

Visibility and Distortion Measurement for No-Reference Dehazed Image Quality Assessment via Complex Contourlet Transform

Tuxin Guan, Chaofeng Li, *Senior Member, IEEE*, Ke Gu, *Member, IEEE*, Hantao Liu, *Member, IEEE*, Yuhui Zheng, *Member, IEEE*, and Xiao-jun Wu

Abstract—Recently, most dehazed image quality assessment (DQA) methods have focused on estimating remaining haze and omitting distortion impact from the side effect of dehazing algorithms, which leads to their limited performance. Addressing this problem, we propose a method for learning both visibility and distortion-aware features no-reference (NR) dehazed image quality assessment (VDA-DQA). Visibility-aware features are exploited to characterize clarity optimization after dehazing, including the brightness-, contrast-, and sharpness-aware features extracted by the complex contourlet transform (CCT). Then, distortion-aware features are employed to measure the distortion artifacts of images, including the normalized histogram of the local binary pattern (LBP) from the reconstructed dehazed image and the statistics of the CCT subbands corresponding to the chroma and saturation map. Finally, all the above features are mapped into quality scores by support vector regression (SVR). Extensive experimental results on six public DQA datasets verify the superiority of the proposed VDA-DQA method in terms of consistency with subjective visual perception and outperform state-of-the-art methods. The source code of VDA-DQA is available at <https://github.com/li181119/VDA-DQA>.

Index Terms—Dehazed image quality assessment, Visibility-aware features, Distortion-aware features, Complex contourlet transform, Support vector regression.

I. INTRODUCTION

Currently, images captured in outdoor scenes are degraded by bad weather, such as fog, haze, or mist. Degraded images usually suffer from low contrast, faint color, and shifted luminance. Under these circumstances, image dehazing methods become highly important for high-level computer vision tasks. Owing to continuous research, many efficient dehazing algorithms have been proposed to handle the degradation problem.

With the development of image dehazing, it is crucial to evaluate the performance of dehazing algorithms. Different from classical image distortion, classical image quality

assessment (IQA) methods are unsuitable for capturing haze removal. Many studies suggest that common IQA methods such as BRISQUE [1] and SSEQ [2] perform poorly in DQA tasks. It is important to develop an effective DQA method to promote dehazing algorithm progress. Generally, the DQA method includes two categories: subjective and objective evaluation methods. Subjective DQA methods based on human visual perception possess high prediction accuracy, but the disadvantage of being time and labor-consuming limits their use. Real-time objective DQA methods that automatically evaluate the dehazed image quality are urgently demanded.

Generally, objective DQA methods can be divided into two categories: full-reference (FR) and no-reference (NR) DQA methods. The difference between FR-DQA and NR-DQA is whether the source hazy image belongs to a synthetic hazy image. FR-DQA methods make full use of the haze-free images to compare with dehazed images and measure the difference between them. Min et al. [3] designed a synthetic haze database including regular and aerial image parts and then integrated the similarity measures of structure recovery, color rendition, and over-enhancement to obtain a quality score. Zhao et al. [4] proposed two novel evaluation criteria, the visibility index (VI) and the realness index (RI), to measure the dehazing effect. Liu et al. [5] utilized a fog-relevant feature-based similarity index (FRFSIM), which focuses on fog density and artifact distortion detection, to conduct quality assessment. Despite the good performance of FR-DQA methods, haze-free images do not exist for real hazy scenes, which hinders the usage of FR-DQA methods.

Without haze-free images, NR-DQA methods can directly predict the quality of dehazed images, as shown in Fig. 1. Choi, You and Bovik [6] proposed a referenceless model FADE for fog density prediction of hazy images, extracting the natural scene statistics features and a series of fog-aware statistical features of hazy images. Jiang et al. [7] extracted a series of fog-relevant features to construct a polynomial regression model for

This work was supported in part by the National Natural Science Foundation of China under Grant 62176150, and in part by the Shanghai Maritime University Graduate Top-notch Innovative Talent Cultivation Project under Grant 2021YBR020. (*Corresponding author: Chaofeng Li and Yuhui Zheng.*)

T. Guan and C. Li are with the Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China (e-mail: txguan6896@163.com; wxlichaofeng@126.com).

K. Gu is with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

H. Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF243AA, U.K (e-mail: liuh35@cardiff.ac.uk).

Yuhui Zheng is with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China and with the Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing 210044, China (e-mail: zheng_yuhui@nuist.edu.cn).

Xiao-Jun Wu is with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China (e-mail: xiaojun_wu_jnu@163.com).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>

the optical depth with a surrogate-based model. Then, the optical depth was used to derive the transmission map for fog density estimation and hazy image restoration. Pan et al. [8] defined a haze distribution map based on the dark channel prior and range channel prior and then deduced a haze density estimator of remote-sensing images HDMHA. Zhang et al. [9] designed an end-to-end deep-learning method, HazDesNet, to evaluate the density score of hazy images, converting the input hazy image into the pixel-level haze density map and taking the average of the refined density map as the haze density. Mahajan et al. [10] presented an end-to-end optimized CNN method DLHPQE for haze density prediction, which refined the pretrained DenseNet201 model in the constructed hazy image dataset for feature learning. The above methods conduct the DQA task based on the haze density estimation, which is not sufficient for the design of the DQA method.

Different from the above works, many researchers have focused on capturing the distortion from the side effect of image dehazing algorithms. In [11], Hautière et al. introduced three indicators, i.e., e , \bar{r} , and σ , for contrast restoration of dehazed images, where e reflects the edge restoration ability of the dehazing method, \bar{r} computes the quality of contrast restoration of a dehazing method, and σ calculates the number of saturated pixels in the dehazed image. Fang et al. [12] developed a dehazed image quality evaluation metric according to the combination of the contrast ascension measure and the structural similarity measure between hazy images and dehazed images. Min et al. [13] extracted haze-removing, structure-preserving, and over-enhancement features from the dehazed images to design an NR-DQA method DHQI. It is worth noting that the works in [12] and [13] calculate the similarity measure between hazy images and dehazed images for capturing uncertain structure distortion. Shen et al. [14] comprehensively analyzed the information loss, contrast, and luminance distortion of dehazed images, extracting a series of fog-aware features for quality prediction. Wu et al. [15] utilized the pretrained VGG-16 network with the transfer learning method to predict the quality of dehazed images and deblur images. Although these works show certain competitiveness, visibility improvement and content preservation measurements were not considered comprehensively.

For example, in Fig. 2, five dehazed images of the DHQ database [13] are ranked in descending order in terms of subjective quality from (a) to (e). To quantitatively evaluate the quality of these images, two quality evaluators, FADE and IL-NIQE [16], are chosen for exemplification. From the view of haze density prediction, the higher the FADE index is, the denser the fog in the image. From the perspective of the distortion measure, the higher the IL-NIQE is, the worse the distortion. It can be found that neither FADE nor IL-NIQE correlates with subjective perception very well, so it is necessary to measure both the haze removal effect and distortion artifacts for the DQA method. Furthermore, it is worth noting that the overall quality rating of dehazed images depends on the haze removal effect and artifact measurements during subjective experiments [3],[13].

Fig. 3 shows example images of common JP2K, Blur, and WN distortions in the IQA database [17] along with structural loss, over-enhancement, and color shift distortion in the DQA

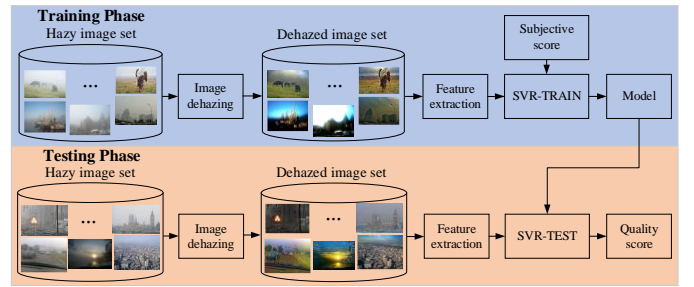


Fig. 1. General framework of no-reference dehazed image quality assessment.

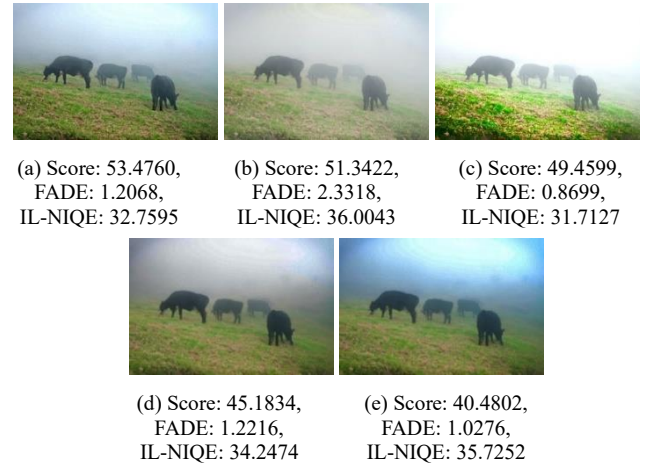


Fig. 2. Performances of dehazed images predicted by different quality estimators.

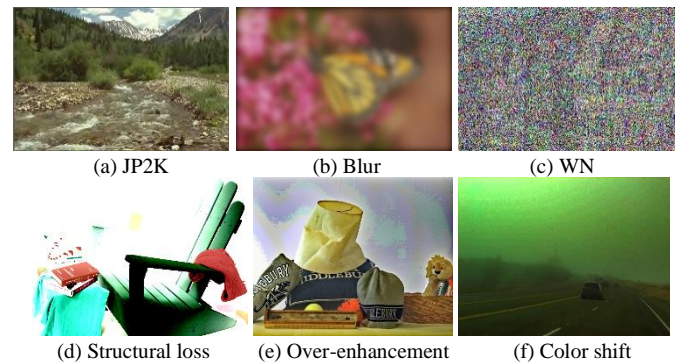


Fig. 3. Comparison between classical distortion types and distortion types in dehazed images.

database. The common distortion types are synthesized to imitate the disturbance occurring at the acquisition, transmission, and image compression steps. However, the distortions of dehazed images are induced by the side effects of dehazing algorithms. Compared with the common distortion types, the distribution of the distortion from dehazing algorithms tends to be heterogeneous, resulting in the inapplicability of IQA metrics for the DQA task.

Inspired by the above findings, we propose a novel NR-DQA method by joint measurement of the visibility and distortion of dehazed images. To imitate the spatial decomposition mechanism that occurs in the primary visual cortex [18] well, the complex contourlet transform is utilized to extract the visibility- and distortion-aware features. The contributions of this paper are as follows:

- 1) We propose a method for learning both visibility and

distortion-aware features DQA (VDA-DQA) for comprehensively considering two aspects of dehazing algorithms: the effect of haze removal and the introduction of distorted artifacts, where sharpness, brightness and contrast features are used for measuring haze removal, and the structure and color features are utilized for capturing the distortion artifacts.

- 2) To measure haze removal, a series of visibility-aware features are discovered: the log energy of high-frequency CCT subbands for sharpness features, the subband entropy of dehazed images and brightness-changed images for brightness features, and the significant differences between the dehazed image and contrast-changed image for contrast calculation.
- 3) To measure distorted artifacts, we use the high-frequency part of CCT subbands to generate a reconstructed image by inverse CCT to highlight the change in structure and take its LBP histogram statistics as structure-aware features. The standard deviation of the joint statistics of the CCT subbands, normalized by the GGSM-DNT, corresponding to the chroma and saturation map, is utilized as a color-aware feature.
- 4) Experimental results on several DQA databases show that the proposed VDA-DQA is highly linear against subjective visual perception and outperforms the state-of-the-art general-purpose IQA, contrast-enhancement IQA, and other DQA-oriented methods.

The rest of the paper is arranged as follows. Section 2 describes the proposed VDA-DQA method in detail. Section 3 presents the experimental results and discussion. Finally, the general conclusion of this article is given in Section 5.

II. COMPLEX CONTOURLET TRANSFORM

The contourlet transform (CT) [19],[20] performs better than the classical wavelet in many image processing tasks due to the characteristics of shift sensitivity and directionality. However, the lack of translation invariance leads to pseudo-Gibbs phenomena through the contourlet transform. Then, Chen et al. [21] proposed a complex contourlet transform to overcome the disadvantages of CT. The CCT utilizes the dual-tree complex wavelet transform (DT-CWT) [22] and direction filter bank (DFB) [23] for scale and orientation decomposition. Among them, DT-CWT is translation invariant, and DFB can strengthen the directional selectivity. Fig. 4 presents the decomposition procedure of CCT.

As shown in Fig. 4, trees A and B represent the real and imaginary parts of DT-CWT, and the DFB operation further improves the direction resolution of the complex wavelet transform. Examples of a complex contourlet decomposition of a hazy image and dehazed image with a 1-scale and 2-orientation bandpass transform are presented in Fig. 5. From Fig. 5, it can be found that the inherent characteristics of CCT are useful for capturing contour information of images that exhibit directional multiscale dependencies, which can be related to the change in image quality.

III. PROPOSED NO-REFERENCE VDA-DQA METHOD

The whole procedure of our proposed VDA-DQA method is

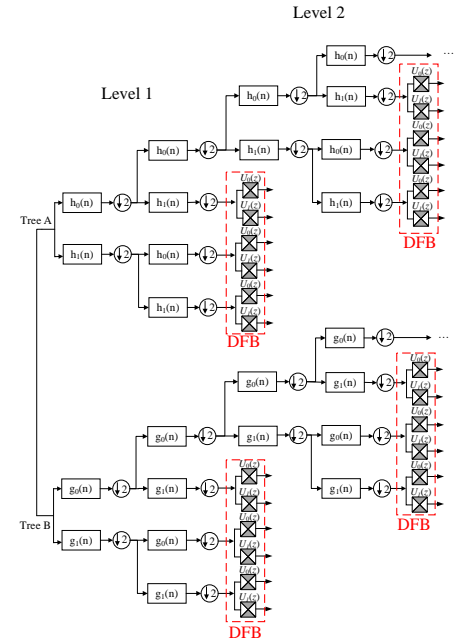


Fig. 4. The CCT decomposition procedure.

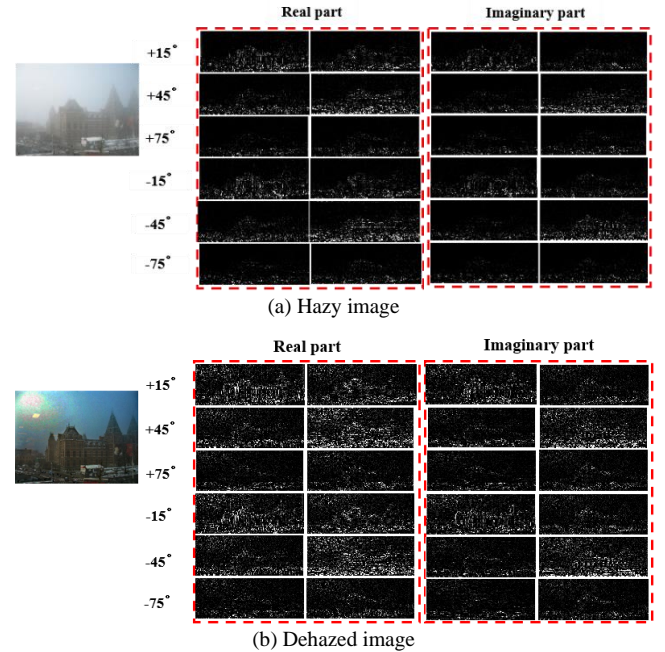


Fig. 5. Examples of complex contourlet transforms.

shown in Fig. 6. First, brightness and contrast-changed images are obtained by adjusting the brightness and contrast of the original dehazed image. Then, a novel reconstructed dehazed image that reflects the structure information variation is generated by reconstructing the high-frequency CCT subbands via the inverse CCT. For visibility-aware features, the brightness and contrast features are extracted from brightness and contrast-changed images in the complex contourlet domain, and the sharpness feature is calculated by using the high-frequency subband energy of the dehazed image. For distortion-aware features, the histogram statistics of LBP are extracted from the reconstructed image to reflect structure variation, and the across-scale and across-orientation statistics are extracted from the CCT subbands of the chroma and saturation map to reveal the quality degradation under

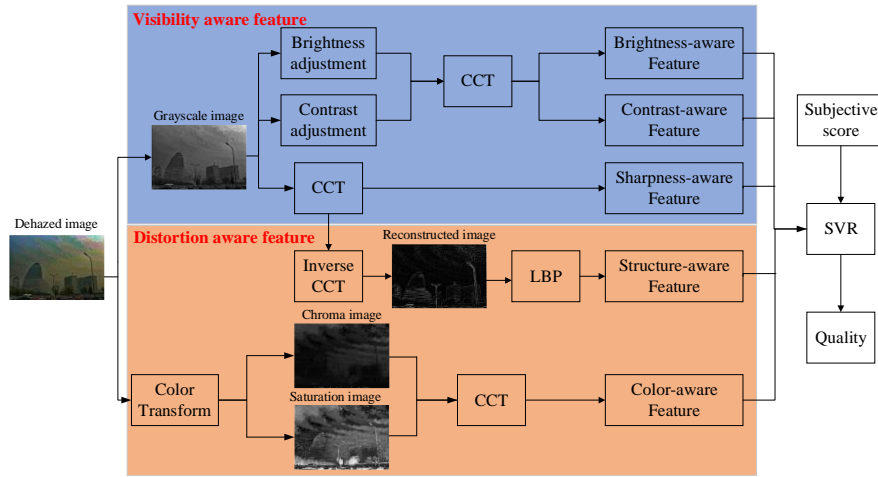


Fig. 6. Framework of the proposed VDA-DQA method.

color distortion. Finally, the SVR model is utilized to conduct regression mapping for the above perceptual features.

A. Visibility-aware Features

For the image dehazing task, the residual haze in the dehazed images has a significant effect on the visibility of the original image, so a high-quality dehazed image should have appropriate brightness and contrast for presenting more details.

1) Sharpness feature

As an important factor related to image quality, the sharpest region in the image is easily identified by the viewers. The degradation of sharpness characteristic reflecting the variation in local details can arise from the atmospheric turbulence or the side effect of image restoration methods [24]. For the dehazed images, the higher the sharpness of the dehazed image is, the more salient edges it preserves, and the higher the quality. In [25], the authors found that sharp images contain more high-frequency content than smooth images owing to the occurrence of the blurring effect in the high-frequency part of the image. Following that, we only compute the log energy of each high-frequency CCT subband at each decomposition level as the sharpness intensity of the dehazed image. The detailed equation is as follows (1):

$$\begin{cases} LE_{LH,l} = \log_{10}[1 + \frac{1}{K_l} \sum_i LH_l^2(i)] \\ LE_{HL,l} = \log_{10}[1 + \frac{1}{K_l} \sum_i HL_l^2(i)] \\ LE_{HH,l} = \log_{10}[1 + \frac{1}{K_l} \sum_i HH_l^2(i)] \end{cases}, \quad (1)$$

where i represents the pixel index of the image and K_l is the number of CCT subband coefficients at decomposition level l . Then, the total log energy of CCT subbands at each decomposition level is calculated by

$$LE_l = \frac{\alpha * LE_{HH,l} + 0.5 * (LE_{LH,l} + LE_{HL,l})}{1 + \alpha}, \quad (2)$$

where the parameter α is used to impose larger weights on the HH part of CCT subbands, which contain more distortion-sensitive high-frequency contents [26]. Here, we set α to 4, using the same method as ref [25]. Fig. 7 shows five different intensities of hazy images at the same scene on LIVE_Image_Defogging_Database [6], and the corresponding log-energy-based sharpness distribution

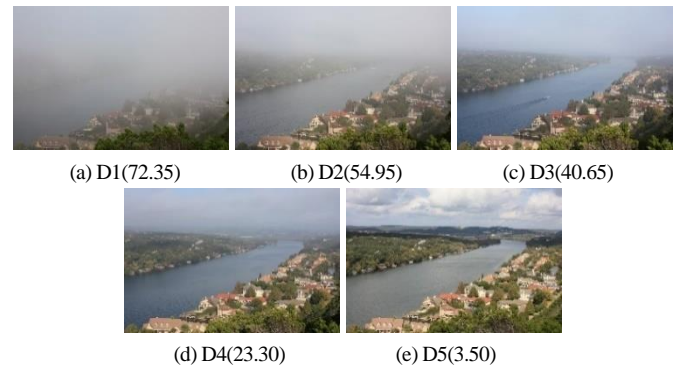


Fig. 7. Examples of hazy images in the LIVE_Image_Defogging_Database.

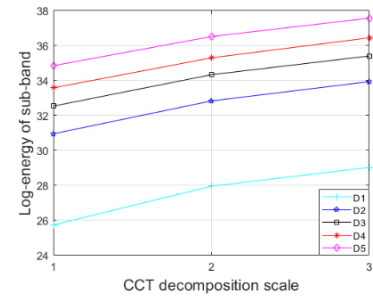


Fig. 8. Subband energy distribution of the examples in Fig. 7.

of these images is shown in Fig. 8. In Fig. 7, the numbers in the brackets indicate the fog density in the hazy image, which continues decreasing from D1 to D5. Additionally, it can be found that the sharpness value decreases with the fog becoming increasingly dense in Fig. 8 and increases with the decomposition scale becoming finer. Thus, the sharpness feature can be utilized as an effective visibility-aware feature.

2) Brightness feature

The brightness characteristics can render dehazed images a broad dynamic range [27]. Different from the local perception of sharpness, a good dehazed image with suitable brightness can help in understanding scene information. However, the existence of residual haze in dehazed images leads to the attenuation of brightness. Moreover, the improper operation of dehazing algorithms can lead to the underexposure and overexposure of dehazed images, so it is essential to measure the intensity of



Fig. 9. Examples of dehazed images in the IVCDehazing Database.

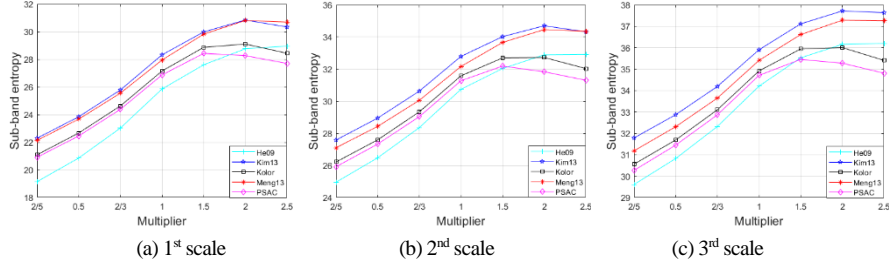


Fig. 10. Entropy distribution of the CCT subbands corresponding to different decomposition scales.

brightness corresponding to the dehazed image. Gu et al. [28] suggested that the tone-mapped image with suitable brightness comprises more valuable information than others, even if the brightness of the original image changes obviously. Referring to [28], we create a series of brightness-changed images by increasing or decreasing the brightness of the original dehazed images, generated by the following (3):

$$I_i = \min(\max(M_i \cdot I, 0), 255), \quad (3)$$

where I represents the grayscale image of the dehazed image and parameter M_i stands for the i -th multiplier. Here, we set $M = \{1, n, \frac{1}{n} | n=1.5, 2, 2.5\}$ by considering both effectiveness and computation complexity. In addition, the operators \min and \max are used to limit the grayscale level of brightness-changed images to the range of 0-255. To measure abundant detail information contained in images with suitable brightness, the entropy of CCT subband coefficients at each decomposition scale and orientation is calculated here. Compared with [28], the entropy of the subbands used here conforms to the multiscale perception method of humans, utilizing the different contributions of the entropy at each scale. Let $H_{s,o}$ denote the entropy of CCT subband coefficients at the s -th decomposition scale and o -th decomposition orientation, and the formula is given by (4)

$$H_{s,o} = -\sum p(I_{CCT}^{s,o}) \log p(I_{CCT}^{s,o}), \quad (4)$$

where $I_{CCT}^{s,o,k}$ denotes the CCT subband coefficients of the corresponding brightness-changed images in Equation (3), $k = 1, \dots, M_{s,o}$ denotes all the values of subband coefficients $I_{CCT}^{s,o,k}$, and $p(\cdot)$ is the probability density function corresponding to $I_{CCT}^{s,o}$. After computing the entropies of CCT subbands at all scales and orientations according to Equation (4), the entropies $H_{s,o}$ are pooled at each scale s , $s = 1, \dots, N_s$ along the decomposition orientations as follows:

$$H_s = \sum_{o=1}^{N_o} \log(1 + H_{s,o}), \quad (5)$$

where H_s denotes the pooled value of entropies at all orientations of the s -th scale and N_o represents the number of decompositions at the s -th scale.

Fig. 9 shows five dehazed images in the IVCDehazing database [29], and their corresponding entropy curve is shown in Fig. 10.

From Fig. 9 and Fig. 10, it can be found that the dehazed image with proper brightness presents more detailed information than other dehazed images, no matter how the luminance intensity is adjusted. Here, the brightness feature is chosen as the second kind of visibility-aware feature.

3) Contrast feature

The goal of image dehazing is to generate a clean image with high visibility and appropriate contrast. As an important attribute of the image, the contrast has a significant effect on the visual perception behaviors of the human brain. A high-quality dehazed image with appropriate contrast appears perceptually pleasing and visually informative, strengthening the ability to distinguish the object [30], so the contrast measure is vital for an image. In our previous work [31], a novel blur index was constructed based on the difference between the blurred image and its reblurred version. Inspired by this, we estimate the contrast of dehazed images based on the significant differences between the original dehazed image and its contrast-changed version.

To reduce the statistical dependencies between subbands, the generalized Gaussian scale mixture (GGSM)-based divisive normalization transform (DNT) [32] is used to normalize the CCT subband coefficients. As shown in [33], the nonlinear behavior of cortical neurons can be characterized by the DNT, modeling the interactions between neighboring neurons and pooling the response of surrounding neurons. The authors of [33] suggest that the generalized Gaussian scale mixture (GGSM)-based divisive normalization transform (GGSM-DNT) is more robust to distorted natural images than GSM-DNT, so we use GGSM-DNT to normalize the CCT subband coefficients to reduce the statistical dependencies between subbands. According to [32], an N -dimensional random vector Y can be expressed as a GGSM only if $Y \equiv z^{1/2} \cdot u$, where symbol \equiv represents equality in the probability distribution, u is a zero-mean multivariate generalized Gaussian (MVG) random vector, whose scale parameter is Σ_u and shape parameter is s , and the mixing multiplier z is a scalar random variable. Given the variance field z , the vector Y corresponding conditional density function is given by the following:

$$p(Y|z) = \frac{\Gamma(N/2) \cdot s \cdot z^{-N/2}}{\pi^{N/2} \Gamma(N/2s) 2^{N/2s} |\Sigma_u|^{1/2}} \cdot \exp\left\{-\frac{z^{-s}}{2} (Y^T \Sigma_u^{-1} Y)^s\right\}, \quad (6)$$

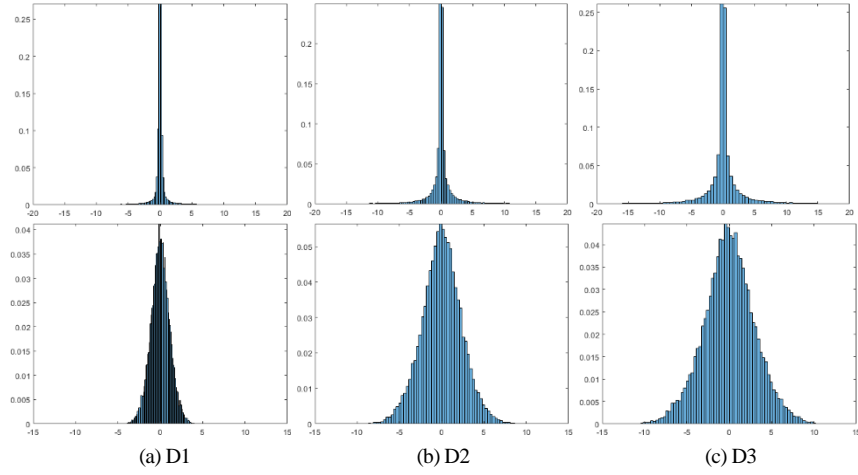


Fig. 11. Comparison between original and normalized subband statistics of hazy images in Fig. 7.

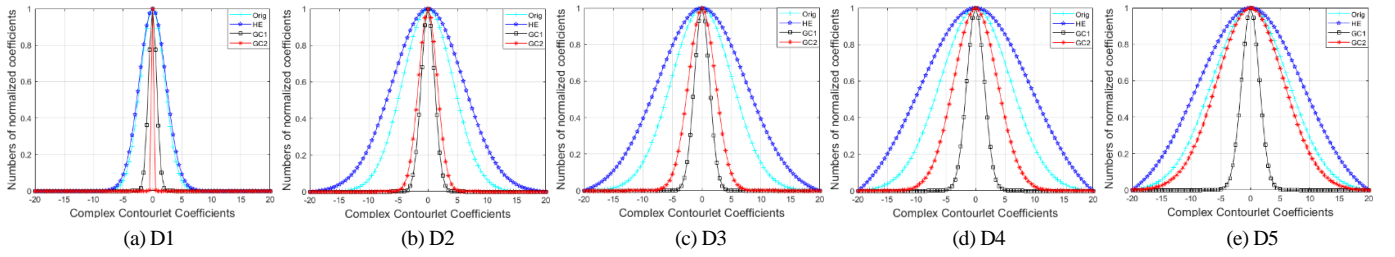


Fig. 12. Marginal distribution of CCT subbands corresponding to the examples in Fig. 7.

For a center coefficient y_c at each CCT subband y , its DNT neighborhood vector Y consists of the coefficients from the 3×3 neighborhood around y_c in the same subband, the parent subband, and the same spatial location in the subbands at the same scale but different orientations. Then, the normalization parameter \hat{z} , computed by the neighborhood vector Y , is exploited to process the coefficients at each subband to generate its corresponding divisively normalized subband $y/(\hat{z})^{1/2}$. The maximum likelihood estimation of factor \hat{z} is given by:

$$\begin{aligned} \hat{z} &= \arg \max_z (\log p(Y|z)) \\ &= \arg \max_z \left(-\frac{N}{2} \log(z) - \frac{1}{2} z^{-s} (Y^T \Sigma_u^{-1} Y)^s \right), \\ &= \arg \min_z \left(\frac{N}{2} \log(z) + \frac{1}{2} z^{-s} (Y^T \Sigma_u^{-1} Y)^s \right) \\ &= (s/N)^{1/s} (Y^T \Sigma_u^{-1} Y) \end{aligned} \quad (7)$$

The effect of GGSM-DNT on the statistics of the CCT subband coefficients is shown in Fig. 11. The statistics of the CCT subband coefficients before and after the GGSM-DNT operation are shown in the first and second rows in Fig. 11, respectively, which suggest that the GGSM-based DNT operation converts the approximate Laplacian nature of the original CCT subband coefficients into the Gaussian-like nature of normalized subbands. The marginal distribution of normalized CCT subbands corresponding to the hazy images in Fig. 7 and their contrast-changed images is shown in Fig. 12. Here, “HE” represents the histogram equalization method. Additionally, “GC1” and “GC2” represent the gamma correction method whose values of gamma are 0.1 and 10, respectively, which ensure that the contrast degrades greatly. It can be seen in Fig. 12 that the distance between the marginal distribution of original hazy images and their contrast-changed

images decreases as the haze becomes dense.

Similar to [34], if the contrast of the image is poor enough, the gap between itself and its contrast change image is also small. Based on the above findings, the difference in the standard deviation and kurtosis parameter of the CCT subbands corresponding to the original image and contrast-changed image is calculated as:

$$D_\sigma(\sigma_o, \sigma_c) = \sum_{i=1}^m |\sigma_{o,i} - \sigma_{c,i}|, \quad (8)$$

$$D_k(k_o, k_c) = \sum_{i=1}^m |k_{o,i} - k_{c,i}|, \quad (9)$$

where $\sigma_{o,i}$ and $\sigma_{c,i}$ represent the standard deviation of the marginal distribution of the i -th subband, and $k_{o,i}$ and $k_{c,i}$ represent the kurtosis parameter of the marginal distribution. In addition, the local entropy is used to estimate the information loss as follows (10):

$$D_e(E_o, E_c) = \sum_{i=1}^m |E_{o,i} - E_{c,i}|, \quad (10)$$

where $E_{o,i}$ and $E_{c,i}$ represent the local entropy of the i -th subband of the original image and contrast-change image, respectively. The formula of local entropy is as follows:

$$E = \frac{1}{N} \sum_{i=1}^N E(B_i), \quad (11)$$

where B_i represents the i -th block of size 5×5 in the CCT subband and N is the number of blocks in the image.

In this paper, the contrast features, including the difference between the standard deviation, kurtosis and local entropy of the subbands corresponding to original and contrast-changed images, are chosen as the third kind of visibility-aware features. Different from the work in [35], the difference between the three above types of parameters is calculated upon the pseudo-reference criterion for estimating the contrast of dehazed images in this paper.

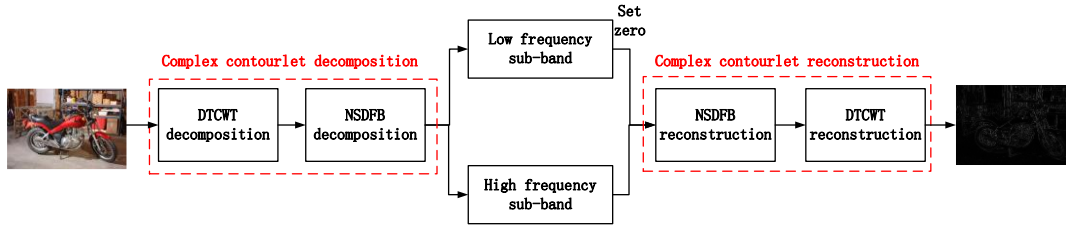


Fig. 13. The generated process of the reconstructed image.

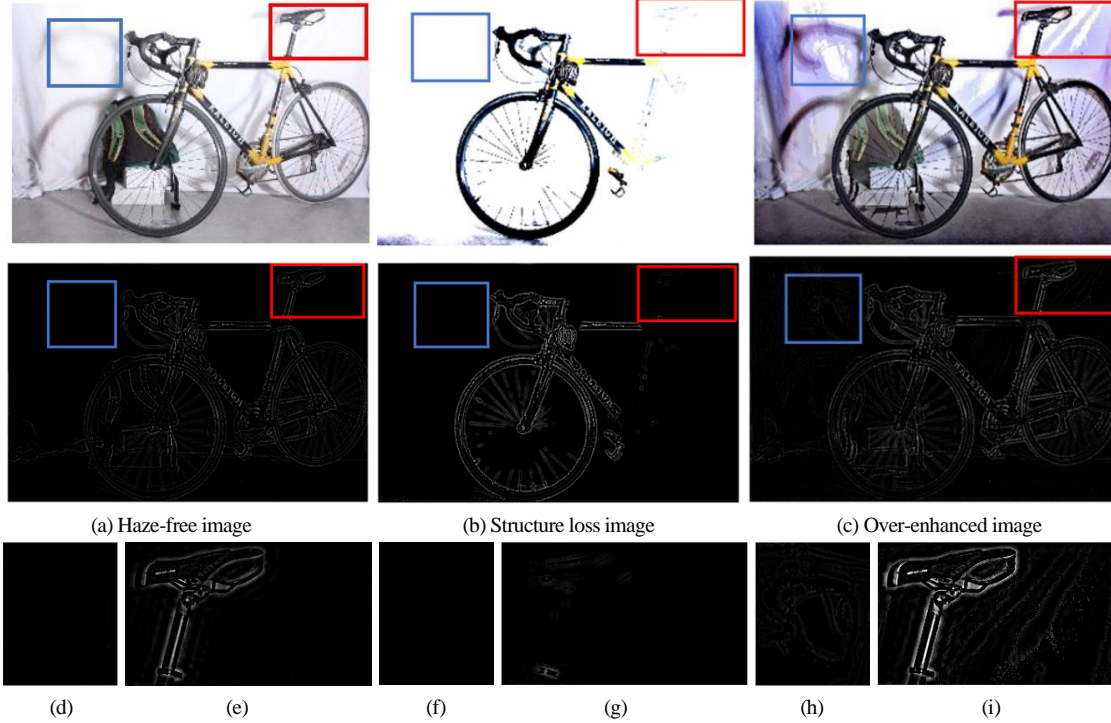


Fig. 14. Examples of dehazed image reconstruction.

B. Distortion-aware Features

Apart from the fine visibility, it assumes that a high-quality dehazed image should look natural as well. This means that the dehazing algorithms should not only restore the clean scene but also preserve the original detailed information without producing any artifacts.

1) Structure-aware feature

① Reconstructed image generation

As the basic component of the image, structural information can be adaptively extracted from the input stimulus by the human vision system, as shown in [36]. During the dehazing process, some dehazing algorithms can damage the genuine structure of the image, resulting in a poor dehazing effect. For example, some dehazing methods blindly pursue the complete removal of haze, ultimately losing the intrinsic structural information of the dehazed image. In contrast, another kind of structure distortion artifact is the over-enhancement phenomenon. Several dehazing methods incorrectly enhance the inapparent detail information in the low-contrast background area of hazy images. All of the above operations result in dehazed image quality degradation.

The characteristics of CCT can be used to capture image structures. The high-frequency component of the CCT subbands contains abundant contour information of the image, while the

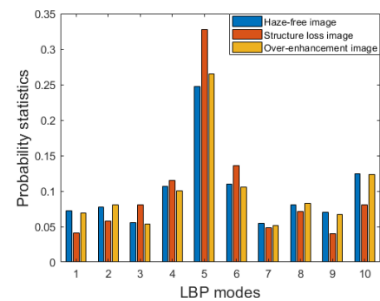


Fig. 15. The LBP histograms of different reconstructed images.

low-frequency component is rich in useless information. Here, the low-frequency CCT subbands are set to zero to preserve only the structure description of the dehazed image. Finally, the high-frequency CCT subbands are converted to the reconstructed image via the inverse CCT. The reconstruction procedure for the dehazed image is shown in Fig. 13. Moreover, examples of reconstructed images corresponding to the haze-free image, structure loss image, and over-enhanced image are shown in Fig. 14.

In Fig. 14, the original input images are shown in the first row, the corresponding reconstructed images are shown in the second row, and Fig. 14 (d), (f), (h) and (e), (g), (i) are the enlarged regions of the blue and red rectangles in the above three images. Comparing Fig. 14 (d), (f), and (h), it can be seen

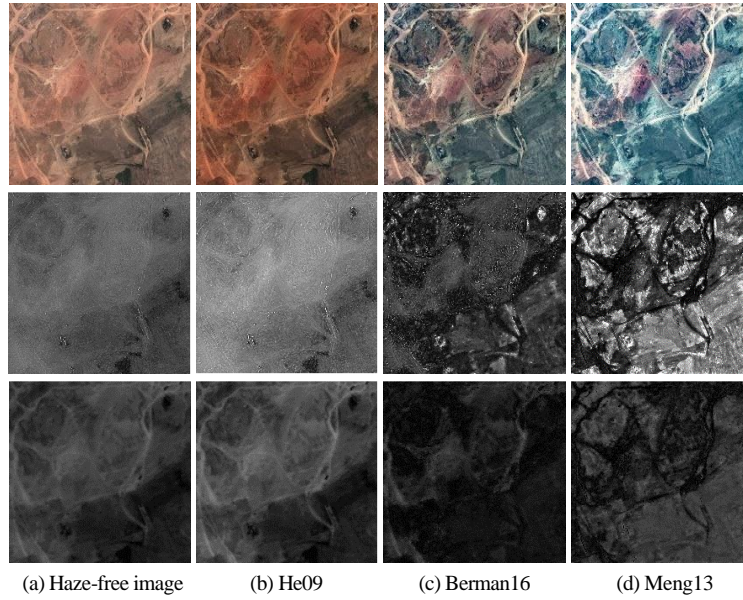


Fig. 16. Examples of haze-free image and corresponding dehazed images in the SHRQA database.

in Fig. 14(h) that the imperceptible details in the image are incorrectly enhanced because of the overoperation of the algorithm. In Fig. 14 (g), the structure information of the bicycle is completely lost. In Fig. 14 (i), it is apparent that the intrinsic structure information of the bicycle is wrongly highlighted. In summary, the reconstructed image can effectively characterize the structural information change.

② LBP statistics calculation

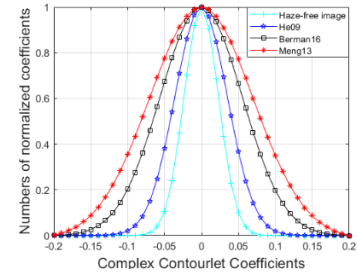
In classical IQA tasks [37],[38], the local binary pattern (LBP) is widely employed to capture local structural relationships in the image, according to the calculation of the gray-level differences between the center pixel and surrounding pixels of a local area. However, the output dimensionality of the original LBP histograms is up to 256, which greatly increases the computational complexity. To achieve rotation-invariance and reduce the dimension, the rotation-invariant and uniform LBP [39] is used to describe the primitive microstructures of the reconstructed image. Compared with the classical version, uniform LBP contains no more than two spatial bitwise transitions. Moreover, the rotation-invariant and uniform variant with a lower feature dimension can further increase the discriminative capability of LBP. The rotation-invariant uniform LBP at one location can be formulated as (12):

$$LBP_{P,R}^{riv2} = \begin{cases} \sum_{i=0}^{P-1} s(I_i - I_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1 & \text{otherwise,} \end{cases} \quad (12)$$

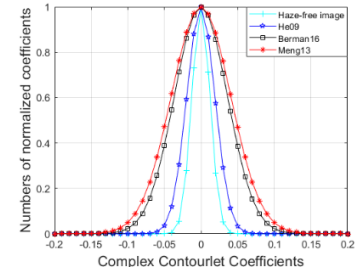
where I_i and I_c are the center location and its neighbor in the reconstructed image, respectively, and P and R are the numbers of neighbors and the neighborhood radius, respectively. U is the uniform measure that calculates the number of bitwise transitions, and $LBP_{P,R}^{riv2}$ has a U value of at most 2. The uniform measure U can be computed as follows (13):

$$U(LBP_{P,R}) = \|s(I_{P-1} - I_c) - s(I_0 - I_c)\| + \sum_{i=0}^{P-1} \|s(I_i - I_c) - s(I_{i-1} - I_c)\|, \quad (13)$$

Then, the normalized probability histogram $H_{LBP}(t)$ is used as the structure-aware feature, and its definition is shown as (14):



(a) Across-scale distribution of normalized subbands



(b) Across-orientation distribution of normalized subbands

Fig. 17. Normalized histogram of coefficients from across-scale and across-orientation for the image in Fig. 16.

$$H_{LBP}(t) = \frac{\sum_{i=1}^M \sum_{j=1}^N f[LBP_{P,R}^{riv2}(i, j), t]}{M * N}, \quad (14)$$

where

$$f[LBP_{P,R}^{riv2}(i, j), t] = \begin{cases} 1, & LBP_{P,R}^{riv2}(i, j) = t \\ 0, & LBP_{P,R}^{riv2}(i, j) \neq t \end{cases} \quad (15)$$

where $f(\cdot)$ is the correspondence between the actual pixel value and the possible value t of each pixel in the image $LBP_{P,R}^{riv2}(i, j)$, and M and N are the two-dimensional sizes of the image. Fig. 15 shows the normalized rotation-invariant and uniform LBP histograms of the examples in Fig. 14. In Fig. 15, it can be found that there are significant differences between the three above images. To accord with the multiscale characteristics of human visual perception, the two scales. The normalized histogram of LBP is chosen as the first

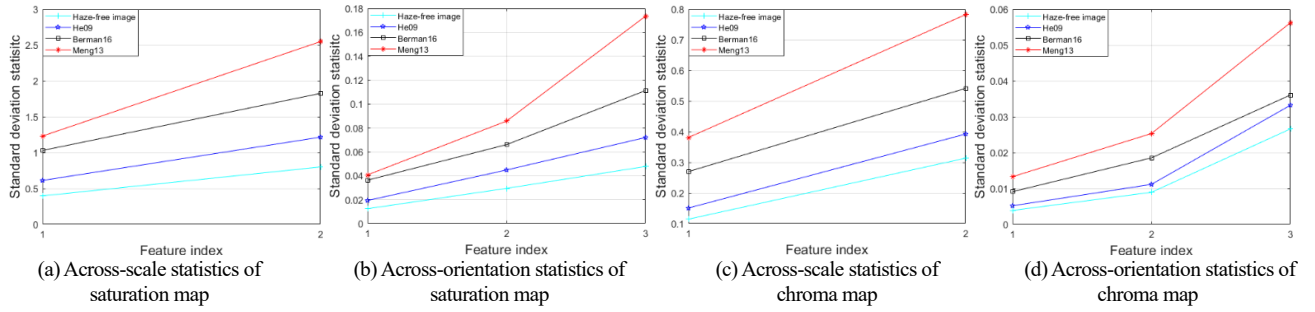


Fig. 18. Statistical analysis of the saturation and chroma map.

kind of distortion-aware feature.

2) Color-aware feature

① Feature map analysis

All of the above factors related to the quality of dehazed images are conducted in the grayscale space, whereas the color information of images reflects the plentiful degree of color scenes and plays a significant role in quality assessment [40]. In addition to the structure distortion artifacts, several dehazing algorithms can inevitably introduce color distortion (color shift, etc.) during the image dehazing process, as shown in Fig. 16. Such artifacts aroused from the side effect of dehazing algorithms destroy the naturalness of the original image, leading to quality deterioration. In the first row, the color distribution of the dehazed image generated by He09 [41] is close to the haze-free image. The color appearance of the dehazed image obtained by Berman16 [42] and Meng13 [43] differs from the original image. Choi et al. [6] found that the saturation-based colorfulness descriptor is an effective fog-aware feature. In addition, a novel chroma image that integrates the saturation and value components of the HSV color space was used for measuring the color change in [5]. The saturation and chroma map of haze-free and different dehazed images are shown in the second and third rows in Fig. 16, respectively. As shown in Fig. 16, the dehazed images corresponding to the latter two methods are far from the ideal haze-free image in terms of the intensity distribution of the saturation and chroma map.

② Across-scale and across-orientation statistics calculation

In this paper, the CCT subband coefficients of the above saturation and chroma map are used to characterize the color distortion. Before researching the statistical properties of the above feature maps, the same GGSM-DNT operation used in calculating the contrast feature is implemented for normalizing the CCT subbands of the saturation and chroma map. There exists a relationship between CCT subbands at each orientation but across scales, which can be affected by distortions. Here, the CCT subbands from different scales but the same orientation are merged to form a 1-D vector. Similarly, the same operation is implemented on the CCT subbands from different orientations but the same scale. The histograms of the across-scale and across-orientation subband coefficients are shown in Fig. 17. As shown in Fig. 17, the normalized histograms of haze-free and dehazed images both obey the generation Gaussian distribution (GGD). The variance in distribution increases when the color distortion of the dehazed image becomes more severe.

In Fig. 18, we plot the distribution of standard deviation parameters corresponding to the across-scale and across-orientation subband coefficients. Confirming the findings in Fig. 17, the change in the standard deviation statistics of the feature

TABLE I
SUMMARY OF THE FEATURES IN THE VDA-DQA MODEL

Feature type	Feature index	Feature description
Visibility-aware features	Sharpness feature F_1-F_3	Log energy of high-frequency CCT subbands
	Brightness feature F_4-F_{21}	Subband entropy of the brightness-changed image
	Contrast feature $F_{22}-F_{33}$	Statistics difference of CCT subbands between dehazed image and contrast-changed image
Distortion-aware features	Structure feature $F_{34}-F_{53}$	LBP histogram statistics of reconstructed image
	Color feature $F_{54}-F_{63}$	Across-scale and across-orientation statistics of saturation and chroma map

maps corresponding to the dehazed images are consistent with the change in color distortion. Thus, the standard deviation corresponding to the distribution of across-scale and across-orientation subband coefficients is chosen as the color-aware feature.

C. Quality Prediction

Here, visibility-aware features, including sharpness, brightness, and contrast features, and distortion-aware features consisting of structure and color features are summarized in Table I. After extracting these perceptual features, a regression model is needed to train for mapping the feature into a quality score Q as follows (16):

$$model = regressor_Train(F_{all}, Q), \quad (16)$$

where $F_{all} = [F_V, F_D]$ denotes the combination of visibility-aware feature set F_V and distortion-aware feature set F_D . Once the model is built, the score of any test dehazed image with its corresponding feature set $F_{all, t}$ can be predicted by Formula (17):

$$Q_t = regressor_Test(F_{all, t}, model), \quad (17)$$

Here, the SVR model implemented by the LIBSVM package [44] is chosen for the quality regression task. The kernel function of SVR is the radial basis function (RBF) kernel.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Database Description

To verify the effectiveness and robustness of the proposed VDA-DQA method, we analyze its performance on the following six publicly available DQA databases: SHRQR database [3], SHRQA database [3], reprocessed D-HAZY database [13], DHQ database [13], IVCDehazing database [29] and exBeDDE database [4].

① The SHRQR database consists of 45 haze-free images, 45 synthetic hazy images, and 360 dehazed images produced from 8 typical image dehazing algorithms.

② The SHRQA database includes 30 haze-free aerial images, 30 synthetic hazy images, and 240 dehazed images produced from the same 8 image dehazing algorithms used in the SHRQR database.

③ The reprocessed D-HAZY database includes 23 haze-free images and the corresponding synthetic hazy images collected from the D-HAZY database [45]. Then, the same 8 dehazing algorithms used in the SHRQR database are also implemented to obtain 184 dehazed images.

④ The DHQ database contains 250 real hazy images and 1,750 dehazed images produced from 7 typical image dehazing algorithms.

⑤ The IVCDehazing database contains 25 hazy images and 200 dehazed images obtained by 7 typical image dehazing algorithms and a Photoshop-based image processing tool.

⑥ The exBeDDE database contains 167 real hazy images captured in several capital cities of China and 1,670 corresponding dehazed images reprocessed by 10 image dehazing algorithms.

The subjective scores of each dehazed image in these databases are provided for DQA studies. Note that the subjective scores of the above six DQA databases, except for the reprocessed D-HAZY database, are in the form of MOS. The subjective score of the reprocessed D-HAZY database is labeled with the predicted results of the DEHAZEfr metric [3].

Among the six above databases, the SHRQR, SHRQA, and reprocessed D-HAZY databases are synthetic DQA databases, and the DHQ, IVCDehazing, and exBeDDE databases are authentic DQA databases. Fig. 19 shows several hazy images in synthetic and authentic DQA databases, and it can be seen that the haze that exists in synthetic hazy images is homogenous, whereas the haze corresponding to authentic hazy images is more complicated than its synthetic counterpart.

Here, three common metrics are chosen to evaluate the performance of the model, namely, the Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SROCC), and root mean square error (RMSE). PLCC and RMSE evaluate the prediction accuracy of the quality metric. SROCC aims to evaluate the monotonic prediction of the model. A value close to 1 for SROCC and PLCC and a value close to 0 for RMSE indicate a superior correlation with human perception. For the PLCC and RMSE criteria, a five-parameter logistic regression function is employed to map the predicted scores to the subjective scores, as follows:

$$q(r) = \beta_1 \left\{ \frac{1}{2} - \frac{1}{1 + \exp[\beta_2(r - \beta_3)]} \right\} + \beta_4 r + \beta_5, \quad (18)$$

where r and $q(r)$ are the original and mapped predicted scores, respectively, and β_1 to β_5 are the fitting parameters of the model.

In this experiment, 80% of the dehazed images from each DQA database are randomly sampled as the training set, and the remaining 20% of the images are sampled as the test set, ensuring independence between the training set and test set. The training-testing procedures are randomly repeated 1,000 times, and the median value of the results is reported.

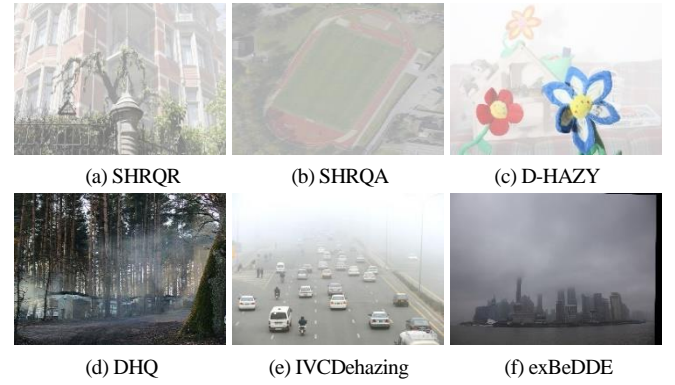


Fig. 19. Examples of hazy images on authentic DQA databases.

B. Comparison on a Single Database

In this section, the performance comparison of 5 general-purpose IQA methods (BRISQUE [1], PSQA-I [46], dipIQ [47], hyperIQA [48] and MetaIQA [49]) and 4 contrast-enhancement and DQA-oriented IQA methods (NIQMC [50], BIQME [27], FADE [6], DHQI [13]) and the proposed VDA-DQA method is evaluated on the six public DQA databases. For the handcrafted feature-based OA-BIQA methods, we retrain them on these DQA databases and obtain the optimum SVR parameters (penalty C and kernel γ) via the grid search method. The SROCC, PLCC and RMSE on synthetic and real haze databases are shown in Tables II and III, and the top two results are marked in bold blue and green, respectively. It is noteworthy that since the exBeDDE database is utilized for pretraining the MetaIQA model, the results of MetaIQA are not reported in Table III.

As shown in Tables II and III, whether in the synthetic or authentic DQA databases, the DQA-oriented methods VDA-DQA and DHQI outperform the state-of-the-art IQA methods, including recent deep-learning-based hyperIQA and MetaIQA, perhaps because it is difficult for them to learn perception rules or meta-knowledge about dehazing methods. These results indicate the large difference between the classical IQA and DQA tasks. Furthermore, DQA-oriented haze density predictor FADE also shows poor performance, which suggests the importance of measuring distortion artifacts in the DQA task.

In Tables II and III, the proposed VDA-DQA shows the best performance on the SHRQR, D-HAZY, IVCDehazing and exBeDDE databases overall, and the DHQI method performs the best only on the SHRQA and DHQ databases. This is because in the SHRQA database, the haze density of hazy images is slight, the structural similarity in the DHQI metric can better measure the structural damage, and the over-enhancement measurement term of the DHQI metric can better deal with the over-enhancement distortion in the DHQ database. However, the overall proposed VDA-DQA is superior to DHQI, which is attributed to both structure-aware features and color-aware features in VDA-DQA, strengthening the discrimination of distortion arising from the side effects of dehazing algorithms.

In addition to the quantitative evaluation results, the scatter plots of objective scores against subjective scores are shown in Fig. 20. Among them, the IQA methods BRISQUE, PSQA-I, dipIQ, hyperIQA, NIQMC and BIQME, along with the DQA-oriented methods FADE, DHQI and VDA-DQA, are utilized for straightforward comparison on the exBeDDE database. As shown

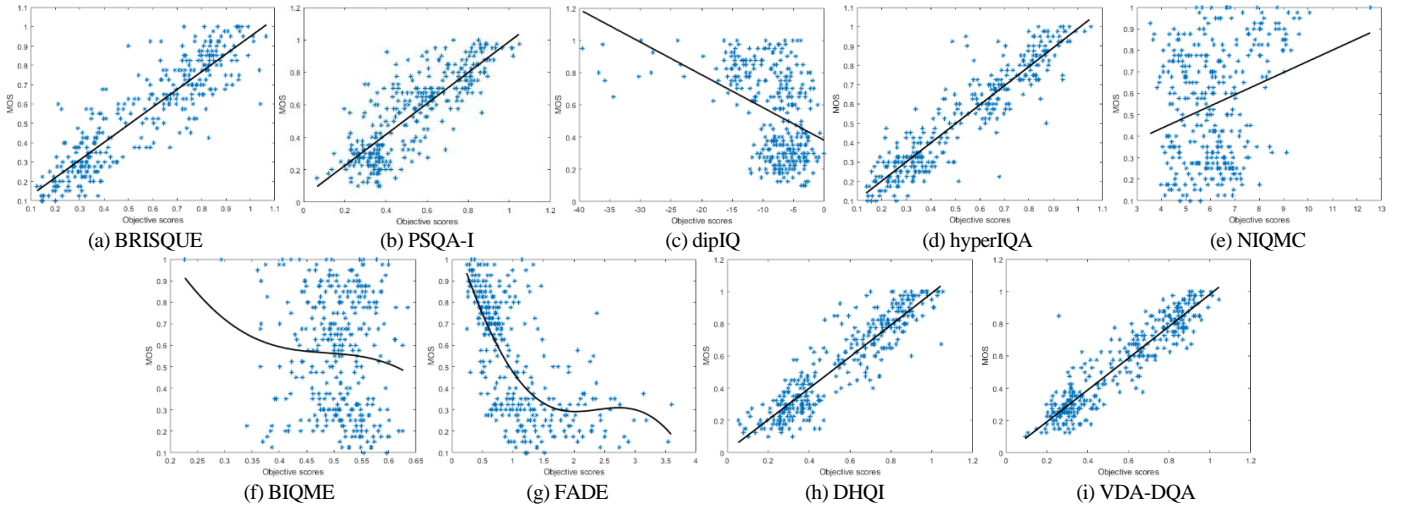


Fig. 20. Scatter plots of predicted scores versus subjective scores on the exBeDDE database.

TABLE II
PERFORMANCE COMPARISON ON SYNTHETIC DQA DATABASES

Database	Metrics	General-purpose IQA methods					Contrast-enhancement IQA methods		DQA-oriented methods		
		BRISQUE [1]	PSQA-I [46]	dipIQ [47]	HyperlQA [48]	MetalQA [49]	NIQMC [50]	BIQME [27]	FADE [6]	DHQI [13]	VDA-DQA
SHRQR	SROCC	0.715	0.743	0.040	0.749	0.757	0.303	0.296	0.358	0.761 ²	0.816 ¹
	PLCC	0.722	0.788	0.224	0.787	0.794	0.304	0.300	0.356	0.830 ²	0.873 ¹
	RMSE	7.889	7.667	13.434	-	-	13.116	13.079	12.951	7.547 ²	6.723 ¹
SHRQA	SROCC	0.671	0.745	0.074	0.781	0.807	0.716	0.710	0.659	0.869 ¹	0.854 ²
	PLCC	0.649	0.735	0.213	0.807	0.818	0.735	0.728	0.698	0.872 ¹	0.853 ²
	RMSE	12.265	8.875	15.694	-	-	10.964	11.070	11.539	7.969 ¹	8.401 ²
Reprocessed D-HAZY	SROCC	0.842	0.788	0.031	0.845	0.852	0.487	0.474	0.242	0.859 ²	0.885 ¹
	PLCC	0.889	0.866	0.126	0.860	0.892	0.513	0.513	0.262	0.916 ²	0.921 ¹
	RMSE	0.094	0.103	0.204	-	-	0.177	0.177	0.199	0.083 ²	0.080 ¹
Avg.	SROCC	0.743	0.759	0.048	0.792	0.805	0.502	0.493	0.420	0.830 ²	0.851 ¹
	PLCC	0.753	0.813	0.187	0.818	0.835	0.517	0.513	0.439	0.873 ²	0.882 ¹
	RMSE	6.749	5.508	9.777	-	-	8.085	8.109	8.230	5.200 ²	5.068 ¹

TABLE III
PERFORMANCE COMPARISON ON AUTHENTIC DQA DATABASES

Database	Metrics	General-purpose IQA methods					Contrast-enhancement IQA methods		DQA-oriented methods		
		BRISQUE [1]	PSQA-I [46]	dipIQ [47]	HyperlQA [48]	MetalQA [49]	NIQMC [50]	BIQME [27]	FADE [6]	DHQI [13]	VDA-DQA
DHQ	SROCC	0.760	0.804	0.163	0.813	0.838	0.373	0.368	0.249	0.876 ¹	0.849 ²
	PLCC	0.784	0.849	0.205	0.852	0.844	0.432	0.425	0.265	0.885 ¹	0.860 ²
	RMSE	8.161	7.516	12.842	-	-	11.864	11.864	12.654	6.271 ¹	6.673 ²
IVCDehazing	SROCC	0.494	0.595 ²	0.007	0.536	0.558	0.282	0.276	0.203	0.561	0.674 ¹
	PLCC	0.542	0.653 ²	0.277	0.572	0.616	0.312	0.298	0.244	0.626	0.705 ¹
	RMSE	1.230	1.115 ²	1.414	-	-	1.394	1.398	1.394	1.138	1.043 ¹
exBeDDE	SROCC	0.890	0.907	0.399	0.917	-	0.148	0.152	0.714	0.919 ²	0.923 ¹
	PLCC	0.906	0.924	0.404	0.926	-	0.177	0.176	0.729	0.939 ²	0.942 ¹
	RMSE	0.113	0.102	0.245	-	-	0.263	0.263	0.183	0.092 ²	0.090 ¹
Avg.	SROCC	0.715	0.769	0.190	0.755	0.698	0.268	0.265	0.389	0.785 ²	0.815 ¹
	PLCC	0.744	0.809	0.296	0.783	0.730	0.307	0.300	0.413	0.817 ²	0.836 ¹
	RMSE	3.168	2.911	4.834	-	-	4.507	4.508	4.744	2.500 ¹	2.602 ²

in Fig. 20, the proposed VDA-DQA performs better than other IQA metrics in terms of convergence and monotonicity, which further indicates that the high correlation of the proposed method with human subjective judgments.

C. Ablation Study

A series of perceptual features, including sharpness, brightness, contrast, structure, and color features, are extracted for designing the DQA method in this paper. To analyze the impact of single

features, they are removed from the feature vectors and tested on the synthetic and authentic DQA databases, and the results are reported in Tables IV and V. In Tables IV and V, the definition of the feature groups is shown as follows:

- Case 1: Only excluding the luminance feature term;
- Case 2: Only excluding the sharpness feature term;
- Case 3: Only excluding the contrast feature term;
- Case 4: Only excluding the structure feature term;
- Case 5: Only excluding the color feature term;

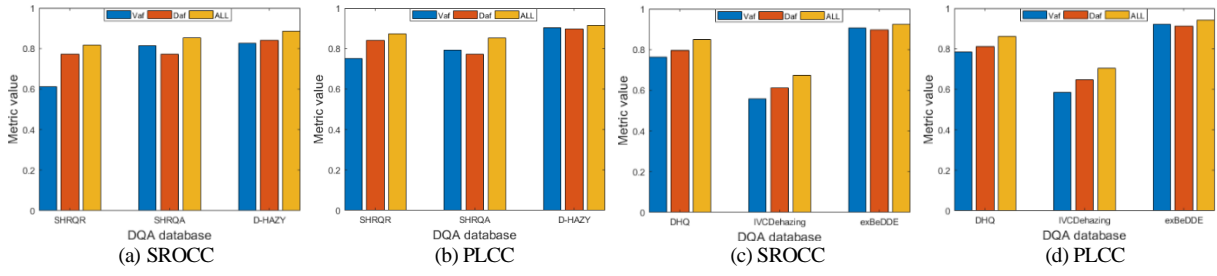


Fig. 21. Comparison of the quality prediction of different features in synthetic and authentic DQA databases.

TABLE IV
RESULTS OF THE ABLATION EXPERIMENTS ON SYNTHETIC DQA DATABASES

Database	Criteria	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
SHRQR	SROCC	0.794	0.731	0.748	0.636	0.765	0.816
	PLCC	0.855	0.808	0.819	0.745	0.815	0.873
SHRQA	SROCC	0.814	0.827	0.805	0.832	0.813	0.854
	PLCC	0.808	0.819	0.796	0.822	0.795	0.853
Reprocessed D-HAZY	SROCC	0.865	0.872	0.840	0.788	0.846	0.885
	PLCC	0.907	0.912	0.897	0.811	0.903	0.916

TABLE V
RESULTS OF THE ABLATION EXPERIMENTS ON AUTHENTIC DQA DATABASES

Database	Criteria	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
DHQ	SROCC	0.806	0.811	0.794	0.764	0.832	0.849
	PLCC	0.823	0.827	0.814	0.789	0.844	0.860
IVCDehazing	SROCC	0.665	0.660	0.617	0.542	0.663	0.674
	PLCC	0.711	0.704	0.653	0.597	0.700	0.705
exBeDDE	SROCC	0.898	0.905	0.885	0.812	0.889	0.923
	PLCC	0.916	0.922	0.903	0.828	0.905	0.942

TABLE VI
RESULTS OF THE PARAMETER SENSITIVITY TEST ON SYNTHETIC AND AUTHENTIC DQA DATABASES.

Database	Criteria	$\alpha = 1$	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$	$\alpha = 8$
IVCDehazing	SROCC	0.648	0.663	0.674	0.670	0.665
	PLCC	0.686	0.690	0.705	0.699	0.698
	RMSE	1.065	1.040	1.043	1.051	1.053

TABLE VII
SROCC RESULTS OF THE INDIVIDUAL DEHAZING ALGORITHM ON THE exBeDDE DATABASE

Categories	Dehazing methods	BRISQUE	FADE	PSQA-I	DHQI	VDA-DQA
Prior-based methods	FVR	0.607¹	0.554	0.562	0.572	0.577²
	DCP	0.162	0.227	0.193	0.291²	0.308¹
	BayD	0.751	0.720	0.740	0.796²	0.842¹
	CAP	0.819	0.604	0.807	0.856¹	0.854²
	NLD	0.527²	0.214	0.586¹	0.586¹	0.502
Deep learning-based methods	MSCNN	0.304	0.107	0.382	0.415²	0.459¹
	DehazeNet	0.371	0.097	0.398	0.425²	0.432¹
	AOD-Net	0.699²	0.450	0.657	0.628	0.713¹
	DCPDN	0.362²	0.108	0.288	0.327	0.392¹
	GFN	0.789	0.226	0.768	0.811²	0.837¹
Avg.		0.539	0.331	0.538	0.571²	0.592¹

Case 6: The proposed VDA-DQA method.

As shown in Tables IV and V, the performance of the metric is inferior to the proposed method while excluding any types of features. It is noteworthy that the structure-aware features

contribute the best to most databases, confirming the finding that humans are highly sensitive to the degradation of structure information in [36]. Contrast-aware features also make a significant contribution to these databases, which is not surprising

TABLE VIII
CROSS-TEST RESULTS ON SYNTHETIC DQA DATABASES

Methods	Metrics	SHRQR (Training)		SHRQA (Training)		Reprocessed D-HAZY (Training)		Avg.
		SHRQA	Reprocessed D-HAZY	SHRQR	Reprocessed D-HAZY	SHRQR	SHRQA	
PSQA-I	SROCC	0.593	0.521	0.469	0.475	0.338	0.553	0.492
	PLCC	0.616	0.546	0.547	0.423	0.387	0.584	0.517
HyperlQA	SROCC	0.712	0.650	0.553	0.540	0.466	0.709	0.605
	PLCC	0.718	0.663	0.557	0.552	0.481	0.719	0.615
MetalQA	SROCC	0.719	0.653	0.558	0.557	0.481	0.712	0.613
	PLCC	0.723	0.662	0.560	0.558	0.505	0.720	0.621
FADE	SROCC	0.659	0.242	0.358	0.242	0.358	0.659	0.420
	PLCC	0.698	0.262	0.356	0.262	0.356	0.698	0.439
DHQI	SROCC	0.722	0.663	0.577	0.565	0.503	0.723	0.626
	PLCC	0.707	0.681	0.610	0.577	0.475	0.723	0.629
VDA-DQA	SROCC	0.740	0.677	0.561	0.563	0.529	0.735	0.634
	PLCC	0.732	0.696	0.566	0.576	0.517	0.731	0.636

TABLE IX
CROSS-TEST RESULTS ON AUTHENTIC DQA DATABASES

Methods	Metrics	DHQ (Training)		IVCDehazing (Training)		exBeDDE (Training)		Avg.
		IVCDehazing	exBeDDE	DHQ	exBeDDE	DHQ	IVCDehazing	
PSQA-I	SROCC	0.454	0.461	0.236	0.268	0.482	0.520	0.404
	PLCC	0.483	0.477	0.258	0.278	0.504	0.535	0.423
HyperlQA	SROCC	0.640	0.612	0.353	0.308	0.628	0.685	0.538
	PLCC	0.678	0.639	0.368	0.322	0.647	0.701	0.559
MetalQA	SROCC	0.647	-	0.379	-	-	-	0.513
	PLCC	0.680	-	0.401	-	-	-	0.541
FADE	SROCC	0.203	0.714	0.249	0.714	0.249	0.203	0.389
	PLCC	0.244	0.729	0.265	0.729	0.265	0.244	0.413
DHQI	SROCC	0.664	0.576	0.431	0.313	0.630	0.701	0.553
	PLCC	0.683	0.589	0.444	0.273	0.650	0.713	0.559
VDA-DQA	SROCC	0.665	0.590	0.415	0.395	0.601	0.722	0.565
	PLCC	0.695	0.618	0.432	0.409	0.606	0.741	0.584

since the primary goal of dehazing is to generate a clean image with high contrast.

After analyzing each type of feature, the performance comparison of visibility-aware features (including sharpness, brightness, and contrast features) and distortion-aware features (structure and color features) are plotted in Fig. 21, where the symbol “Vaf” denotes the visibility-aware features, and “Daf” stands for the distortion-aware features. It can be seen in Fig. 21 that Daf is more important than Vaf, and they are complementary, which supports our motivation that DQA should comprehensively measure haze effects and distortion artifacts.

The only parameter in the VDA-DQA is the weighting factor α in Equation (2), which is utilized to balance the LH, HL and HH parts of the CCT subbands. Here, it is set to 4, which is the same as ref [25]. The comparisons of different values of α are shown in Table VI. It can be seen from Table VI that when α is set to 4, the performances reach the best, which balances the LH, HL and HH parts the best.

D. Performance Evaluation of the Dehazing Algorithms

One purpose of DQA is to evaluate the performance of dehazing methods for optimal parameters. Here, we compare 10 classical dehazing algorithms using the exBeDDE database, 5 of

which are prior knowledge-based methods and the other 5 deep learning-based methods. Table VII shows the SROCC result on each image dehazing method, and the top two performances are marked in bold blue and green, respectively. Avg. is the direct average results of ten dehazing methods.

From Table VII, it can be found that the proposed VDA-DQA performs the best on most dehazing algorithms. Additionally, the result of the VDA-DQA method on the NLD [42] method is comparable to the PSQA-I and DHQI methods. Furthermore, on average, the proposed VDA-DQA far outperforms other QA methods, which further indicates that measuring both haze removal and distortion artifacts is necessary for human visual perception.

E. Cross-database Comparison

Since the VDA-DQA is a training-based method, we test its generalization ability through cross dataset validation on synthetic and authentic DQA databases. The results of the cross-test are listed in Tables VIII and IX, and the top results are marked in blue and bold.

From Table VIII, it can be concluded that the proposed VDA-DQA method outclasses other methods on average. Regardless of when the model is trained on the SHRQR or D-HAZY database and tested on the remaining databases, the VDA-DQA performs

TABLE X
RESULTS OF F-TEST BWTWEEN DIFFERENT MODELS ON THE SHRQR DATABASE

Method	dipIQ	BIQME	NIQMC	FADE	BRISQUE	PSQA-I	HyperIQA	MetaIQA	DHQI	VDA-DQA
dipIQ	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
BIQME	1	0	-1	-1	-1	-1	-1	-1	-1	-1
NIQMC	1	1	0	-1	-1	-1	-1	-1	-1	-1
FADE	1	1	1	0	-1	-1	-1	-1	-1	-1
BRISQUE	1	1	1	1	0	-1	-1	-1	-1	-1
PSQA-I	1	1	1	1	1	0	-1	-1	-1	-1
HyperIQA	1	1	1	1	1	1	0	-1	-1	-1
MetaIQA	1	1	1	1	1	1	1	0	-1	-1
DHQI	1	1	1	1	1	1	1	1	0	-1
VDA-DQA	1	1	1	1	1	1	1	1	1	0

TABLE XI
RESULTS OF F-TEST BWTWEEN DIFFERENT MODELS ON THE ExBeDDE DATABASE

Method	NIQMC	BIQME	dipIQ	FADE	BRISQUE	PSQA-I	HyperIQA	DHQI	VDA-DQA
NIQMC	0	-1	-1	-1	-1	-1	-1	-1	-1
BIQME	1	0	-1	-1	-1	-1	-1	-1	-1
dipIQ	1	1	0	-1	-1	-1	-1	-1	-1
FADE	1	1	1	0	-1	-1	-1	-1	-1
BRISQUE	1	1	1	1	0	-1	-1	-1	-1
PSQA-I	1	1	1	1	1	0	-1	-1	-1
HyperIQA	1	1	1	1	1	1	0	-1	-1
DHQI	1	1	1	1	1	1	1	0	-1
VDA-DQA	1	1	1	1	1	1	1	1	0

TABLE XII
COMPARISON OF THE COMPUTATIONAL COMPLEXITIES BETWEEN DIFFERENT IQA METRICS.

IQA metrics	BRISQUE	PSQA-I	dipIQ	NIQMC	BIQME	FADE	DHQI	VDA-DQA
Time(s)	0.285	1.401	3.867	2.921	3.173	1.504	1.671	5.479

better than the other five models. The performance of the VDA-DQA is only inferior to the DHQI metric when the model is trained on the SHRQA database and tested on the other two databases. The reason may be that there exists an additional over-enhancement distortion in the other two databases compared with the SHRQA database, and the learning ability of the over-enhancement distortion of the VDA-DQA method is weaker than that of the DHQI. Limited by insufficient training samples in synthetic DQA databases, both the deep-learning-based hyperIQA and MetaIQA metrics are inferior to the two DQA-oriented methods overall.

Since MetaIQA needs the exBeDDE database for pretraining, its cross-database experiments are only conducted on the other two authentic DQA databases. As shown in Table IX, the VDA-DQA is superior to all compared metrics when the model is tested on IVCDehazing, which is attributed to the contrast and color distortion measurement terms of VDA-DQA. The FADE shows advantages on the exBeDDE, perhaps because the opinion-free FADE can adaptively learn the pattern of dehazed images in the exBeDDE database. Furthermore, due to the elaborate over-enhancement term, the DHQI achieves the best performance on the DHQ dataset. Overall, the proposed VDA-DQA shows the best performance, which further illustrates that the joint measurement of the visibility and distortion of images strengthens its robustness.

F. Statistical Significance Testing

On the SHRQR database and the exBeDDE database, an F-test based on the discrepancy between the subjective and predicted quality scores was conducted to verify whether the differences

between different IQA models were statistically significant. Among the F-tests, the null hypothesis means that the differences between a pair of IQA models sharing the same distribution are statistically indistinguishable with a confidence of 95%. The results of the F-test depend on the discrepancy between the subjective and predicted quality scores, as shown in Tables X and XI.

In Tables X and XI, the symbol ‘1’ indicates that the method in a row is significantly superior to the counterpart in a column, while the symbol ‘-1’ indicates that the row method is significantly inferior to the column method. The symbol ‘0’ indicates that the methods in the row and column are statistically indistinguishable. From Tables X and XI, the VDA-DQA is significantly superior to other methods on the SHRQR and exBeDDE databases, which suggests that VDA-DQA is the best performer on both the SHRQR and the exBeDDE databases.

G. Running Time Comparison

Apart from the high accuracy, computational efficiency is also a very important quantitative index for IQA metrics in practical applications. To verify the computational efficiency of the IQA metrics, the average time of 100 images randomly selected from the SHRQR database is computed as the complexity index, as shown in Table XII. Noting that training the hyperIQA and MetaIQA methods relies on GPU devices, we only compare the running time of handcrafted feature-based methods here. All methods are tested with MATLAB R2017b on a 3.2-GHz processor with 8 GB RAM. As shown in Table XII, the running

time of the proposed VDA-DQA increased because of the frequency domain decomposition and the GGSM-DNT operation.

V. CONCLUSION

In this paper, we propose a VDA-DQA method by learning both the visibility and distortion measurement of dehazed images. For the visibility measure, we extracted a series of perceptual features in the complex contourlet domain, including the entropy of brightness-changed image, the significant difference between the original image and the contrast-changed image and the high-frequency logarithmic energy of the dehazed image. For distortion perception, two kinds of perceptual features are deployed for realness measurement of dehazed images, consisting of the histogram statistics of LBP computed from the novel reconstructed image and the across-scale and across-orientation statistics of chroma and saturation map. Finally, the SVR model is used to map perceptual features into quality scores. The experimental results verify the good correlation of the proposed VDA-DQA method with human subjective assessment.

In the future, we plan to extend our work from the following two parts. First, inspired by the success of the deep learning-based method in deraining image quality assessment [51], we will utilize powerful deep neural networks to capture abstract semantic features for dehazed image quality assessment. Second, the features extracted in our proposed method will be inserted into some existing models [52]-[54] to boost the detection and recognition performance of particulate matter or smoke.

REFERENCES

- [1] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695-4708, Dec. 2012.
- [2] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process. Image Commun.*, vol. 29, no. 8, pp. 856-863, Sept. 2014.
- [3] 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 Trans. Multimedia.*, vol. 21, no. 9, pp. 2319-2333, Sept. 2019.
- [4] S. Zhao, L. Zhang, S. Huang, Y. Shen, and S. Zhao, "Dehazing Evaluation: Real-World Benchmark Datasets, Criteria, and Baselines," *IEEE Trans. Image Process.*, vol. 29, pp. 6947-6972, 2020.
- [5] W. Liu, F. Zhou, T. Lu, J. Duan, and G. Qiu, "Image Defogging Quality Assessment: Real-World Database and Method," *IEEE Trans. Image Process.*, vol. 30, pp. 176-190, 2021.
- [6] L. K. Choi, J. You, and A. C. Bovik, "Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888-3901, Nov. 2015.
- [7] Y. Jiang, C. Sun, Y. Zhao, and L. Yang, "Fog Density Estimation and Image Defogging Based on Surrogate Modeling for Optical Depth," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3397-3409, July. 2017.
- [8] X. Pan, F. Xie, Z. Jiang, Z. Shi, and X. Luo, "No-Reference Assessment on Haze for Remote-Sensing Images," *IEEE Geoscience and Remote Sensing Letters.*, vol. 13, no. 12, pp. 1855-1859, Dec. 2016.
- [9] 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 Trans. Intelligent Transportation Systems.*, pp. 1-16, 2020.
- [10] P. Mahajan, V. Jakhethiya, P. Abrol, P. K. Lehana, B. N. Subudhi, and S. C. Guntuku, "Perceptual Quality Evaluation of Hazy Natural Images," *IEEE Trans. Industrial Informatics.*, pp. 1-1, 2021.
- [11] N. Hautière, J.-P. Tarel, D. Aubert, and É. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Image Anal. Stereol.*, vol. 27, no. 2, pp. 87-95, 2011.
- [12] S. Fang, J. Yang, J. Zhan, H. Yuan, and R. Rao, "Image quality assessment on image haze removal," in *Proc. Chinese Control and Decision Conference*, Mianyang, China, 2011, pp. 610-614.
- [13] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective Quality Evaluation of Dehazed Images," *IEEE Trans. Intelligent Transportation Systems.*, vol. 20, no. 8, pp. 2879-2892, Aug. 2019.
- [14] 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, Feb. 2017.
- [15] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "Q-DNN: A quality-aware deep neural network for blind assessment of enhanced images," in *Proc. Visual Communications and Image Processing*, Chengdu, China, 2016, pp. 1-4.
- [16] L. Zhang, L. Zhang, and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579-2591, Aug. 2015.
- [17] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.
- [18] B. A. Olshausen and D. J. Field, "How close are we to understanding V1?" *Neural Computation*, vol. 17, no. 8, pp. 1665-1699, 2005.
- [19] C. Li, T. Guan, Y. Zheng, X. Zhong, X. Wu, and A. C. Bovik, "Blind image quality assessment in the contourlet domain," *Signal Process. Image Commun.*, vol. 91, pp. 116064, Feb. 2021.
- [20] C. Li, T. Guan, Y. Zheng, B. Jin, X. Wu, and A. C. Bovik, "Completely blind image quality assessment via contourlet energy statistics," *IET Image Process.*, vol. 15, pp. 443-453, Feb. 2021.
- [21] D. Chen, and Q. Li, "The Use of Complex Contourlet Transform on Fusion Scheme," in *Proc. world academy of science, engineering and technology*, Prague, Czech Republic, 2005, pp. 342-347.
- [22] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Sig. Process. Mag.*, vol. 22, no. 6, pp. 123-151, Nov. 2005.
- [23] A. L. Da Cunha, J. Zhou, and M. N. Do, "The Nonsubsampled Contourlet Transform: Theory, Design, and Applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089-3101, Oct. 2006.
- [24] R. Hassen, Z. Wang, and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798-2810, Jul. 2013.
- [25] P. V. Vu, and D. M. Chandler, "A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation," *IEEE Signal Processing Letters.*, vol. 19, no. 7, pp. 423-426, Jul. 2012.
- [26] D. D. Sandić-Stanković, D. D. Kukolj, and P. Le Callet, "Fast Blind Quality Assessment of DIBR-Synthesized Video Based on High-High Wavelet Subband," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5524-5536, Nov. 2019.
- [27] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301-1313, Apr. 2018.
- [28] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, W. Lin, W. Zhang, and W. Gao, "Blind Quality Assessment of Tone-Mapped Images Via Analysis of Information, Naturalness, and Structure," *IEEE Trans. Multimedia.*, vol. 18, no. 3, pp. 432-443, Mar. 2016.
- [29] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *Proc. International Conference on Image Processing*, Quebec City, QC, Canada, 2015, pp. 3600-3604.
- [30] M. H. Khosravi, and H. Hassanpour, "Blind Quality Metric for Contrast-Distorted Images Based on Eigendecomposition of Color Histograms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 48-58, Jan. 2020.
- [31] C. Li, W. Yuan, A. C. Bovik, and X. Wu, "No-reference blur index using blur comparisons," *IET Electron. Lett.*, vol. 47, no. 17, pp. 962-963, 2011.
- [32] P. Gupta, A. K. Moorthy, R. Soundararajan, and A. C. Bovik, "Generalized Gaussian scale mixtures: A model for wavelet coefficients of natural images," *Signal Process. Image Commun.*, vol. 66, pp. 87-94, Aug. 2018.
- [33] A. K. Moorthy, and A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350-3364, Dec. 2011.
- [34] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind Quality Assessment Based on Pseudo-Reference Image," *IEEE Trans. Multimedia.*, vol. 20, no. 8, pp. 2049-2062, Aug. 2018.
- [35] Q. Li, and Z. Wang, "Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation," *IEEE J. Select. Topics Signal Process.*, vol. 3, no. 2, pp. 202-211, April 2009.

- [36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [37] Q. Li, W. Lin, and Y. Fang, "No-Reference Quality Assessment for Multiply-Distorted Images in Gradient Domain," *IEEE Signal Processing Letters.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
- [38] L. Liu, J. Zhang, M. A. Saad, H. Huang, and A. C. Bovik, "Blind S3D image quality prediction using classical and non-classical receptive field models," *Signal Process. Image Commun.*, vol. 87, pp. 115915, Sep. 2020.
- [39] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [40] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness and structure," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3784–3793, Jul. 2018.
- [41] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [42] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 1674–1682.
- [43] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. International Conference on Computer Vision*, Sydney, NSW, Australia, 2013, pp. 617–624.
- [44] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.
- [45] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. International Conference on Image Processing*, Phoenix, AZ, USA, 2016, pp. 2226–2230.
- [46] L. Liu, T. Wang, and H. Huang, "Pre-attention and spatial dependency driven no-reference image quality assessment," *IEEE Trans. Multimedia.*, vol. 21, no. 9, pp. 2305–2318, Sept. 2019.
- [47] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.
- [48] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network," in *Proc. Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 3664–3673.
- [49] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetalQA: Deep Meta-Learning for No-Reference Image Quality Assessment," in *Proc. Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 14131–14140.
- [50] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.
- [51] Q. Wu, L. Wang, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Subjective and Objective De-Raining Quality Assessment Towards Authentic Rain Image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3883–3897, Nov. 2020.
- [52] K. Gu, Z. Xia, J. Qiao, and W. Lin, "Deep dual-channel neural network for image-based smoke detection," *IEEE Trans. Multimedia.*, vol. 22, no. 2, pp. 311–323, Feb. 2020.
- [53] K. Gu, Y. Zhang, and J. Qiao, "Ensemble meta-learning for few-shot soot density recognition," *IEEE Trans. Industrial Informatics.*, vol. 17, no. 3, pp. 2261–2270, Mar. 2021.
- [54] K. Gu, H. Liu, Z. Xia, J. Qiao, W. Lin, and D. Thalmann, "PM2.5 monitoring: Use information abundance measurement and wide and deep learning," *IEEE Trans. Neural Networks & Learning Systems.*, vol. 32, no. 10, pp. 4278–4290, Oct. 2021.



Tuxin Guan received the bachelor's degree in automation from Jinling Institute of Technology, China, in 2018. He is currently pursuing the Ph.D. degree with the Institute of Logistics Science and Engineering, Shanghai Maritime University. His current research direction is image quality assessment.



Chaofeng Li (Senior Member, IEEE) received the B.S. and M.S. degrees and the Ph.D. degree from the Chinese University of Mining and Technology, Xuzhou, China, in 1995, 1998, and 2001, respectively. From July 2001 to June 2003, he has been finishing postdoctoral research work at the Nanjing University of Science and Technology. He is an Associate Professor and a Professor with the School of Internet of Things Engineering, Jiangnan University, from 2003 to 2017. From 2008 to 2009, he was a Visiting Research Scholar with Laboratory for Image & Video Engineering, the University of Texas at Austin, U.S. He is currently a Professor with the Institute of Logistics Science and Engineering, Shanghai Maritime University. His current research interests include image processing, deep learning, and pattern recognition.



Ke Gu (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He is currently a Professor with the Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning. He received the Best Paper Award from the IEEE TRANSACTIONS ON MULTIMEDIA (T-MM), the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016.



Hantao Liu (Member, IEEE) 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 of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and IEEE SIGNAL PROCESSING LETTERS.



Yuhui Zheng (Member, IEEE) received the B.S. degree in chemistry and the Ph.D. degree in computer science from the Nanjing University of Science and Technology, Nanjing, China, in 2004 and 2009, respectively. From 2014 to 2015, he was a Visiting Scholar with the Digital Media Laboratory, School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology. His current research interests include inverse problems in image processing, data mining, and pattern recognition.



Xiao-jun Wu received the B.Sc. degree in mathematics from Nanjing Normal University, Nanjing, China, in 1991, and the M.S. and Ph.D. degrees in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, China, in 1996 and 2002, respectively. From 1996 to 2006, he was with the School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, China, where he was promoted to Professor. Since 2006, he has been with the School of Information Engineering, Jiangnan University, Wuxi, China, where he is currently a Professor of pattern recognition and computational intelligence. From 2003 to 2004, he was a Visiting Researcher with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, U.K. He has authored or coauthored more than 300 papers in his fields of research. His current research interests include pattern recognition, computer vision, fuzzy systems, neural networks, and intelligent systems.