}$ of size $X \times H \times W$. Notably, the leaky rectified linear unit (LReLU) activation function [57] is used after all convolutional layers. We will discuss how the residuals are connected and verify the best choice between splicing and summation in Section IV-C2.

D. Improved Regression Network (IRNet)

1) Structure of Previous Regression Network: Specifically, in the previous CNN-based IQAs [30], [47], each patch inherits the same annotation score. Distortions in an image are usually unevenly distributed, which means that the level

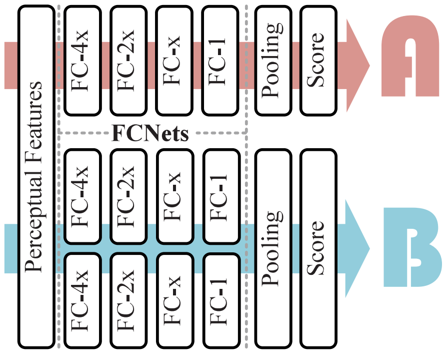


Fig. 5. Two mainstream regression network structures for CNN-based IQAs.

of the visual quality of each patch is not flush with the whole image. Fig. 4 shows two examples. The topmost images in Fig. 4 (a) and (b) are the DHIs from the SHRQ-Regular [16] dataset, and the three rows of patches below are taken from the corresponding positions of the DHIs, haze-free images and hazy images, respectively. These two examples show that the color, texture, and structure of the dehazed patches are close to those of haze-free patches. However, there are still some areas that are visually different between DHIs and haze-free images. The uneven distribution of visual quality makes it difficult to estimate the visual quality of the entire DHI.

As shown in Fig. 5, there are two mainstream regression networks in existing CNN-based IQAs based on FC networks (FCNets). One is the single pooling approach [30], [31], [58] in Fig. 5 (A). The input and the output of FCNet are the perceptual features and the learned patch score o_i. FCNet consists of several cascaded FC layers, the number of FC layers can be adaptively adjusted. We set it to 4 here. The final quality score O of the DHI is generated by pooling the scores of all sampled patches, which is defined as:

$$O = \sum_{i=1}^{N_p} o_i \tag{6}$$

where N_p is the number of sampled image patches. The other one is the weighted pooling method [59] in Fig. 5 (B), where two parallel FCNets are responses to learning the quality score o_i and the weight score w_i for each image patch, respectively. The weight score represents the proportion of the image patch in the overall quality of the entire image. The final quality score O of the DHI is defined by:

$$O = \frac{\sum_{i=1}^{N_p} o_i w_i}{\sum_{i=1}^{N_p} w_i} \tag{7}$$

where N_p is the number of sampled image patches.

There are ill-posed requirements in the abovementioned two types of regression networks, that is, each sampled image patch needs to be labeled with subjective scores for training. The literature [30] usually adopts the way that an image patch inherits the subjective score (ground truth) of its parent image. However, this approach is not appropriate for DIQA. Due to the uneven distribution of haze density in DHIs, different image patches with different haze densities should have different quality scores. Therefore, assigning the same labeled subjective score to all patches sampled from an image may lead to inaccurate and unreliable prediction results.

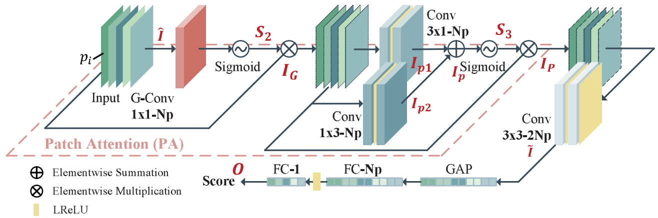


Fig. 6. The architecture of IRNet.

2) IRNet: Motivated by the above observations, we propose IRNet to avoid the mismatch between local and overall image quality due to uneven distribution of haze density and to overcome the difficulty of obtaining the ground truth of the patches for training. The proposed IRNet aims to map all patch features of the DHI to the final perceptual quality score. Fig.2 illustrates the main structure of IRNet, while Fig.6 refines each part to make it easier to understand. The input of IRNet is the perceptual features of size $N_p \times X \times H \times W$ from the output of HFNet. To combine the features of each patch to estimate the final score of the image, a reshaping operation is required to convert the size to $1 \times N_p \times h \times w$, where $X = h \times w$. The GAP operation of IRNet in Fig.2 ensures that the values of H and W are 1.

PA Module: We design a PA mechanism to adaptively aggregate perceptual features of all patches sampled from one DHI. For a given perceptual input feature $\widehat{I}$ of size $N_p \times h \times w$, the feature map I_P is obtained by the PA mechanism. In detail, $\widehat{I}$ generates an attention vector S_2 by 1×1 group convolution (groups=N_p) and sigmoid activation function. Then, the first attention feature map I_G of size $N_p \times h \times w$ is generated with the attention weights S_2 as follows:

$$I_G = S_2 \otimes \widehat{I} \tag{8}$$

where $\otimes$ denotes the elementwise multiplication operation. Next, multiscale convolutions are conducted by using 3×1 and 1×3 kernel sizes for I_{p1} and I_{p2}, respectively. The multiscale feature maps are aggregated into I_p:

$$I_p = I_{p1} \oplus I_{p2} \tag{9}$$

where $\oplus$ denotes the elementwise summation operation. Then, a sigmoid activation function is applied to I_p to generate the attention vector S_3. Similarly, the PA feature map I_P can be obtained by:

$$I_P = S_3 \otimes I_G \tag{10}$$

where $\otimes$ denotes the elementwise multiplication operation.

Last, a 3×3 convolutional layer is applied on I_P to obtain $\widetilde{I}$ of size $2N_p \times h \times w$. The GAP and two FC layers (FC-N_p, FC-1) are used to learn the final quality score O from the features $\widetilde{I}$. In the proposed IRNet, instead of predicting the visual score of each patch individually, we fuse the perceptual features of all sampled patches to obtain the visual quality score of the whole image. In other words, the object of our IRNet is the whole image rather than a single patch.

E. Network Training

For the proposed BDQM, our target is to minimize the loss function. Given an input image I_i, O_i and S_i are its predicted quality score and subjective quality score, respectively. We adopt the l_1-norm loss to measure the difference between O_i and S_i, which is formulated as

$$loss(i) = \|O_i - S_i\|_1 \tag{11}$$

By minimizing the loss function, we can obtain the optimal network parameters as follows:

$$\theta^* = \underset{min}{\theta} \frac{1}{N} \sum_{i=1}^{N} loss(i) \tag{12}$$

where N denotes the number of training images.

The backpropagation algorithm [60] is employed to iteratively train the proposed model over multiple epochs. We plit the training set into minibatches for batch optimization, and each sample in the training set is used only once in an epoch. In our experiments, image patches sampled from the same image must be distributed in the same minibatch, because IRNet integrates the features of the input patches for overall score prediction. The Adam Optimizer (ADAM) [61] is utilized to change the conventional stochastic gradient descent method for better convergence of batch optimization.

IV. EXPERIMENTS AND ANALYSIS

In this section, we conduct extensive experiments to demonstrate the performance of our proposed BDQM on five test datasets and compare it with that of state-of-the-art IQAs.

A. Experimental Protocols

1) Test Datasets: Five benchmark dehazed image datasets are used to evaluate the performance of our proposed method. They are two synthetic dehazed image datasets SHRQ-Regular [16] and SHRQ-Aerial [16], two natural dehazed image datasets DHQ [17] and exBeDDE [18], and one hybrid dehazed image dataset IVCDD [40]. The details of the five datasets are presented in Table I.

SHRQ-Regular: The SHRQ-Regular consists of 360 DHIs created from 45 synthetic-haze images using 8 DHAs. The synthetic-haze images are derived mainly from high-quality haze-free images and the corresponding depth maps from [62] and Middlebury Stereo datasets [63]. The standard double-stimulus method with a five-grade continuous quality scale is used to conduct the subjective experiment test [64]. First, 38 observers are invited to rate a quality score for each dehazed image based on their won visual experience. Then the outlier detection operations are employed to process the obtained raw subjective scores, and invalid observers are rejected. Finally, the MOS value of each image is generated by averaging all the rating scores from the valid observers. The MOS values lie in the range [10, 80], where a higher value indicates better visual quality.

SHRQ-Aerial: The SHRQ-Aerial dataset consists of 240 DHIs created from 30 synthetic-haze images using the same 8 DHAs in the SHRQ-Regular dataset. The synthetic-haze images are generated from high-quality aerial images in the AID [65] dataset. The subjective test is the same as the SHRQ-Regular. The MOS values in SHRQ-Aerial are in the range of [10, 80], with higher values indicating better visual quality.

exBeDDE: The exBeDDE dataset is an extension of BeDDE [18]. It contains 167 hazy images from 12 cities in BeDDE. Of these, 1670 DHIs are generated from 10 representative DHAs. In the subjective experiment, the images were divided into 12 hazy groups and 167 dehazing groups. 10 subjects rank each dehazing group based on realness and visibility by the double-stimulus method. The final MOS of an image is converted from its rank in a group by a specially designed mapping function [18]. The MOS values in exBeDDE lie in the range [0, 1], where a higher value indicates better visual quality.

DHQ: The DHQ dataset contains 1750 DHIs generated from 250 hazy images of various haze densities using 7 representative DHAs. The 250 hazy images selected from the total 500 hazy images [66] are obtained from the real world. Subjective experiments are performed by 54 observers by using the double-stimulus method with a five-grade categorical rating scale. After the outlier detection, three subjects are rejected. The MOS values are in the range [20, 80], where a higher value indicates better visual quality.

IVCDD: The IVCDD dataset includes 25 hazy images and 200 DHIs created by 8 DHAs. Most images in this dataset are captured in the real world, but the haze of indoor static objects is simulated artificially. This single-stimulus method is used in the subjective experiment test. 24 observers are participated to rate their quality of each image based on an integer scale from 1 to 10, where higher values indicate better visual quality.

2) Evaluation Criteria: Based on the recommendations of the video quality expert group (VQEG) [67] for the first phase of FR-TV testing, four common performance evaluation criteria are used to evaluate the performance of different DIQA metrics. Pearson linear correlation coefficients (PLCCs) and Spearman rank order correlation coefficients (SRCCs) are employed to evaluate the prediction monotonicity, while the Kendall rank order correlation coefficients (KRCCs) and root mean squared error (RMSE) are used to evaluate the prediction accuracy. To obtain accurate PLCC and RMSE, the predicted score is processed by a five-parameter nonlinear logistic mapping function [68]:

$$O' = \beta_1\left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(O-\beta_3)}}\right) + \beta_4 O + \beta_5 \tag{13}$$

where O denotes the predicted score, and O' denotes the fitted predicted score. $\beta_i(i = 1, 2, ..., 5)$ are the parameters to be fitted. A better DIQA method should have a higher SRCC, KRCC, and PLCC but a lower RMSE.

3) Protocol Configuration: We select five dehazed image datasets in Table I for our experiments. The performance results on each dataset are obtained by training the proposed model on 80% images of the dataset while testing on the remaining 20%. According to patch sampling method in Section III-B1, we divide the input DHI into nonoverlapping patches of size $m \times m$ $(m = 32)$ and select $N_p = N_m$ patches for training. The MAE loss function with a learning rate of $\eta = 10^{-4}$ is adopted by the ADAM optimizer with

TABLE II
THE RESULTS OF PARAMETER ANALYSIS

Parameter	Setting	SHRQ-Aerial			
		SRCC	KRCC	PLCC	RMSE
Patch Size	48	0.9673	0.8635	0.9575	0.0715
	64	0.9645	0.8564	0.9460	0.0790
	80	0.9633	0.8475	0.9591	0.0685
	96	0.9436	0.8191	0.9432	0.0788
	112	0.9429	0.8121	0.9375	0.0825
	128	0.9307	0.7801	0.8900	0.1069
Patch Number	3/4	0.9491	0.8209	0.9388	0.0912
	1/2	0.9357	0.7943	0.9206	0.0918
	1/4	0.9332	0.8032	0.9331	0.0933
Color Space	HSV	0.9657	0.8528	0.9445	0.0952
	LAB	0.8909	0.7199	0.8422	0.1278
Kernel Size	5x5	0.9525	0.8245	0.9457	0.0835
	7x7	0.9629	0.8511	0.9511	0.0820
Residual Connection	Summation	0.9718	0.8723	0.9635	0.0634
	False	0.9694	0.8670	0.9409	0.0796
BDQM		**0.9767**	**0.8848**	**0.9613**	**0.0682**

the parameter settings recommended in [61]. Moreover, we add the exponential decay method to reduce the learning rate based on the training progress. Furthermore, all convolution layers and FC layers in our BDQM network are followed by LReLU activation function. Dropout regularization is applied to each FC layer with a probability of 0.5. All experiments are performed under 10 iterative operations with randomly divided training data and median scores for the evaluations are reported. The best scores for each evaluation criterion are highlighted in bold in the tables.

B. Parameters Analysis

Experiments demonstrate that different parameters have different impacts on the performance of our proposed BDQM. In this section, we investigate and discuss how these parameters affect the performance of BDQM and determine the optimal configuration of these parameters. We conduct experiments in this section on the SHRQ-Aerial [16] dataset. Table II shows the results of the experiments with various parameters. The parameters of the proposed BDQM are set as follows: patch size: 32, patch number: all patches, color space: RGB, kernel size: 3×3, residual connection: concatenation.

1) Patch Size: We split the DHI into patches with nonoverlapping contents for training. As described in Section III-B1, for a given DHI, the maximum number of patches varies according to patch size. The purpose of segmenting patches is to capture the inhomogeneous quality of the DHI; patches that are too small increase computational overhead, while patches that are too large lead to unstable quality estimates. We examine the performance of BDQM under different patch sizes m. Here, all the patches sampled from the image are fed into the network for training. Table II shows that BDQM achieves the best performance on the SHRQ-Aerial dataset with patch size $m = 32$.

2) Patch Number: Once the patch size m is determined, the number of patches N_p is also determined. Obviously, there is an upper limit to the number of sampled patches since we split them so that the content is not duplicated. Patches selected for training directly impact the quality assessment of the whole image. If too few patches are selected, other information about the image may be lost in training. Table II shows the performance results of selecting 1/4, 1/2, and 3/4 of the patches, and all the patches. The best result is achieved by using all image patches for training.

3) Color Space: To explore the impact of the image color space, we compare the network performance in the RGB, HSV, and LAB color spaces. RGB is defined by the chromaticity of the three primary colors: red, green, and blue. HSV represents the hue, saturation, and lightness. In LAB color space, L denotes the luminance, and A and B denote the opposing color dimensions. We normalized the color space components to [0, 1] for training, and the experimental results are shown in Table II. Obviously, the proposed model trained in the RGB color space has better results.

4) Kernel Size: To verify the impact of the kernel size, we use different kernel sizes to train our model and test the corresponding performance. Here, we change only the kernel size while keeping the other structures unchanged. The experimental results of different kernel sizes in Table II show that the network performance is sensitive to kernel size. Therefore, we use the kernel size 3×3 in our method since it achieves the best performance on the SHRQ-Aerial dataset.

C. Ablation Study

We conduct an ablation study to demonstrate the effectiveness of our proposed PIE module and IRNet. The experiments are conducted by comparing the proposed model with several baseline models on SHRQ-Aerial [16] and SHRQ-Regular [16]. The same experimental conditions are set as in Section III-E to ensure the validity and usability of the results.

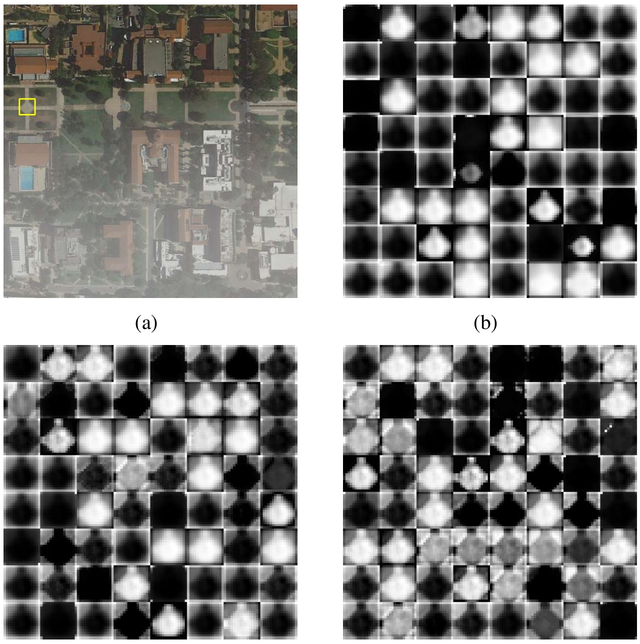
1) Effectiveness of PIE Module: To verify the effectiveness of the proposed PIE module, we replace it in the HFNet with a convolutional module. The convolutional module contains two 3×3 convolutional layers and one maxpooling layer. Similarly, an LReLU activation function is added after the convolutional layer. In total, five different combination schemes are generated, as shown in Table III, where P_i $(i = 1, 2, 3, 4)$ denotes the PIE module, C_i $(i = 1, 2, 3, 4)$ denotes the convolutional module, and $+$ denotes that the module is tandem. Table III gives the results of these five schemes based on patch correlation pooling, where the best scores are marked in bold.

From Table III, we infer the following conclusions. First, based on the results of HFNet schemes A to C, we find that network performance gradually improves with an increase in stacked PIE modules and is best when stacking 3 PIE modules. Second, observing the results of HFNet schemes C and D, we speculate that the complex PIE modules are not suitable for extracting low-level features of the images, which leads to degradation of the network performance. Finally, stacking PIE modules are more effective than spacing PIE modules based on the results of scheme C and scheme E. Combining the above information, we adopt HFNet scheme C as the final structure of our proposed HFNet.

2) Residual Concatenation: In the proposed PIE module, we introduce a residual concatenation to enhance the feature extraction capability and stabilize network training. To evaluate the effect of the residual concatenation, we compare BDQM performance under three cases: without an operation,

TABLE III
DIFFERENT DESIGN SCHEMES FOR HFNET

Model	Combination	SHRQ-Regular				SHRQ-Aerial			
		SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
HFNet Scheme A	$C_1 + C_2 + C_3 + P_4$	0.8478	0.6549	0.7998	0.1260	0.9662	0.8493	0.9611	0.0697
HFNet Scheme B	$C_1 + C_2 + P_3 + P_4$	0.8573	0.6706	0.8014	0.1504	0.9669	0.8582	0.9567	0.0685
HFNet Scheme C	$C_1 + P_2 + P_3 + P_4$	**0.8676**	**0.6917**	**0.8397**	**0.1182**	**0.9767**	**0.8848**	**0.9613**	**0.0682**
HFNet Scheme D	$P_1 + P_2 + P_3 + P_4$	0.8670	0.6737	0.8208	0.1221	0.9630	0.8475	0.9513	0.0743
HFNet Scheme E	$C_1 + P_2 + C_3 + P_4$	0.8637	0.6808	0.8500	0.1243	0.9704	0.8617	0.9577	0.0785



(a) (b) (c) (d)

Fig. 7. Examples of the proposed BDQM for learning patch features in three scenarios. (a) is the DHI from the SHRQ-Aerial [16] dataset. (b), (c), and (d) are the learned feature maps of the yellow rectangle in (a) of the models without an operation, with residual connection, and with residual concatenation, respectively.

with residual connection, and with residual concatenation. In our experiments, residual connection denotes the elementwise summation operation, while residual concatenation denotes the channelwise splicing operation. For a fair comparison, all the other experimental settings are the same for the three cases.

To intuitively compare model performance under the three cases, we train models on the SHRQ-Aerial [16] dataset, and the learned feature maps in the first PIE module are given in Fig. 7. In this figure, (a) is a DHI of the SHRQ-Aerial [16] dataset, (b), (c), and (d) are the corresponding learned feature maps of the yellow rectangle in (a) of the models without an operation, with residual connection, and with residual concatenation, respectively. The visualization results indicate that using residual concatenation can preserve more structural information. The feature maps learned by the model with residual concatenation are sharper than those learned by the model with residual connection.

3) Effectiveness of IRNet: To verify the effectiveness of our proposed IRNet, we compare two main regression network structures for the CNN-based IQA task, whose architectures are described in Section III-D1. We abbreviate the single

TABLE IV
RESULTS OF THREE REGRESSION NETWORKS

Dataset		SPNet	WPNet	IRNet
SHRQ-Regular	SRCC	0.8188	0.8190	**0.8676**
	KRCC	0.6424	0.6416	**0.6917**
	PLCC	0.8092	0.8093	**0.8397**
	RMSE	0.1878	0.1409	**0.1182**
SHRQ-Aerial	SRCC	0.9021	0.9083	**0.9767**
	KRCC	0.7589	0.7606	**0.8848**
	PLCC	0.8924	0.8791	**0.9613**
	RMSE	0.1514	0.1309	**0.0682**

pooling-based regression network as SPNet and the weighted pooling-based regression network as WPNet. We compare these two regression network structures with HFNet based on scheme C in IV-C1 and keep other settings unchanged. From the experimental results shown in Table IV, we obtain the following three observations.

First, the single pooling operation exhibits relatively poor performance on these two datasets when we adopt HFNet scheme C. The final score obtained by summing or averaging the patch quality scores ignores the contribution of superior local quality to the overall image. It also weakens the existence of inferior local quality. Second, using weights in single pooling can effectively alleviate the neglect of local quality. The weighted pooling operation emphasizes local properties and effectively stretches the quality distance between patches. According to Table IV, the results of four performance criteria of weighted pooling are significantly better than those of single pooling. Finally, compared to the above two methods, the results show significant improvements in patch-related pooling operations. The higher SRCC values on the SHRQ-Regular and SHRQ-Aerial datasets show conspicuous progress. The single pooling and weighted pooling methods compute individual patch scores only without considering the underlying relationship between patches in the entire image. The patch-related pooling operation strengthens the feature connections between each patch. As discussed in Section III-D1, the inexact MOS value of each patch can also cause performance degradation. The proposed scheme can avoid this awkward situation to produce better results.

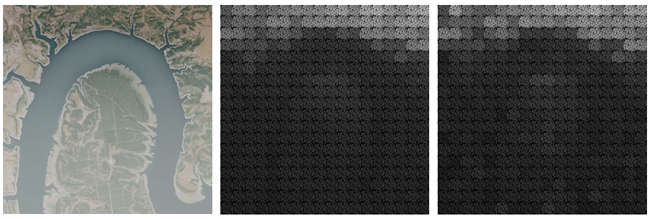
4) Patch Attention (PA): In the proposed IRNet, we introduce a PA mechanism to determine the different contributions of local quality to the overall quality prediction of the DHIs. To verify the effectiveness of the proposed PA mechanism, we compare the feature maps from HFNet before and after passing PA. Fig. 8 shows the results, where (a) is a DHI and (b) and (c) are the corresponding feature maps before PA and after PA, respectively. The difference between the river and the land is more significant in (c) than in (b), which is more

TABLE V
THE RESULTS OF DIFFERENT METHODS ON THE SHRQ-REGULAR AND SHRQ-AERIAL DATASETS

Metric			SHRQ-Regular				SHRQ-Aerial				exBeDDE			
			SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
GIQA	PSNR	FR	0.6308	0.4516	0.6942	0.1486	0.8350	0.6371	0.8151	0.1421	0.3918	0.2698	0.4020	0.2718
	SSIM [22]	FR	0.5627	0.3991	0.6201	0.1619	0.8207	0.6267	0.8166	0.1416	0.5208	0.3595	0.5222	0.2532
	DISTS [69]	FR	0.7246	0.5375	0.8047	0.1225	0.8409	0.6489	0.8265	0.1381	0.4929	0.3404	0.5000	0.2571
	VSI [70]	FR	0.6849	0.5065	0.7706	0.1315	0.7071	0.5158	0.6987	0.1755	0.5931	0.4164	0.6057	0.2362
	LPIPS [15]	FR	0.7221	0.5421	0.8130	0.1202	0.8351	0.6382	0.8187	0.1409	0.3845	0.2644	0.4142	0.2702
	PSIM [23]	FR	0.6238	0.4580	0.7580	0.1346	0.7593	0.5755	0.7338	0.1667	0.3993	0.2736	0.3654	0.2763
	IL-NIQE [32]	NR	0.3372	0.2347	0.5935	0.1661	0.4306	0.2955	0.4746	0.2159	0.0988	0.0673	0.1046	0.2952
	SNP-NIQE [33]	NR	0.4300	0.3036	0.5873	0.1671	0.4258	0.2944	0.4677	0.2168	0.2458	0.1683	0.2646	0.2863
	BPRIp [24]	NR	0.0159	0.0053	0.1638	0.2036	0.2388	0.1605	0.3568	0.2292	0.2647	0.1820	0.3203	0.2812
	BMPRI [25]	NR	0.5148	0.3666	0.6740	0.1650	0.5493	0.3901	0.5448	0.2240	0.6108	0.4360	0.6567	0.2256
DIQA	FRDQ [16]	FR	0.8292	0.6430	**0.8656**	**0.1033**	0.8615	0.6685	0.8554	0.1271	0.1461	0.0994	0.0474	0.2969
	FRDQa [16]	FR	0.7703	0.5908	0.8302	0.1151	0.9028	0.7219	0.9017	0.1061	0.2385	0.1633	0.1588	0.2931
	VI [18]	FR	0.6176	0.4537	0.6974	0.1479	0.7250	0.5335	0.7078	0.1733	0.4630	0.3199	0.4283	0.2682
	RI [18]	FR	0.5527	0.3953	0.6781	0.1517	0.6416	0.4515	0.6411	0.1883	0.4818	0.3342	0.4992	0.2572
	FRFSIM [19]	FR	0.5862	0.4264	0.6754	0.1522	0.8065	0.6245	0.7844	0.1522	0.1978	0.1344	0.3230	0.2809
	DPI [20]	RR	0.5519	0.3961	0.6254	0.1611	0.7075	0.5776	0.7295	0.1678	0.2774	0.1925	0.3350	0.2797
	PDIQA [21]	RR	0.4427	0.3061	0.4319	0.1862	0.6189	0.4508	0.5920	0.1977	0.4296	0.2976	0.4410	0.2664
	DHQI [17]	RR	0.4240	0.3000	0.6289	0.1605	0.5675	0.4341	0.5726	0.2011	0.6792	0.4701	0.6909	0.2146
	BQMD [27]	NR	0.5643	0.4161	0.7620	0.1337	0.8604	0.6745	0.8441	0.1315	0.6244	0.4455	0.6318	0.2301
	FADE [14]	NR	0.3425	0.2395	0.3358	0.1944	0.6632	0.5076	0.6759	0.1808	0.4991	0.3508	0.5168	0.2541
	HazDesNet [51]	NR	0.4045	0.2791	0.4196	0.1874	0.5472	0.3944	0.5547	0.2041	0.2132	0.1450	0.2253	0.2892
	BDQM	NR	**0.8676**	**0.6917**	0.8397	0.1182	**0.9767**	**0.8848**	**0.9613**	**0.0682**	**0.9156**	**0.7487**	**0.9369**	**0.1070**

TABLE VI
THE RESULTS OF DIFFERENT METHODS ON THE DHQ AND IVCDD DATASETS

Metric			DHQ				IVCDD			
			SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
GIQA	IL-NIQE [32]	NR	0.5970	0.4276	0.6548	0.1693	0.0952	0.1429	0.7950	0.1802
	SNP-NIQE [33]	NR	0.4740	0.3346	0.5447	0.1878	0.1905	0.1429	0.8636	**0.0962**
	BPRIp [24]	NR	0.2529	0.1767	0.3512	0.2097	0.0238	0.0714	0.7955	0.1800
	BMPRI [25]	NR	0.7265	0.5417	0.7496	0.1542	0.1413	0.0978	0.2468	0.2441
DIQA	DPI [20] [20]	RR	0.2710	0.1859	0.2727	0.2155	0.6429	0.5000	0.6072	0.2360
	PDIQA [21]	RR	0.4468	0.3002	0.4291	0.2023	0.8810	**0.7857**	**0.9101**	0.1231
	DHQI [17]	RR	0.8757	0.6994	0.8843	0.1046	0.5000	0.3571	0.6464	0.2267
	BQMD [27]	NR	0.6537	0.4844	0.7103	0.1640	0.4451	0.3276	0.5766	0.2124
	FADE [14]	NR	0.2526	0.1757	0.2540	0.2166	0.4762	0.3571	0.5903	0.2398
	HazDesNet [51]	NR	0.3303	0.2248	0.3430	0.2104	0.6429	0.5000	0.7928	0.1810
	BDQM	NR	**0.8936**	**0.7267**	**0.9008**	**0.0972**	**0.8955**	0.7404	0.9083	0.1077



(a) (b) (c)

Fig. 8. Examples of the proposed BDQM for learning patch features in two cases. (a) is the DHI from the SHRQ-Aerial [16] dataset. (b) shows the patch feature maps without the PA mechanism. (c) shows the patch feature maps after processing with the PA mechanism.

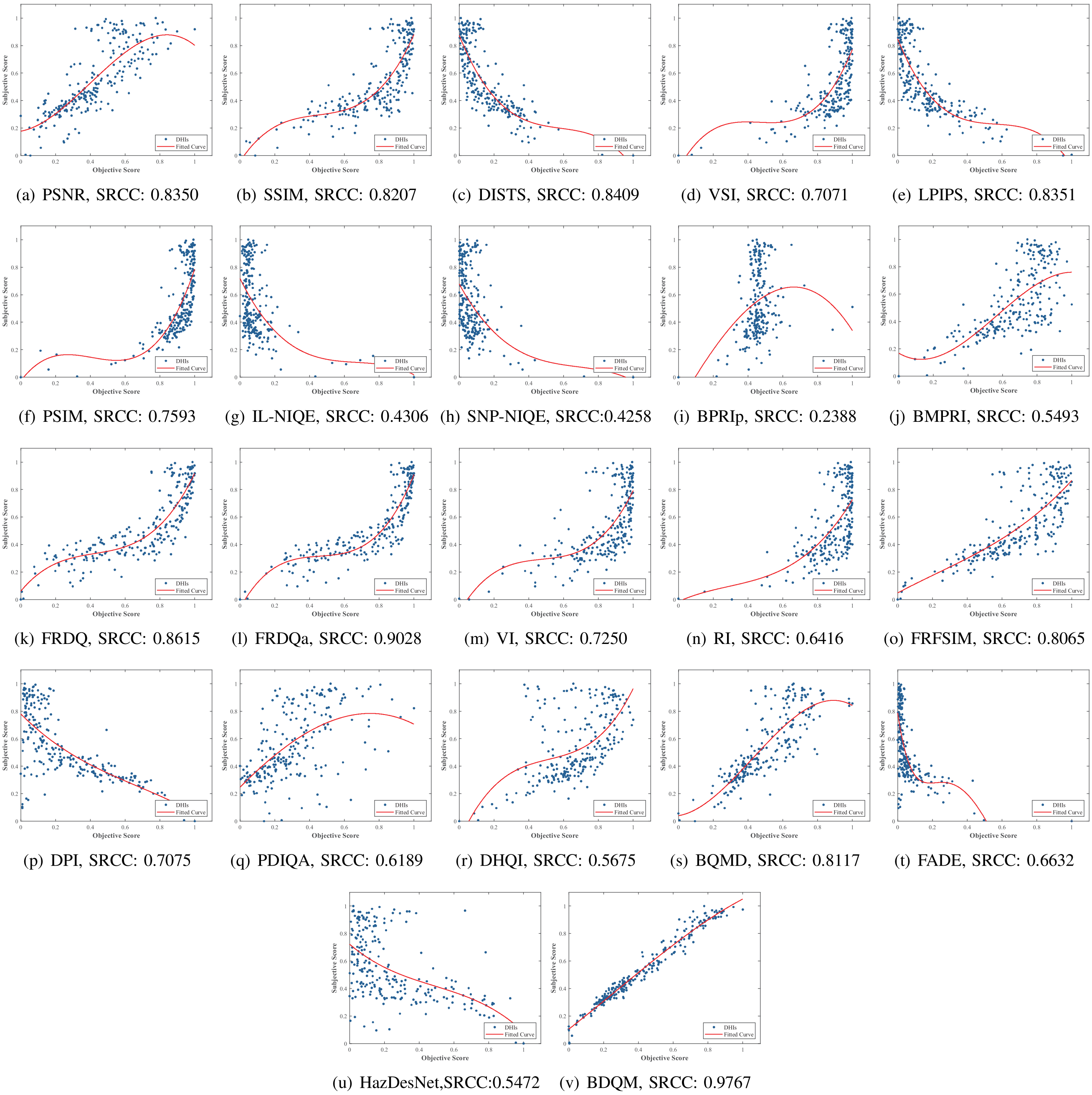
in line with the visual perception of the HVS in (a). Thus, applying the PA mechanism in IRNet has a significant impact on the accurate evaluation of the overall image.

D. Performance Evaluation

In this section, experiments are conducted to illustrate the advantages of the proposed BDQM for DIQA. Two types of IQAs are used for comparison, including 10 general-purpose IQA methods (GIQAs) and 11 DIQAs. The 10 GIQAs contain 6 FR-IQAs: PSNR, SSIM [22], DISTS [69], VSI [70], LPIPS [15] and PSIM [23], and 4 NR-IQAs: IL-NIQE [32], SNP-NIQE [33], BPRIp [24] and BMPRI [25]. The 11 DIQAs contain 5 FR-DIQAs: FRDQ [16], FRDQa [16], VI [18], RI [18] and FRFSIM [19], 3 RR-DIQAs: DPI [20], PDIQA [21] and DHQI [17], and 3 NR-DIQAs: BQMD [27], FADE [14] and HazDesNet [51].

1) Evaluation of Individual Dataset: We compare the proposed BQDM with 10 GIQAs and 11 DIQAs on the SHRQ-Regular, SHRQ-Aerial, exBeDDE, DHQ, and IVCDD datasets. Tables V and VI show the experimental results; the best results in each dataset are marked in bold. From the results, we can obtain the following observations. First, for these two types of IQAs, the FR and RR methods show better performance than the NR methods with the help of the reference information. Second, compared to the state-of-the-art IQAs (including GIQAs and DIQAs), our proposed BDQM achieves the best performance on SHRQ-Aerial, exBeDDE and DHQ and comparable performance on SHRQ-Regular and IVCDD. The results show that the proposed method can effectively evaluate the quality of DHIs without references.

In addition, we also employ scatter plots to graphically show the disparity between the proposed BDQM and 21 competitors. Fig. 9 shows the scatter plots of all examined methods on the SHRQ-Aerial dataset, where each symbol represents the



(a) PSNR, SRCC: 0.8350 (b) SSIM, SRCC: 0.8207 (c) DISTS, SRCC: 0.8409 (d) VSI, SRCC: 0.7071 (e) LPIPS, SRCC: 0.8351

(f) PSIM, SRCC: 0.7593 (g) IL-NIQE, SRCC: 0.4306 (h) SNP-NIQE, SRCC:0.4258 (i) BPRIp, SRCC: 0.2388 (j) BMPRI, SRCC: 0.5493

(k) FRDQ, SRCC: 0.8615 (l) FRDQa, SRCC: 0.9028 (m) VI, SRCC: 0.7250 (n) RI, SRCC: 0.6416 (o) FRFSIM, SRCC: 0.8065

(p) DPI, SRCC: 0.7075 (q) PDIQA, SRCC: 0.6189 (r) DHQI, SRCC: 0.5675 (s) BQMD, SRCC: 0.8117 (t) FADE, SRCC: 0.6632

(u) HazDesNet,SRCC:0.5472 (v) BDQM, SRCC: 0.9767

Fig. 9. Scatter plots of subjective scores against the objective scores predicted by representative IQA methods on the SHRQ-Aerial dataset. The solid red line and blue triangle symbols represent the fitted curve and the DHIs, respectively. The SRCC score for each metric is also shown.

subjective and objective scores for a single DHI. The solid red lines are the fitted curves of the scatter plots. The gap between the symbol and the fitted curves visually indicates the performance of the IQA model. The smaller the gap is, the better the performance is. Comparing the distribution characteristics of the scatter plots in Fig. 9, it is easy to see that the proposed BDQM has a great performance advantage.

2) Evaluation of different DHAs on DHIs: To further check the performance of the proposed BDQM, we conducted comparative experiments on two datasets with images applying specific DHAs. Eight DHAs are applied to SHRQ-Regular and SHRQ-Aerial datasets, including Fattal [1], Tarel [2], He [3], Xiao [4], Meng [5], Lai [6], Berman [7] and Cai [8]. For the sake of brevity, we present the results of SRCC and PLCC, and similar conclusions are drawn for KRCC and RMSE. The SRCC and PLCC results for the 10 GIQAs and 11 DIQAs based on different DHAs are presented in Tables VII and VIII.

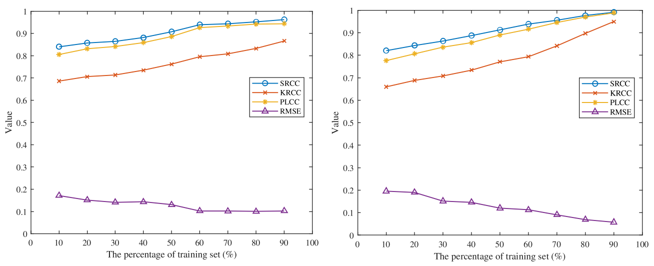
From the results presented in Tables VII and VIII, we can observe that the proposed BDQM has the highest SRCC and PLCC hit counts. In particular, the proposed BDQM shows promising performance for Fattal [1] and Cai [8] DHAs, while other available metrics show poor results for both DHAs. The reason for this result is that the images after applying DHAs Fattal [1] and Cai [8] have uneven haze residue.

TABLE VII
PERFORMANCE COMPARISON (SRCC) OF DIFFERENT DHAS

Metric	SHRQ-Regular								SHRQ-Aerial							
	Fattal	Tarel	He	Xiao	Meng	Lai	Berman	Cai	Fattal	Tarel	He	Xiao	Meng	Lai	Berman	Cai
PSNR	0.3775	0.6203	0.2347	0.4688	0.6074	0.2551	0.5748	0.5484	0.7428	0.8861	0.6592	0.2178	0.8839	0.4656	0.8158	0.6561
SSIM [22]	0.1485	0.5036	0.1748	0.5809	0.5611	0.2165	0.5437	0.4129	0.7357	0.8020	0.7913	0.5604	0.7264	0.3976	0.7397	0.6400
DISTS [69]	0.3876	0.6964	0.7850	0.5535	0.6920	0.3897	0.3574	0.4067	0.5293	0.8834	0.8527	0.3290	0.8776	0.4091	0.8612	0.5769
VSI [70]	0.5928	0.6750	0.5588	0.6275	0.6382	0.3567	0.4486	0.5506	0.6828	0.8211	0.8772	0.4158	0.8487	0.3771	0.5502	0.0478
LPIPS [15]	0.5744	0.6838	0.6057	0.6569	0.6821	0.5054	0.2578	0.5638	0.6592	0.9017	0.8443	0.2899	0.8994	0.4385	0.7989	0.4251
PSIM [23]	0.4457	0.5610	0.6584	0.4806	0.5499	0.2872	0.4043	0.4473	0.6392	0.9008	0.9212	0.2187	0.8879	0.3286	0.7264	0.1115
IL-NIQE [32]	0.2220	0.2349	0.6816	0.0605	0.1099	0.0121	0.2572	0.0735	0.2347	0.4545	0.6894	0.0736	0.3895	0.1911	0.4180	0.1097
SNP-NIQE [33]	0.3563	0.3896	0.7331	0.2603	0.2569	0.2908	0.1179	0.1067	0.1662	0.2529	0.5475	0.1706	0.3264	0.1880	0.1212	0.0638
BPRIp [24]	0.1439	0.0153	0.3934	0.0982	0.0092	0.0215	0.1040	0.0480	0.2365	0.3041	0.7068	0.4594	0.4581	0.0002	0.4345	0.3669
BMPRI [25]	0.6866	0.4984	0.8198	0.5257	0.3457	0.6507	0.3486	0.4846	0.0905	0.7526	0.8981	0.2970	0.6383	0.3152	0.7192	0.5987
FRDQ [16]	0.6982	0.7502	0.6686	**0.8407**	0.7443	**0.7486**	0.6462	0.6559	0.7544	0.8843	0.8474	0.7188	0.9168	0.4638	0.8376	0.7566
FRDQa [16]	0.6148	0.7833	0.7148	0.7424	0.7108	0.4826	0.6183	0.7308	0.8131	0.9181	0.8011	0.7682	0.9199	0.7192	0.8518	0.8545
VI [18]	0.3337	0.6538	0.3984	0.4958	0.7021	0.3389	0.3758	0.2155	0.5689	0.8407	0.8429	0.4527	0.7815	0.2267	0.8256	0.2022
RI [18]	0.4880	0.4980	0.4806	0.4671	0.3933	0.2195	0.1874	0.4561	0.6409	0.7197	0.7882	0.2694	0.7526	0.4407	0.3455	0.4852
FRFSIM [19]	0.2249	0.5515	0.7590	0.4051	0.4095	0.3038	0.4199	0.4117	0.5800	0.8808	0.9066	0.5497	0.8465	0.4162	0.6712	0.5230
DPI [20]	0.5722	0.4551	0.4524	0.4935	0.4250	0.5788	0.2968	0.3980	0.2974	0.9253	0.3686	0.2400	**0.9582**	0.1960	0.8349	0.8051
PDIQA [21]	0.4349	0.2684	0.3004	0.2769	0.3999	0.4133	0.3424	0.2047	0.5537	0.2494	0.2307	0.1662	0.4474	0.1947	0.6053	0.0683
DHQI [17]	0.1146	0.3859	0.7804	0.1292	0.2111	0.3381	0.2668	0.1576	0.0220	0.8986	0.9284	0.1297	0.7419	0.4105	0.7931	0.7308
BQMD [27]	0.3688	0.2693	**0.8710**	0.1235	0.2725	0.5410	0.1542	0.2174	0.6792	0.9417	0.9399	0.1987	0.9199	0.5662	0.6392	0.8042
FADE [14]	0.5314	0.3870	0.0722	0.5625	0.3812	0.4627	0.1651	0.3847	0.1657	0.9043	0.4336	0.1613	0.9479	0.2792	0.7717	0.7686
HazDesNet [51]	0.3759	0.4325	0.3029	0.1810	0.3292	0.0289	0.0208	0.3109	0.0563	0.8283	0.6156	0.3513	0.8945	0.3918	0.6436	0.5155
BDQM	**0.7717**	**0.7937**	0.7585	0.8347	**0.7527**	0.7050	**0.6792**	**0.7618**	**0.9769**	**0.9640**	**0.9662**	**0.8977**	0.9546	**0.9493**	**0.8834**	**0.9702**

TABLE VIII
PERFORMANCE COMPARISON (PLCC) OF DIFFERENT DHAS

Metric	SHRQ-Regular								SHRQ-Aerial							
	Fattal	Tarel	He	Xiao	Meng	Lai	Berman	Cai	Fattal	Tarel	He	Xiao	Meng	Lai	Berman	Cai
PSNR	0.4590	0.7265	0.2470	0.5509	0.6276	0.3154	0.5769	0.5886	0.7302	0.8601	0.6785	0.6843	0.8920	0.5008	0.7381	0.6797
SSIM [22]	0.3283	0.5740	0.2501	0.5878	0.6256	0.2596	0.5517	0.4978	0.7210	0.7973	0.8492	0.7297	0.7886	0.5946	0.7459	0.6472
DISTS [69]	0.5502	0.8151	0.7743	0.6024	0.7283	0.4229	0.4712	0.6502	0.5568	0.9208	0.8091	0.4893	0.8832	0.6335	0.9033	0.7238
VSI [70]	0.5926	0.7583	0.5665	0.6072	0.6753	0.3992	0.4985	0.6535	0.7038	0.8529	0.9020	0.5173	0.8336	0.6042	0.6232	0.1186
LPIPS [15]	0.6690	0.7454	0.5877	0.7605	0.7325	0.6045	0.3788	0.6085	0.7013	0.8936	0.8612	0.4836	0.9133	0.6699	0.8821	0.4770
PSIM [23]	0.5461	0.6507	0.6993	0.5846	0.5950	0.3481	0.3969	0.4706	0.6633	0.8854	0.9315	0.2642	0.8944	0.6401	0.7160	0.0799
IL-NIQE [32]	0.2907	0.3967	0.6897	0.1053	0.2771	0.0069	0.3704	0.3606	0.5235	0.4770	0.8056	0.2368	0.5424	0.2478	0.4485	0.3065
SNP-NIQE [33]	0.4219	0.3744	0.7422	0.1865	0.3101	0.3683	0.1891	0.2556	0.4761	0.4821	0.7861	0.0807	0.2881	0.3394	0.0830	0.1604
BPRIp [24]	0.4656	0.2705	0.4160	0.3196	0.2770	0.3030	0.0544	0.1357	0.3372	0.5736	0.7578	0.4810	0.5276	0.4563	0.5500	0.4793
BMPRI [25]	0.7174	0.5621	0.8676	0.5604	0.3296	0.6975	0.3240	0.5846	0.3081	0.7581	0.8798	0.7969	0.7348	0.4886	0.7625	0.6790
FRDQ [16]	0.7368	0.7904	0.7164	0.8472	0.7870	**0.7778**	0.6815	0.6756	0.7581	0.9241	0.8595	0.7919	0.9078	0.6988	0.7940	0.7855
FRDQa [16]	0.7160	**0.8391**	0.7295	0.7801	0.7553	0.5602	0.6054	0.7457	0.8113	0.9513	0.8302	0.8194	0.9167	0.9340	0.8001	0.9128
VI [18]	0.3789	0.7500	0.4446	0.4927	0.7350	0.4460	0.4043	0.4053	0.5649	0.8844	0.8777	0.5971	0.7486	0.4244	0.9189	0.3712
RI [18]	0.5742	0.7243	0.5380	0.5386	0.4047	0.3595	0.2895	0.4790	0.6806	0.7794	0.8136	0.4205	0.7903	0.6913	0.1899	0.5759
FRFSIM [19]	0.4303	0.5725	0.7689	0.5150	0.4766	0.5347	0.4543	0.4781	0.6297	0.8550	0.9172	0.6040	0.8482	0.6089	0.6619	0.5619
DPI [20]	0.6287	0.5024	0.4181	0.5350	0.5333	0.6105	0.3662	0.6008	0.3632	0.9167	0.4452	0.6692	0.9555	0.3388	**0.9214**	0.8176
PDIQA [21]	0.4823	0.2842	0.3600	0.5750	0.4501	0.5331	0.5101	0.4137	0.6640	0.3432	0.3684	0.2082	0.5688	0.2843	0.5637	0.3111
DHQI [17]	0.1192	0.5331	0.8126	0.3603	0.3682	0.3924	0.3527	0.3964	0.2321	0.8925	0.9429	0.3611	0.7402	0.5512	0.8883	0.7669
BQMD [27]	0.5332	0.4288	**0.8991**	0.3903	0.4370	0.5689	0.1367	0.3411	0.6886	0.9438	0.9401	0.3939	0.9205	0.7385	0.6908	0.8007
FADE [14]	0.6722	0.5498	0.3939	0.6059	0.3358	0.4329	0.3437	0.6015	0.2479	0.9007	0.6602	0.3010	0.9337	0.3800	0.7910	0.7964
HazDesNet [51]	0.4542	0.5763	0.4664	0.3637	0.3698	0.1399	0.0714	0.3641	0.5120	0.8736	0.7439	0.4255	0.8843	0.4585	0.7851	0.3797
BDQM	**0.8182**	0.7890	0.7803	**0.8713**	**0.7935**	0.7370	**0.8227**	**0.8120**	**0.9755**	**0.9780**	**0.9735**	**0.9388**	**0.9632**	**0.9849**	0.9091	**0.9739**



(a) (b)

Fig. 10. Performance results on the SHRQ-Aerial dataset using training sets of different sizes. The training sets are selected by the two strategies: (a) randomly selecting samples; (b) selecting samples based on expected error reduction.

Existing metrics fail to obtain accurate results because they do not capture the inhomogeneous haze features. In contrast, our proposed BDQM can learn visual features without any constraints to produce satisfactory results. However, for He [3], the performance of BDQM is slightly worse. Because He [3] generates additional artifacts during the dehazing process, our model learns these artifacts as perceptual features, which leads to inaccurate results. In general, BDQM exhibits stable and superior performance on most types of DHAs compared to the state-of-the-art IQA metrics

3) Studies on Different Size Training Sets: In this task, we investigate the impact of different-size training sets. Inspired by active learning [71], the training sets are selected by the following two strategies: (a) randomly selecting samples; (b) selecting samples based on expected error reduction. $10\%, 20\%, \ldots, 90\%$ subsets from a given dataset are selected as the training sets, and our proposed BDQM is trained independently on these training sets of different sizes, without introducing redundant variables. Then the model is tested on

TABLE IX
STATISTICAL SIGNIFICANCE TEST RESULTS FOR THE SHRQ-AERIAL DATASET. SYMBOL 1/0/- AT (I, J) INDICATES THAT THE MODEL IN ROW I IS STATISTICALLY SUPERIOR/INFERIOR/INCOMPARABLE TO THE MODEL IN COLUMN J.

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
PSNR	A	-	1	-	0	-	0	-	-	0	0	1	1	0	0	1	-	0	0	1	-	-	0
SSIM [22]	B	0	-	-	0	-	0	-	-	0	1	1	0	0	0	1	-	0	0	0	-	-	0
DISTS [69]	C	-	-	-	0	1	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0
VSI [70]	D	1	1	1	-	-	1	0	0	0	1	0	0	1	1	0	-	1	1	0	0	-	0
LPIPS [15]	E	-	-	0	-	-	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	1	0
PSIM [23]	F	1	1	1	0	1	-	0	0	0	1	0	0	1	1	0	-	1	1	0	0	-	0
IL-NIQE [32]	G	-	-	1	1	1	1	-	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
SNP-NIQE [33]	H	-	-	1	1	1	1	0	-	0	0	0	0	0	0	0	0	0	0	0	1	0	0
BPRIp [24]	I	1	1	1	1	1	1	1	1	-	0	0	0	0	0	0	-	0	0	0	0	-	0
BMPRI [25]	J	1	0	-	0	-	0	1	1	1	-	0	0	1	1	0	-	0	0	0	0	-	0
FRDQ [16]	K	0	0	1	1	1	1	1	1	1	1	-	0	0	0	1	-	0	0	0	-	-	0
FRDQa [16]	L	0	1	1	1	1	1	1	1	1	1	1	-	0	0	1	-	0	0	1	-	-	0
VI [18]	M	1	1	1	0	1	0	1	1	1	0	1	1	-	1	0	-	0	0	0	0	-	0
RI [18]	N	1	1	1	0	1	0	1	1	1	0	1	1	0	-	0	-	1	1	0	0	-	0
FRFSIM [19]	O	0	0	1	1	1	1	1	1	1	1	0	0	1	1	-	-	0	0	0	-	-	0
DPI [20]	P	-	-	1	-	1	-	1	1	-	-	-	-	-	-	-	-	0	0	0	0	1	0
PDIQA [21]	Q	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	-	1	0	0	-	0
DHQI [17]	R	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1	0	-	0	0	-	0
BQMD [27]	S	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	-	-	-	0
FADE [14]	T	-	-	1	1	1	1	0	0	1	1	-	-	1	1	-	1	1	1	-	-	0	0
HazDesNet [51]	U	-	-	1	-	0	-	1	1	-	-	-	-	-	-	-	0	-	-	-	1	-	0
BDQM	V	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-

the remaining subset of each random division. Theoretically, if more data are used for training, the model is fitted better and results in better performance.

Fig. 10 shows the performance results on the SHRQ-Aerial dataset under two data selection strategies. We repeat each random division of different sizes 10 times; the average result is plotted in Fig. 10. Obviously, the proposed model can still obtain satisfactory SRCC values even if only 10% of the data are used for training. A similar result can be found for the PLCC curve. The experimental results show that the proposed model can achieve satisfactory results even with a small amount of training data.

4) Statistical Significance Test: We perform a statistical significance test on the SHRQ-Aerial dataset to verify that the performance of all models is significantly different. The F-test method [16] is used by comparing residual variances between the subjective and the objective score. Table IX shows the result, where the symbols 1, -1, and - at (i, j) indicate that the model in row i performs statistically superior, inferior, and incomparable to the model in column j, respectively. From Table IX, we can make two observations. First, most competitors are statistically incomparable with each other. Second, the proposed BDQM has significant advantages over all considered state-of-the-art IQAs.

5) Computational Complexity: The computational complexity of IQAs is worth analyzing since running time is a critical influencing factor in many real-time applications. In our experiments, the computational complexity of a model is measured by computing the running time of predicting the quality of a single image. To eliminate the bias caused by specific image selections, we choose 240 images from the SHRQ-Aerial dataset for testing and utilize the average running time of each model as the computational time cost. Table X lists the computational complexity of the proposed model and the compared IQA methods.

From this table, we can draw the following conclusions. On the one hand, PSNR and SSIM [22] have the lowest computational complexity because they contain only simple functions among the seven GIQA metrics. DISTS [69] takes the longest time because its features are obtained through deep learning models. In addition, using the MVG models imposes a cost on the computation time of IL-NIQE [32] and SNP-NIQE [33]. On the other hand, for the remaining nine DIQAs that predict image quality by manually extracting features, PDIQA [21] has the highest computational complexity because it extracts features at the pixel level. In contrast, the proposed model shows moderate operating speed among the compared IQAs. This result suggests that our model does not incur an unacceptably heavy computational overhead to improve prediction accuracy.

TABLE X
COMPUTATIONAL COST OF BDQM AND OTHER IQA METRICS

GIQA	Time(s)	DIQA	Time(s)
PSNR	0.0104	FRDQ [16]	0.0631
SSIM [22]	0.0163	FRDQa [16]	0.0574
DISTS [69]	3.2901	VI [18]	0.1807
VSI [70]	0.1006	RI [18]	0.1752
LPIPS [15]	0.0162	FRFSIM [19]	0.1295
PSIM [23]	0.0562	DPI [20]	0.7796
IL-NIQE [32]	2.8047	PDIQA [21]	2.8798
SNP-NIQE [33]	1.9111	DHQI [17]	0.2844
BPRIp [24]	0.2590	BQMD [27]	0.3604
BMPRI [25]	0.3406	FADE [14]	0.5283
-	-	HazDesNet [51]	0.8249
-	-	**BDQM**	**0.0226**

V. CONCLUSION

In this paper, we propose BDQM, a deep CNN-based model for blind dehazed image quality assessment. To overcome the difficulty of acquiring haze-free and hazy image pairs, our proposed BDQM is an end-to-end model that does not require a reference image. BDQM contains three core components: image preprocessing, HFNet and IRNet. The HFNet consists of low-level feature extraction, PIE modules and a feature pooling step, which is beneficial for learning powerful feature representations and enhancing the feature extraction capability of the network. To consider the characteristics of the uneven quality distribution of DHIs and to avoid inaccurate estimation of the subjective scores of image patches during training, we design a PA mechanism in IRNet. Extensive experiments are conducted on five dehazed image datasets to verify the performance of our proposed BDQM and state-of-the-art IQA methods. The results show that BDQM outperforms the compared methods in terms of accuracy and monotonicity with moderate computational cost.

REFERENCES

[1] R. Fattal, "Single image dehazing," *ACM transactions on graphics*, vol. 27, no. 3, pp. 1–9, 2008.

[2] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proceedings of the IEEE international conference on computer vision*, 2009, pp. 2201–2208.

[3] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[4] C. Xiao and J. Gan, "Fast image dehazing using guided joint bilateral filter," *The Visual Computer*, vol. 28, no. 6, pp. 713–721, 2012.

[5] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 617–624.

[6] Y.-H. Lai, Y.-L. Chen, C.-J. Chiou, and C.-T. Hsu, "Single-image dehazing via optimal transmission map under scene priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 1–14, 2014.

[7] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1674–1682.

[8] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[9] J. Park, D. K. Han, and H. Ko, "Fusion of heterogeneous adversarial networks for single image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 4721–4732, 2020.

[10] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M. Yang, "Semi-supervised image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 2766–2779, 2020.

[11] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, no. 11, pp. 1–52, 2020.

[12] X. Min, K. Gu, G. Zhai, X. Yang, W. Zhang, P. Le Callet, and C. W. Chen, "Screen content quality assessment: overview, benchmark, and beyond," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1–36, 2021.

[13] G. Zhai, X. Min, and N. Liu, "Free-energy principle inspired visual quality assessment: An overview," *Digital Signal Processing*, vol. 91, pp. 11–20, 2019.

[14] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, 2015.

[15] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.

[16] X. Min, G. Zhai, K. Gu, Y. Zhu, J. Zhou, G. Guo, X. Yang, X. Guan, and W. Zhang, "Quality evaluation of image dehazing methods using synthetic hazy images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2319–2333, 2019.

[17] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2879–2892, 2019.

[18] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing Evaluation: Real-world benchmark datasets, criteria, and baselines," *IEEE Transactions on Image Processing*, vol. 29, pp. 6947–6962, 2020.

[19] W. Liu, F. Zhou, T. Lu, J. Duan, and G. Qiu, "Image Defogging Quality Assessment: Real-world database and method," *IEEE Transactions on Image Processing*, vol. 30, pp. 176–190, 2021.

[20] C. Hsieh, S. Horng, Z. Huang, and Q. Zhao, "Objective haze removal assessment based on two-objective optimization," in *Proceedings of the IEEE International Conference on Awareness Science and Technology*, 2017, pp. 279–283.

[21] C. Wang, H. Fan, H. Zhang, and Z. Li, "Pixel-level dehazed image quality assessment based on dark channel prior and depth," in *Proceedings of the IEEE International Conference on Parallel and Distributed Processing with Applications*, 2019, pp. 1545–1549.

[22] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[23] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 3903–3912, 2017.

[24] 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, 2017.

[25] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 508–517, 2018.

[26] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proceedings of the IEEE International Conference on Image Processing*, 2016, pp. 2226–2230.

[27] W. Shen, S. Hao, J. Qian, and L. Li, "Blind quality assessment of dehazed images by analyzing information, contrast, and luminance," *Journal of Network Intelligence*, vol. 2, no. 1, pp. 139–146, 2017.

[28] Y. Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1482–1495, 2017.

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

[30] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2200–2211, 2019.

[31] F. Li, Y. Zhang, and P. C. Cosman, "MMMNet: An end-to-end multi-task deep convolution neural network with multi-scale and multi-hierarchy fusion for blind image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4798–4811, 2021.

[32] 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.

[33] Y. Liu, K. Gu, Y. Zhang, X. Li, G. Zhai, D. Zhao, and W. Gao, "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.

[34] J. Tarel, N. Hautière, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2010, pp. 478–485.

[35] J. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, no. 2, pp. 6–20, 2012.

[36] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 973–992, 2018.

[37] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2019.

[38] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 754–762.

[39] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, 2018, pp. 620–631.

[40] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proceedings of the IEEE International Conference on Image Processing*, 2015, pp. 3600–3604.

[41] 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.

[42] M. C. Morrone, J. Ross, D. C. Burr, and R. Owens, "Mach bands are phase dependent," *Nature*, vol. 324, no. 6094, pp. 250–253, 1986.

[43] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and H. Geerts, "Color invariance," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 23, no. 12, pp. 1338–1338, 2001.

[44] Y. Song, H. Luo, R. Lu, and J. Ma, "Dehazed image quality assessment by haze-line theory," *Journal of Physics: Conference Series*, vol. 844, pp. 12–45, 2017.

[45] S. Fang, J. Yang, J. Zhan, H. Yuan, and R. Rao, "Image quality assessment on image haze removal," in *Proceedings of the Chinese Control and Decision Conference*, 2011, pp. 610–614.

[46] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication Technology*, vol. 15, no. 1, pp. 52–60, 1967.

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

[48] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018.

[49] J. Yang, T. Liu, B. Jiang, W. Lu, and Q. Meng, "Panoramic video quality assessment based on non-local spherical cnn," *IEEE Transactions on Multimedia*, vol. 23, pp. 797–809, 2020.

[50] L.-M. Po, M. Liu, W. Y. Yuen, Y. Li, X. Xu, C. Zhou, P. H. Wong, K. W. Lau, and H.-T. Luk, "A novel patch variance biased convolutional neural network for no-reference image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1223–1229, 2019.

[51] J. Zhang, X. Min, Y. Zhu, G. Zhai, J. Zhou, X. Yang, and W. Zhang, "Hazdesnet: An end-to-end network for haze density prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3087–3102, 2022.

[52] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 990–998.

[53] Y. Yang, T. Xiang, H. Liu, and X. Liao, "Convolutional neural network for visual security evaluation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3293–3307, 2020.

[54] J. Yang, S. Xiao, A. Li, W. Lu, X. Gao, and Y. Li, "Msta-net: forgery detection by generating manipulation trace based on multi-scale self-texture attention," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[55] J. Yang, A. Li, S. Xiao, W. Lu, and X. Gao, "Mtd-net: learning to detect deepfakes images by multi-scale texture difference," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4234–4245, 2021.

[56] X. Ding, Y. Guo, G. Ding, and J. Han, "Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1911–1920.

[57] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[58] W. Zhou and Z. Chen, "Deep multi-scale features learning for distorted image quality assessment," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2021, pp. 1–5.

[59] L. Shen, X. Chen, Z. Pan, K. Fan, and J. Lei, "No-reference stereoscopic image quality assessment based on global and local content characteristics," *Neurocomputing*, vol. 424, pp. 132–142, 2021.

[60] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[61] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. 2014, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980

[62] Raanan and Fattal, "Dehazing using color-lines," *ACM Transactions on Graphics*, vol. 34, no. 1, pp. 1–14, 2014.

[63] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[64] R. I.-R. BT, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2002.

[65] G. S. Xia, J. Hu, F. Hu, B. Shi, and L. Zhang, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[66] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, 2015.

[67] V. Q. E. Group. Final report from the video quality experts group on the validation of objective models of video quality assessment. [Online]. Available: http://www.vqeg.org/

[68] 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.

[69] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image Quality Assessment: Unifying structure and texture similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

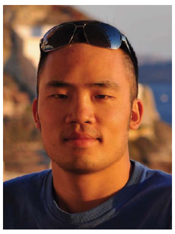
[70] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.

[71] Y. Li, J. Yang, and J. Wen, "Entropy-based redundancy analysis and information screening," *Digital Communications and Networks*, 2021.

Xiao Lv received the B.Sc. degree in computer science from Chongqing University, China, in 2020, where she is currently pursuing the Ph.D. degree with the College of Computer Science. Her current research interests include image quality assessment and multimedia security.

Tao Xiang (Senior Member, IEEE) received the BEng, MS and PhD degrees in computer science from Chongqing University, China, in 2003, 2005, and 2008, respectively. He is currently a Professor of the College of Computer Science at Chongqing University. Dr. Xiang's research interests include multimedia security, cloud security, data privacy and cryptography. He has published over 150 papers on international journals and conferences. He also served as a referee for numerous international journals and conferences.

Ying Yang received the B.Sc. and Ph.D. degrees in computer science from Chongqing University, China, in 2016 and 2021, respectively. She is currently a Post-Doctoral Research Fellow with the Department of Mathematics, the Chinese University of Hong Kong, Hong Kong. Her main research interests include image processing, quality evaluation, and multimedia security.



Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Associate Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is an Associate Editor for the IEEE Transactions on Human-Machine Systems and IEEE Transactions on Multimedia.