EHNQ: Subjective and Objective Quality Evaluation of Enhanced Night-Time Images

Ying Yang, Tao Xiang, *Senior Member, IEEE,* Shangwei Guo, Xiao Lv,
Hantao Liu, and Xiaofeng Liao, *Fellow, IEEE*

Abstract—Vision-based practical applications, such as consumer photography and automated driving systems, greatly rely on enhancing the visibility of images captured in night-time environments. For this reason, various image enhancement algorithms (EHAs) have been proposed. However, little attention has been given to the quality evaluation of enhanced night-time images. In this paper, we conduct the first dedicated exploration of the subjective and objective quality evaluation of enhanced night-time images. First, we build an enhanced night-time image quality (EHNQ) database, which is the largest of its kind so far. It includes 1,500 enhanced images generated from 100 real night-time images using 15 different EHAs. Subsequently, we perform a subjective quality evaluation and obtain subjective quality scores on the EHNQ database. Thereafter, we present an objective blind quality index for enhanced night-time images (BEHN). Enhanced night-time images usually suffer from inappropriate brightness and contrast, deformed structure, and unnatural colorfulness. In BEHN, we capture perceptual features that are highly relevant to these three types of corruptions, and we design an ensemble training strategy to map the extracted features into the quality score. Finally, we conduct extensive experiments on EHNQ and EAQA databases. The experimental and analysis results validate the performance of the proposed BEHN compared with the state-of-the-art approaches. Our EHNQ database is publicly available for download at https://sites.google.com/site/xiangtaooo/.

Index Terms—Enhanced night-time image, blind image quality assessment, two-dimensional entropy, biorder structures, ensemble learning.

I. INTRODUCTION

Images captured in night-time environments usually require postprocessing for better visualization and improved utility in practical application scenarios, such as consumer photography and automatic driving systems. In this context, enhancement becomes highly desirable for improving the quality and availability of night-time images. The last few years have witnessed significant advances in the exploration of enhancement algorithms (EHAs) for night-time images [1]–[9]. However, the visual quality evaluation of enhanced night-time images has been largely overlooked in the literature. Therefore, in this

paper, we explore the quality evaluation of enhanced night-time images from both subjective and objective aspects.

In the quality evaluation of enhanced night-time images, two major challenges are encountered. The first challenge is the absence of an enhanced night-time image database with subjective quality scores. A subjective database provides a benchmark for conducting research on objective evaluation metrics. Once a new objective quality evaluation algorithm is proposed, it can be tested on a constructed subjective database to verify its performance. The second challenge is the lack of an objective quality assessment index specially designed for enhanced night-time images. Enhanced night-time images usually possess unique attributes and complicated distortions (e.g., inappropriate brightness and contrast, deformed structure, and unnatural colorfulness), and they need special attention when evaluating their visual quality. An objective metric has beneficial applications in real-time image processing and image-driven systems, such as objective detection and autonomous driving systems. An efficient objective metric can be used to not only assess the performance of different algorithms (systems) but also automatically optimize and select parameters to achieve the best performance and meet different application requirements.

Nowadays, three possible objective solutions can be adapted to estimate the visual quality of enhanced night-time images. The first possible solution is to use traditional general-purpose image quality assessment methods (IQAs) [10]–[16] that are not restricted by image contents or distortions. However, the traditional general-purpose IQAs mainly focus on natural images with commonly encountered artifacts and show unsatisfactory performance on enhanced night-time image evaluation. The second solution is to use contrast-related IQAs, which are specifically designed for gauging the visual quality of contrast-changed/-enhanced images [17]–[21]. However, these contrast-related approaches mainly focus on the contrast distortion and ignore other distortions introduced by the enhancing procedure for night-time images, such as deformed structure and unnatural colorfulness. The third solution is to use IQAs specifically designed for natural night-time images [22]. Since enhancement may change the intrinsic properties and generate additional distortions, the third solution is not up to the task of quality evaluation of enhanced night-time images.

In this paper, we investigate the quality evaluation of enhanced night-time images from both subjective and objective aspects. First, we construct a public large-scale enhanced night-time image quality (EHNQ) database. Images in the EHNQ database are created by enhancing natural night-time

images via different types of EHAs. To obtain the mean opinion scores (MOSs) (ground truth) of each enhanced night-time image, we conduct a carefully designed subjective experiment according to the double stimulus method recommended by ITU-R BT.500-13 [23]. Finally, necessary outlier detection and correction procedures are executed to address the mistakes generated in the subjective experiments.

Second, we conduct an objective evaluation on the constructed EHNQ database using a blind quality evaluation index for enhanced night-time images (BEHN), which is designed by considering three factors that highly define the visual quality of enhanced night-time images: the enhancement effect, the structure deformation caused by the enhancement, and the overall color naturalness. In BEHN, we capture the enhancement-aware features, structure-preserving features, and colorfulness, respectively, to quantify these three factors. Specifically, the enhancement-aware features measure the degree of brightness and contrast. The structure-preserving features characterize the structural deformation, which are represented by the statistical information of the biorder structures. The colorfulness attempts to quantify the color naturalness of the entire enhanced image in three different color spaces. Hereafter, an ensemble training strategy is designed to map the extracted features into a quality score based on the adaptive boosting (AdaBoost) and random forest (RF) algorithms.

The main contributions of this paper are summarized below.

- We construct a large-scale enhanced night-time image quality database, the EHNQ database. The EHNQ database includes 1,500 enhanced images generated from 100 natural night-time images of two different resolutions using 15 well-known EHAs. To the best of our knowledge, the EHNQ database is the largest enhanced night-time image quality database that contains human-labeled ground-truth quality scores.
- We conduct subjective experiments to evaluate the quality of enhanced night-time images in the EHNQ database. In the experiments, we adopt the double stimulus method recommended by ITU under well-prepared settings, and each observer gives a score for every image. After some necessary postprocessing, we obtain the ground truth, i.e. MOS, for every image by averaging all valid scores.
- We perform an objective evaluation of the EHNQ database using our proposed blind quality evaluation index for enhanced night-time images, BEHN. BEHN first extracts enhancement-aware, structure-preserving, and colorfulness features. Then, an ensemble training strategy is designed to map the extracted features into a quality score. Extensive experiments verify the superiority of our approach over state-of-the-art methods.

The remainder of this paper is structured as follows. Section II briefly introduces the related work, including the EHAs used in the EHNQ database and the related quality assessment metrics. Section III presents the construction of the EHNQ database and its subjective evaluation experiments. Section IV elaborates the proposed BEHN scheme. Section V experimentally evaluates the performance of the proposed BEHN. We present the conclusions of the paper in Section VI.

II. RELATED WORK

In this section, we introduce the 15 enhancement algorithms (EHAs) that are used in our constructed EHNQ database and the related subjective and objective evaluations of EHAs.

A. Enhancement Algorithms

With the demand for image enhancement in practical application scenarios, such as consumer photography, automatic driving systems, and monitoring systems, numerous EHAs have been proposed during the past decades. For brevity, in this section, we only describe the 15 representative EHAs employed in our constructed EHNQ database. These 15 EHAs consist of 3 contrast EHAs (CEHAs), 7 low-light image EHAs (LEHAs), and 5 night-time image EHAs (NEHAs). An overview of the selected 15 EHAs is shown in Table I.

CEHAs are proposed to mainly enhance image contrast. For example, histogram equalization (HE) is a typical and effective enhancement technique based on the image histogram [24]. This method is implemented by spreading the gray-level clusters in the histogram. However, HE without any modification usually leads to excessive enhancement, which, in turn, makes the processed image look unnatural and creates visual artifacts. In [25], Lee et al. presented a contrast enhancement algorithm based on the layered difference representation (LDR). LDR attempts to amplify the gray-level differences between the neighboring pixels of the image for contrast enhancement. Wang et al. [26] presented a fusion-based strategy for signal backlit image enhancement (FSBE). The fundamental strategy of FSBE is to blend different features into a single feature to improve the specific quality of the image.

LEHAs are designed for images captured in low-light conditions, such as underwater or in rainy, hazy, or night-time conditions. The seven LEHAs contained in the EHNQ database are described as follows. In [27], Fu et al. proposed a variational framework for enhancing low-light images by employing a bright channel prior (BCP). Wang et al. [28] proposed a naturalness-preserving enhancement (NPE) method for non-uniform illumination images. The NPE method designs a bi-log transformation to map the illumination such that image details and naturalness are balanced. A retinex-based enhancement method via illumination adjustment (REIA) is presented in [29], which can well preserve image edges by exploiting a model without a logarithmic transformation. A probabilistic enhancing model (PEM) is developed by simultaneously estimating the illumination and reflectance in the linear domain [1], which demonstrates that the linear domain exhibits superior performance in representing the prior information for the reflectance and illumination estimation compared to the logarithmic domain. Furthermore, Fu et al. [30] introduced a weighted variational enhancing model (WVEM) by also estimating both the reflection and illumination. Different from traditional variational models, WVEM can suppress noise and preserve the estimated reflectance with more details. A fusion-based method is proposed for weak illumination image enhancement (FWIE) [31]. The fundamental strategy of FWIE is to blend different inputs and weights derived from a single estimated illumination. Guo et al. [2] presented a

TABLE I
AN BRIEF OVERVIEW OF THE SELECTED 15 ENHANCEMENT METHODS IN EHNQ

Database	Category	Algorithm	Abbreviation
EHNQ	Contrast EHAs (CEHAs)	Histogram Equalization [24]	HE
		Layered Difference Representation [25]	LDR
		Fusion-Based Strategy for Signal Backlit Image Enhancement [26]	FSBE
	Low-Light Image EHAs (LEHAs)	Bright Channel Prior [27]	BCP
		Naturalness-Preserving Enhancement [28]	NPE
		Retinex-Based Enhancement via Illumination Adjustment [29]	REIA
		Probabilistic Enhancement Model [1]	PEM
		Weighted Variational Enhancing Model [30]	WVEM
		Fusion-Based Method for Weak Illumination Image Enhancement [31]	FWIE
		Low-Light Image Enhancement [2]	LIME
	Night-Time Image EHAs (NEHAs)	Color Night-Time Image Enhancement via Fuzzy subsets [32]	CNEFS
		Improved Multi-Scale Retinex-Based Enhancing Model [33]	IMSRE
		Photographic Negative Imaging-Inspired Enhancement [4]	PNIE
		Retinex-Based Algorithm Based on Guided Filters [34]	RGF
		Improved Multi-Scale Retinex [3]	IMSR
EAQA	Low-Light Image Subset	Histogram Equalization [24]	HE
		Classic Histogram Adjustment	CHA
		Classic Gamma Correlation	CGC
		Enhancing Algorithm for Low-Light Videos [35]	ELLV
		Color Night-Time Image Enhancement via Fuzzy Sets [32]	CNEFS

low-light image enhancement (LIME) method. In LIME, the illumination is first estimated individually by computing the maximum value in RGB channels, and then the illumination map is refined by imposing a structure prior on it.

NEHAs are specially designed for night-time images. In [32], Cai and Qian proposed a color night-time image enhancement method based on a fuzzy set (CNEFS). CNEFS enhances the dark regions, restrains the glare regions, and adjusts the contrast, which can make the image have uniform brightness and more details. In [33], Lin and Shi proposed an improved multi-scale retinex-based enhancing model (IMSRE) for night-time images. Since the original multi-scale retinex (MSR) is sensitive to the noise speckles that usually appear in night-time images, IMSRE uses a customized sigmoid function instead of a logarithm function in MSR to achieve better performance. Inspired by photographic negative imaging, Shi et al. [4] proposed PNIE. First, a negative image is generated by reversing the input night-time image. Then, an image dehazing method is used to rectify the negative image. Finally, the rectified negative image is revised to create an enhanced image. A retinex-based scheme based on a guided filter (RGF) is proposed for night-time image enhancement in [34]. In [3], Kuang et al. used an improved multi-scale retinex (IMSR) for night-time image enhancement, which can extract the regions of interest (ROIs) accurately and enhance images effectively for accurate night-time vehicle detection.

B. Quality Evaluation of EHAs

In addition to EHAs themselves, the performance evaluation of EHAs and the quality assessment of enhanced images are also critical. A proper IQA index for enhanced images not only can be used to quantify the quality of experience (QoE) of consumers, but also can be applied to effectively optimize and benchmark practical applications, such as autonomous driving systems and visual surveillance. In the current literature, most EHAs are qualitatively evaluated by human observers and quantitatively evaluated by computer algorithms.

1) Subjective Evaluation: Subjective evaluation is a straightforward and accurate method because images are visually perceived by humans. Subjective evaluation is of great

significance for the research community from two aspects. First, subjective evaluation is necessary for building databases. Constructing a quality assessment database requires subjective rating scores, which are generated via a subjective evaluation experiment. Such a created subjective database provides ground-truth scores with which the scores obtained by objective algorithms should be highly consistent. Chen et al. [36] proposed an enhancement algorithm quality assessment (EAQA) database consisting of three subsets: haze, underwater, and low-light images. Each subset consists of 500 images enhanced by 5 EHAs from 100 original images. The 5 EHAs used to enhance the low-light images in the EAQA database are listed in Table I. However, the type of EHAs contained in the EAQA database is inadequate. Moreover, the EAQA database only gives the performance ranks of the 5 EHAs rather than the absolute quality scores of enhanced images.

Second, subjective evaluation can serve as a qualitative tool for comparing the performance of different EHAs. A common practice is to list a few images with poor visibility and their corresponding enhanced versions for subjective comparison [2], [25], [28]. Although this method complies with intuition, there are two problems: The number of the listed images is extremely limited, which cannot demonstrate the general performance of these EHAs. Moreover, subjective evaluation is time-consuming and labor-expensive, making it infeasible for vast amounts of data.

2) Objective Evaluation: Objective evaluation provides a quantitative method for measuring the performance of EHAs and the quality of enhanced images. Recent years have witnessed the rapid development of objective IQAs for enhanced images. In [19], the authors presented the RIQMC method, which is a reduced reference (RR) method for evaluating the quality of contrast-changed images using image histograms and phase congruency. In [18], the authors proposed an RR method, namely, C-PCQI, and a no-reference/blind (NR/blind) method, namely, BIQME, for contrast-enhanced images. C-PCQI measures the similarity of the mean intensity, contrast change, and structural distortion between original and enhanced images, and BIQME analyzes the contrast, sharpness, brightness, and other features. In [17], the authors designed

NIQMC to assess the quality of contrast-distorted images according to information maximization theory. Fang et al. [37] proposed NR-CDIQA for contrast distortion via natural scene statistics (NSS) regulation, involving the mean, standard deviation, skewness, kurtosis, and entropy.

Although the above-mentioned IQAs achieve promising results on contrast-changed or contrast-enhanced images, they perform unsatisfactorily on enhanced night-time images. The reason is that they mainly focus on contrast changes and enhancement and do not consider the diverse and complex distortions that are highly related to enhanced night-time images, such as structural deformation and unnatural colorfulness.

III. EHNQ Database

In this section, we build the EHNQ database and conduct subjective experiments to obtain ground-truth quality scores.
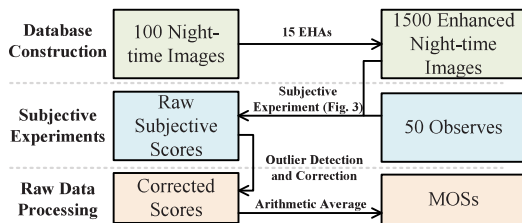

Fig. 1. Pipeline of the subjective evaluation experiments.

A. Pipeline

The procedure of the subjective evaluation experiment for constructing a quality assessment database should strictly follow the recommendation of ITU-R BT.500-13 [23], as our work and other related works [38], [39] do. Fig. 1 gives the pipeline of the subjective evaluation experiments, where three steps are mainly contained for the construction of the EHNQ database. The first step is database construction, which collects a large number of enhanced night-time images with diverse types and contents. Then, a sufficient number of observers are invited to rate the visual quality of the collected images by choosing an appropriate evaluation method under a carefully-arranged experimental environment, and obtain the raw subjective scores. Last, the raw subjective scores are further processed by outlier detection and correction to generate the final MOS value for each image.

B. Database Construction

We select 100 natural night-time images from NNID [22], which is the first public database for night-time image quality evaluation. NNID contains 2,240 natural night-time images with 448 different image contents captured by three different photographic devices in real-world scenarios. The selected 100 images cover diverse night-time scenes that often appear in daily life, including buildings, humans, traffic signs, landscapes, and roads, among other objects. According to the selected 100 natural night-time images, we generate 1,500 enhanced images using 15 representative state-of-the-art EHAs, which are described in Section II-A. The 15 EHAs consist of
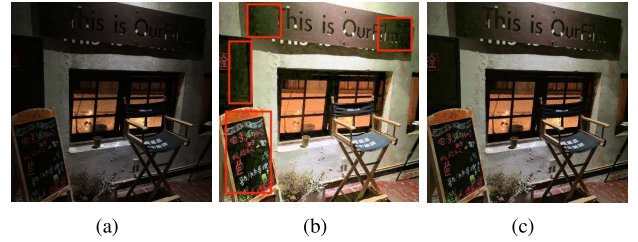

Fig. 2. Role of the original night-time image in subjective evaluation. (a) Original night-time image; (b) enhanced image 1; and (c) enhanced image 2.

3 CEHAs, 7 LEHAs, and 5 NEHAs, and basic information on them is listed in Table I. Our constructed EHNQ database contains two subsets with different resolutions: subset-512 and subset-1024. The resolution of half of the images in the EHNQ database is 512×512 pixels, and the resolution of the other half is 1024×1024 pixels.

C. Subjective Evaluation Methodology

Subjective experiments are strictly conducted based on the standard double-stimulus method [23] recommended by ITU-R BT.500-13, which compares the image that need to be evaluated with the reference image (ground-truth). However, a perfect high-quality (night-time or enhanced) image is unavailable as a reference. Following [38], [39], which constructed a subjective quality assessment database by comparing enhanced images with their original low-quality versions, we use the original night-time image as the reference here. The double-stimulus method is chosen are two-folds. First, the double stimulus method is a more intuitive method for the observers to detect and rate the changes in an image with respect to its reference [40]. Second, the double stimulus method can reflect additional distortions caused by the enhancement algorithms. Fig. 2 shows the role of the original night-time image in the subjective test. If the original night-time image Fig. 2(a) is unavailable, the subjective score of Fig. 2(b) would be better than that of Fig. 2(c). However, Fig. 2(b) suffers from additional distortions (blur and color distortions), which are labeled in the red boxes. Therefore, the enhanced quality of Fig. 2(c) is better than that of Fig. 2(b) in our subjective experiments, whereas we would probably obtain the opposite conclusion without the presentation of Fig. 2(a).

There is another way to conduct the subjective experiment by providing two enhanced images with their reference image for comparison, but it will affect the final quality evaluation of the individual enhanced image. For instance, we have three enhanced images A, B, and C (the enhancement quality scores are 1, 2, and 5, respectively). If we want to assess the quality of A and B by giving the enhanced image C and the original reference image for compassion, the observer may pay more attention to comparing A and C or B and C instead of focusing on the difference between the original image and the enhanced image we want to evaluate because C is much better than A and B. This may result in biased scores. Thus, in our subjective test, we only provide the corresponding reference image of an enhanced image for comparison. The rating scales include 9 discrete points from 1 to 5, which correspond to five quality
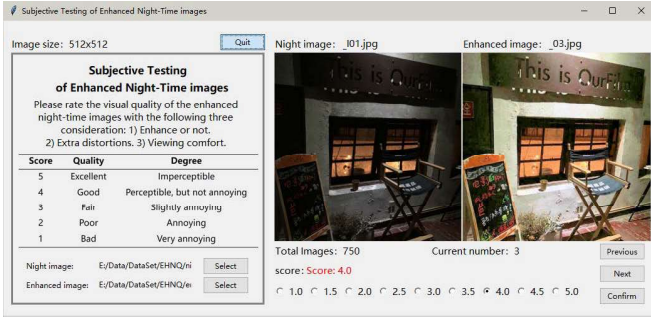
Fig. 3. Graphical user interface for subjective evaluation.

levels: bad, poor, fair, good, and excellent. A higher score indicates better visual quality.

We recruit 50 observers to participate in our subjective experiments, including 21 females and 29 males. They are divided into two groups: 25 observers (11 females and 14 males) rate the enhanced images in subset-512, and the remaining 25 observers (10 females and 15 males) participate to rate the images in subset-1024. All the observers are college students majoring in business, science, art, engineering, or similar fields; only four observers have background knowledge of image processing. They all have normal or corrected vision, and they range from 20 to 30 years old.

We design a graphical user interface (GUI) for conducting the subjective evaluation process, which is exhibited in Fig. 3. Both the enhanced image to be rated and its original night-time image are shown on a 4K monitor with a resolution of 3840×2160. All the images are shown at the original resolution to prevent extra distortions. The viewing conditions of the subjective experiments are consistent with the recommendations of ITU-R BT.500-13 [23]. The observers are required to sit at a fixed viewing distance of approximately four times the image height in a laboratory environment with normal indoor illumination conditions. The incident light falling on the screen is 80 lux, and the environmental illumination from behind the monitor is 240 lux.

The observers are asked to rate the visual quality for at least 5 seconds per image (including the switching time between enhanced images). To avoid visual fatigue, the observers are required to take a break every 20 minutes. A short training session is given to exhibit the approximate range of visual quality of enhanced night-time images. The observers are told to rate the quality of each enhanced night-time image by considering three main aspects. First, compared to the original night-time image, are the brightness and contrast enhanced or not in the enhanced image? Second, does the enhancement introduce additional distortions? Third, how natural is the enhanced night-time image?

D. Raw Data Processing

1) Outlier Detection and Correction: After completing the subjective rating, we obtain the original raw subjective scores given by all the observers. However, it is difficult for the observers to give accurate scores for all enhanced night-time images. Some mistakes may happen in the subjective rating

procedure, which are defined as the outliers of subjective scores. Specifically, if a subjective score is outside the 95% confidence interval of the mean subjective value, it is defined as an outlier. In this work, we follow the method in [23] to automatically detect and correct the outlier scores.

2) MOS: MOS calculation is a widely used method to generate subjective scores. When evaluating an image, different observers may assign different quality scores. To remove the different scale associated with each observer, z-scores [41] are applied to calibrate the corrected scores. The arithmetic average value of the z-scores from all observers on each enhanced night-time image is scaled to the MOS:

$$s_j = \frac{1}{N} \sum_{i=1}^{N} s_{i,j} \qquad (1)$$

where $s_{i,j}$ is the z-score of the jth enhanced image given by the ith observer, N is the number of observers after outlier rejection, and s_j denotes the MOS value of the jth image.

E. MOS Analysis

Generally, the MOS values of a good IQA database should cover the entire range of visual quality, from bad to excellent [22]. Since our EHNQ database consists of enhanced night-time images, the visual quality of the images tends to be higher. Fig. 4 shows the histograms of the MOS values of the entire EHNQ database, subset-512, and subset-1024. The MOS values of the entire database span the whole visual quality scale [1, 5] and have a good spread at different quality levels. Furthermore, the MOS histograms of sub-512 and sub-1024 exhibit similar behavior. This phenomenon shows that the constructed EHNQ database is valuable for quality evaluation, since the result of the subjective test is not affected by the differences in image resolution and content.

In Fig. 5, we also show the histograms of the MOS values of images enhanced by different EHAs in the EHNQ database. Based on the results of Fig. 5, we roughly compare the performance of each algorithm. For the three CEHAs (Fig. 5(a)-5(c)), we find that the MOS values of images enhanced by conditional HE [24] range from 3 to 5, those of the images enhanced by FSBE [26] range from 2 to 4.5, and those of the images enhanced by LDR [25] range from 1 to 3.5. Therefore, we conclude that the performance ranking order of the three CEHAs is HE > FSBE > LDR. Similarly, for the seven LEHAs, NPE [28], REIA [29] and LIME [2] outperform the other LEHAs. Among the NEHAs, PNIE [4] and IMSR [3] exhibit better enhancement performance since the MOS values of images enhanced by these two algorithms are distributed at higher quality levels.

IV. PROPOSED METHOD

In this part, we introduce in detail the proposed BEHN, which is a blind quality evaluation index for enhanced night-time images. First, we present its basic definition and framework. Then, we present the processes of extracting enhancement-aware features, structure-preserve features, and colorfulness. Last, we introduce in detail the proposed ensemble training strategy for feature pooling and quality prediction.
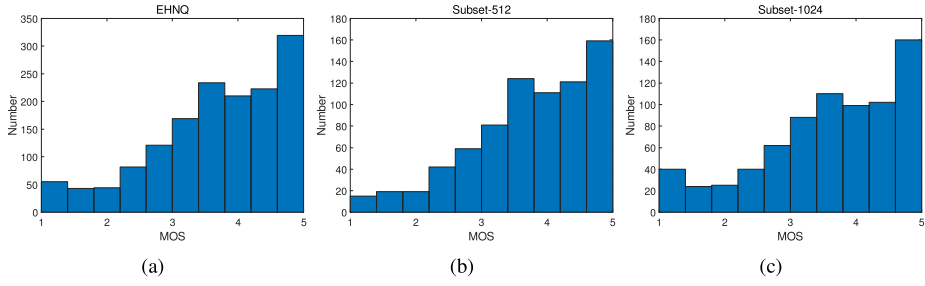
Fig. 4. Histogram of MOS values in the EHNQ database. (a) MOS distribution of the entire EHNQ database; (b) MOS distribution of the images of size 512×512; and (c) MOS distribution of the images of size 1024×1024.
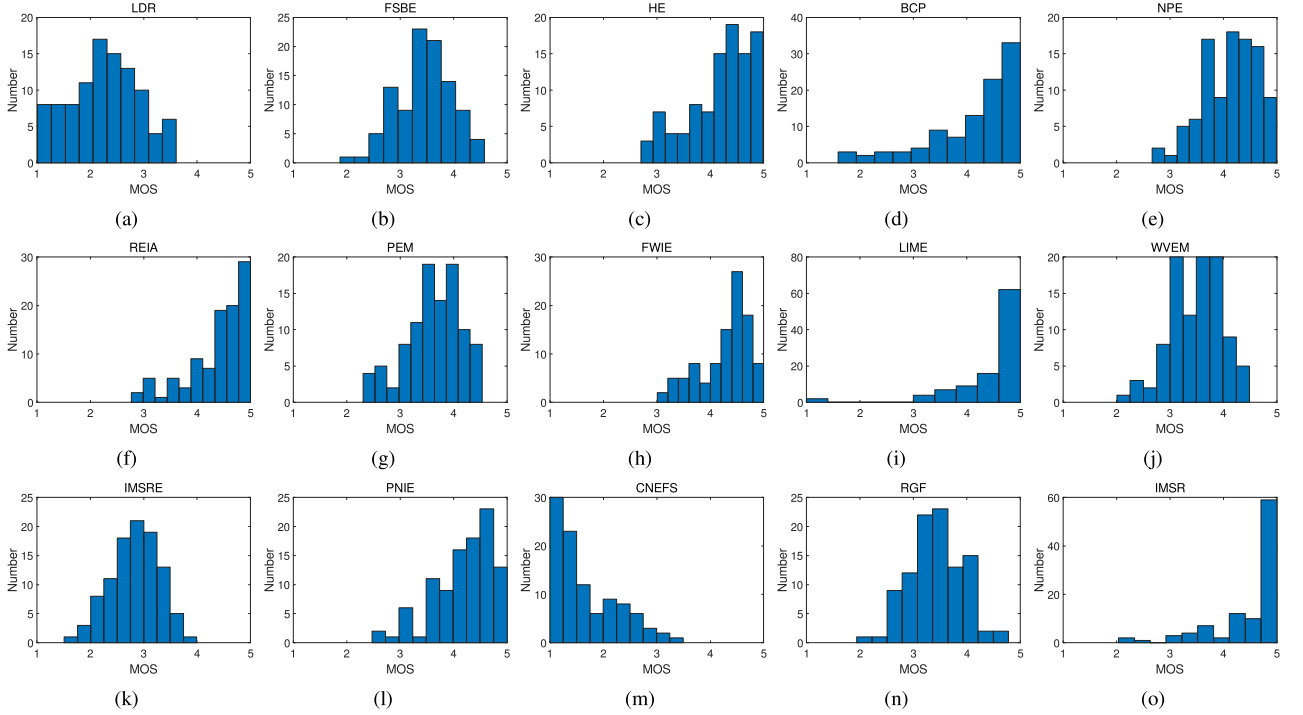

Fig. 5. Histograms of MOS values of the images enhanced by different EHAs. (a)-(c) Three CEHAs; (d)-(j) seven LEHAs; and (k)-(o) five NEHAs.

A. Basic Definition and Framework

Fundamentally, our proposed BEHN measures the perceptual features that are characteristic of enhanced night-time images and critical for visualization. Night-time images suffer from visible issues such as decreased brightness and contrast and loss of image details, while enhancement algorithm attempts to improve the brightness and contrast, and recover image details. Therefore, a good EHA for night-time images is expected to enhance the image content as much as possible, preserve the image structure from damage, and prevent unnatural colorfulness. Based on these considerations, we design BEHN by extracting and fusing three groups of features, namely, enhancement-aware features, structure-preserving features, and colorfulness. Subsequently, we combine the extracted features into the final BEHN index by training a regression model with an ensemble training strategy. Fig. 6 shows the framework of the proposed BEHN, which works as follows. First, we extract the brightness and contrast as the enhancement-aware features. The brightness is captured based on the dark channel prior (DCP) and the luminance channel of the LMN color space. The contrast is characterized by the two-dimensional (2D) entropy and the normalized singular value (NSV) of multiple intermediate images. Second, we extract the structure-preserving feature based on the statistical information of the biorder structures. Third, we capture the color naturalness in three different color spaces. Since the human visual system (HVS) perceives an image at multiple scales, we extract features at three scales (with a down-sampling factor of 0.5) to alleviate the variations in image resolution and viewing distance. Finally, an ensemble training strategy, AdaBoost-RF, is proposed for pooling the extracted features into the overall quality score based on the AdaBoost algorithm and RF regression.

B. Enhancement-Aware Features

Night-time images exhibit visibility problems, such as brightness and contrast reduction, while EHAs try to recover the lost brightness and contrast. Thus, we extract the brightness
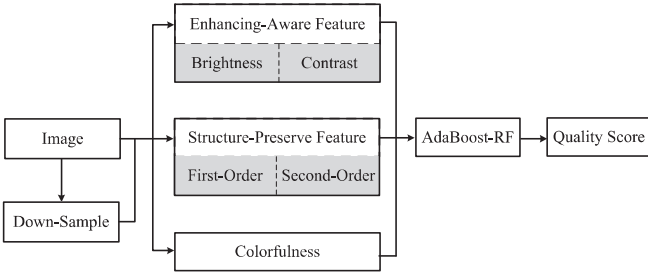
Fig. 6. Framework of the proposed BEHN.

and contrast features to measure the enhancing effect of enhanced night-time images.

1) Brightness: Brightness is a critical property for the visualization of enhanced night-time images. We capture the brightness of each image in two steps. First, we compute the DCP map and the luminance map of LMN color space. Then, the arithmetic average value, logarithmic average value, and skewness of these two maps are calculated to obtain a brightness feature vector.

DCP denotes the lowest intensity among the R, G, B color channels in an image [42]. Night-time images usually exhibit a darker DCP map, and enhancement generally leads to a brighter DCP map. Inspired by this regular change, the local pixel-wise DCP map is leveraged to measure the enhancement degree, which is defined as

$$I^D(i,j) = \min_{\Omega(i,j)} \left[\min_{o \in \{R,G,B\}} (I^o(i,j)) \right] \quad (2)$$

where $\Omega(i,j)$ denotes a 15×15 window centered at (i,j) and $o \in \{R, G, B\}$ denotes the red, green and blue channels.

LMN color space is well-defined on a physical basis and optimized on the HVS [43]. Converting RGB color space to LMN color space can eliminate the correlation between the luminance component (L) and chrominance components (M, N). The luminance component can well characterize the brightness of the image. This conversion is defined as

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & \text{-}0.35 \\ 0.34 & \text{-}0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (3)$$

Then, we compute the arithmetic average value, logarithmic average value, and skewness of the DCP map I^D and the luminance map I^L as the brightness feature. The arithmetic average value represents the absolute light condition across the whole image, which is formulated as

$$BA^c(i,j) = E(I^c(i,j)) = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} I^c(i,j) \quad (4)$$

where $c \in \{D, L\}$ and H and W denote the height and weight of the image. The logarithmic average value is the conjunct representation for the brightness and dynamic range of an image [44], which can be represented as

$$BL^c(i,j) = \frac{R^c}{HW} \exp \left(\sum_{i=1}^{H} \sum_{j=1}^{W} \log \left(\epsilon + \frac{I^c(i,j)}{R^c} \right) \right) \quad (5)$$

where $c \in \{D, L\}$, R^c is the maximum dynamic range of the DCP map or luminance map (i.e., $R^D = 255$ and $R^L = 255$) and ϵ is a small number to avoid computing $\log(0)$. The HVS employs the skewness or a similar measure of histogram asymmetry to perceive the surface quality of an image. An image with a small skewness value tends to brighter and glossier [45]. Thus, we also compute the skewness of these two maps for brightness measurement:

$$BS^c(i,j) = \frac{E[I^c(i,j) - E(I^c(i,j))]^3}{\sigma^3(I^c(i,j))} \quad (6)$$

where $\sigma(I^c(i,j))$ denotes the variance value of image $I^c(i,j)$.

After the above calculation, we obtain six brightness components and combine them to obtain a brightness vector

$$F_{brightness} = \begin{bmatrix} BA^D, BA^L, BL^D, BL^L, BS^D, BS^L \end{bmatrix}. \quad (7)$$

2) Contrast: Contrast refers to the perceived magnitude of the visually meaningful differences within an image, and it has a special significance in gauging the effect of image enhancement. Enhanced night-time images generated by a good EHA should have appropriate contrast and comfort perception. Here, we discover that the 2D entropy [46] and NSV of multiple intermediate images are highly related to the contrast of enhanced night-time images.

By gradually raising/reducing the original luminance of an image, we can obtain multiple intermediate images. The degree of the variation can reflect the contrast of the original image. Thus, we capture the contrast features from multiple intermediate images for quality evaluation of enhanced night-time images here. The intermediate image I_r of an enhanced night-time image I can be obtained by

$$I_r = \max\left(\min(m_r \cdot I, t_u), t_l\right) \quad (8)$$

where m_r denotes the rth multiplier, for $r = \{1, 2, ..., N\}$; t_u and t_l denote the upper and lower bounds, respectively; and the max and min operators are used to confine the intermediate images to the range of $[t_l, t_u]$. Here, we set $t_l = 0$ and $t_u = 255$ for an 8-bit image, and we select $N = 9$ multipliers $m_r = \{1/n, 1, n | n = 3, 5, 7, 9\}$ by considering the balance between efficacy and efficiency. Note that $I_r = I$ when $m_r = 1$.

The 2D entropy of intermediate images can be used to effectively distinguish images with different contrast because compared to traditional entropy, 2D entropy can reflect both the information amount and distribution characteristics of gray values. Fig. 7 exhibits the relation between the multiplier and the corresponding 2D entropy value, where $m_r < 1$ indicates decreasing luminance and $m_r > 1$ indicates increasing luminance. The 2D entropy of Fig. 7(c) quickly falls down to a low level when the luminance is increased. This is because Fig. 7(c) already has proper contrast, and increasing its luminance harms its visual quality. Thus, we use 2D entropy to deduce whether an enhanced night-time image has proper contrast.

The 2D entropy of an image is calculated as follows. First, for a pixel (i, j) of the intermediate image $I_r(i,j)$, the average value of all the adjacent pixels of $I_r(i,j)$ in a window is defined as

$$\widetilde{I_r}(i,j) = \frac{\sum_{(k,l) \in \Omega} I_r(k,l) - I_r(i,j)}{W^2} \quad (9)$$
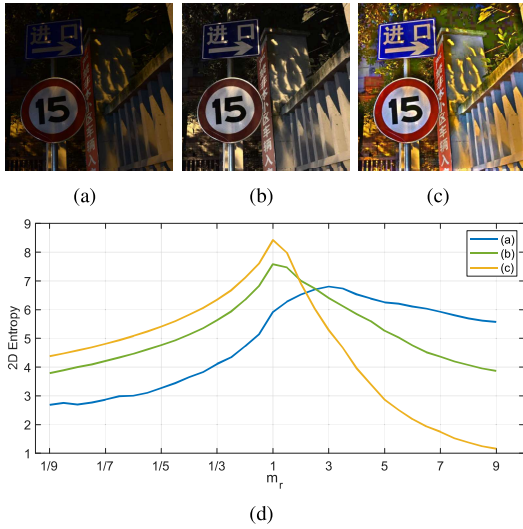
(a)   (b)   (c)

(d)

Fig. 7. Illustrations of the variance of the 2D entropy with the multiplier. (a) An original night-time image, (b)-(c) enhanced images with MOS=3.05 and MOS=4.96, and (d) the relation between the 2D entropy and the multiplier.

where Ω is the set of adjacent pixels of $I_r(i,j)$ in a window of size $W \times W$. In our work, we set $W = 3$. Then, we define a function $f_{m,n}$ to count the number of pixels that simultaneously satisfy $I_r(i,j) = m$ and $\widetilde{I}_r(i,j) = n$ for all (i,j). The probability is computed as

$$P_{m,n} = \frac{f_{m,n}}{HW} \tag{10}$$

where H and W are the height and the width of the image. After that, the 2D entropy of the image is calculated as

$$E_{2D} = -\sum_{m=0}^{255} \sum_{n=0}^{255} P_{m,n} \log P_{m,n} \tag{11}$$

In addition to the 2D entropy, we also compute the singular values of multiple intermediate images to capture the contrast change degree among these intermediate images. A given an intermediate image I_r can be decomposed into

$$I_r = USV^T \tag{12}$$

where U and V are orthogonal matrices. The singular values are $S = diag\{\xi_1, \xi_2, ..., \xi_z\}, z = rank(I_r)$. To reduce the effect of image content on contrast measurement, we normalize the average singular value by

$$\hat{S} = \frac{\sum_{t=1}^{z} \xi_t}{z\sigma^2} \tag{13}$$

where σ is the variance of the image.

For each intermediate image I_r, we obtain a 2D entropy E_{2D}^r and a singular value $\hat{S}^r$. Since we have $N = 9$ intermediate images, we combine all these components to form the contrast feature vector:

$$F_{contrast} = \left[E_{2D}^1, \hat{S}^1, ..., E_{2D}^r, \hat{S}^r, ..., E_{2D}^9, \hat{S}^9 \right]. \tag{14}$$

Finally, the enhancement-aware feature is generated by combining the brightness feature vector $F_{brightness}$ and contrast feature vector $F_{contrast}$, which is expressed as

$$F_e = [F_{brightness}, F_{contrast}]. \tag{15}$$

C. Structure-Preserving Feature

Image structure is a significant cue for the quality prediction of enhanced night-time images because the HVS is highly sensitive to structural degradation [47], [48]. Enhanced night-time images usually suffer from structural artifacts and deformations caused by the enhancement procedure, which severely damage their visual quality. Fig. 8 presents two typical examples of structural artifacts introduced by the enhancement of night-time images. One typical distortion is intrinsic structural damage, which is found in Fig. 8(a)-8(b). The second structural artifact is over-enhancement, which usually occurs in low contrast regions. For example, the image details in the upper-left area are hardly observed in Fig. 8(d), however, this area is enhanced in the image structure in Fig. 8(e).

Therefore, in this paper, we extract the structure-preserving features to capture the over-enhancement and intrinsic structural distortion of enhanced night-time images. The human visual cortex has been demonstrated to perceive the first-order and second-order structures (biorder structures) individually [49]. We leverage statistical information on the biorder structures to measure the structure deformation, which is represented by the histograms of the high-order gradient magnitude (GM) and completed local binary pattern (CLBP).

The first-order structure refers to the dominant contents of an image, such as edges, which convey the dominant structure that is highly sensitive to visual perception [50]. We propose a high-order GM map for capturing the first-order structure, which well captures the structural differences of enhanced night-time images. Examples of high-order GM maps of a night-time image and its enhanced images Figs. 8(a)-8(c) are shown in Figs. 8(g)-8(i). We observe that the high-order GM map captures intrinsic structural information in Fig. 8(h), and its histogram also reflects the structural deformation because the distribution deviates greatly from that of the original night-time image. In contrast, the histogram of the high-order GM map of the enhanced image in Fig. 8(o) without structural distortion has a similar distribution to that of Fig. 8(m). Thus, the statistical information of the high-order GM can be used to distinguish structural corruptions, which is conductive to the quality evaluation of enhanced night-time images. The GM of an image is given by

$$GM(i,j) = \sqrt{[I(i,j) \otimes h_x]^2 + [I(i,j) \otimes h_y]^2} \tag{16}$$

where $\otimes$ is the linear convolution operator and $h_d, d \in (x,y)$ denotes the Gaussian partial derivation filter applied along the horizontal (x) and vertical (y) directions:

$$h_d(i,j|\theta) = -\frac{1}{2\pi\theta^2} \frac{d}{\theta^2} exp\left(-\frac{i^2+j^2}{2\theta^2}\right)^{d \in \{x,y\}} \tag{17}$$

where θ denotes the scale coefficient. The high-order GM maps are obtained by taking $GM(i,j)$ as input; these maps are denoted $GM^m(i,j)$, where $h = 1, 2, ..., H$. The final first-order structure maps can be obtained by

$$HGM(i,j) = \sum_{h=1}^{H} GM^h(i,j) \tag{18}$$

(a) Night-Time Image   (b) MOS=2.18   (c) MOS=3.25   (d) Night-Time Image   (e) MOS=3.67   (f) MOS=4.45

(g)   (h)   (i)   (j)   (k)   (l)

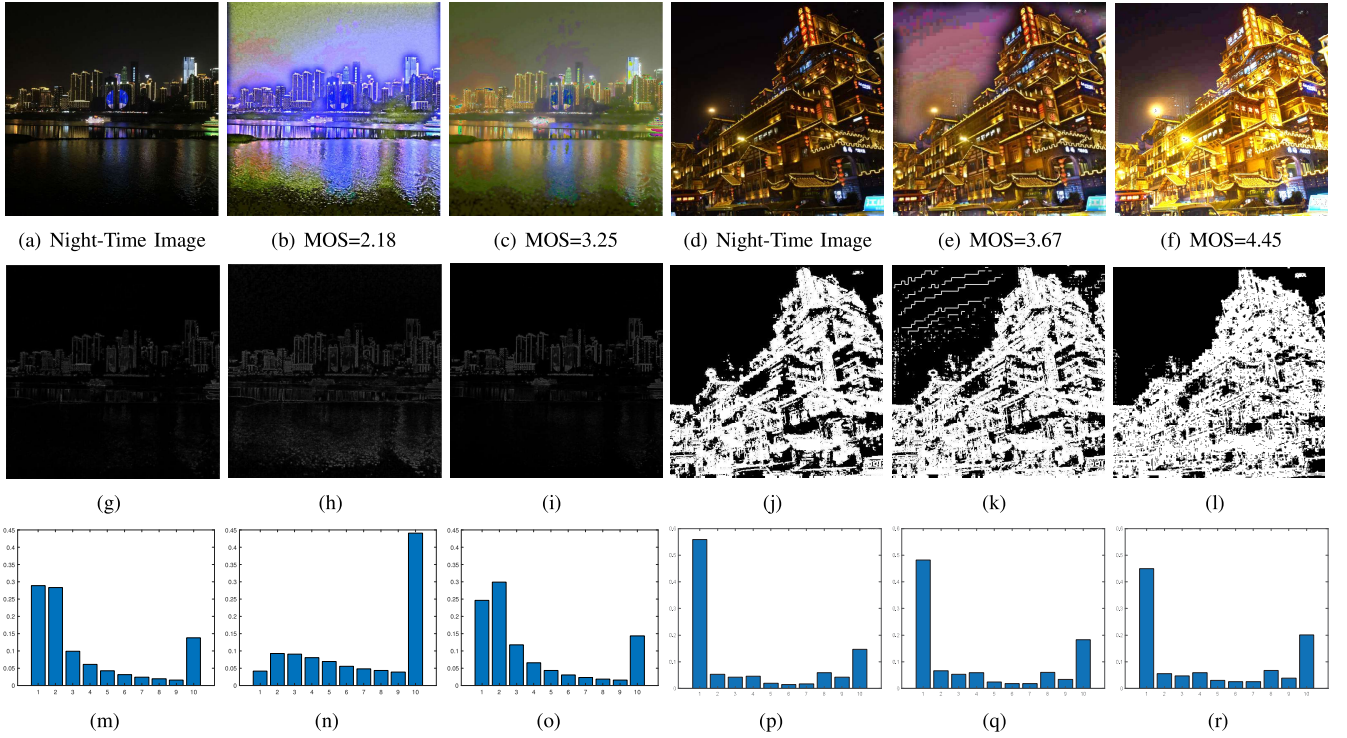(m)   (n)   (o)   (p)   (q)   (r)

Fig. 8. Two typical structural artifacts introduced by EHAs. (a) A night-time image; (b) image enhanced by [27] with intrinsic structural damage; (c) image enhanced by [34]; (d) night-time image; (e) image enhanced by [27] with over-enhancement; (f) image enhanced by citekuang2016; (g)-(i) GM maps of (a)-(c); (j)-(l) $CLBP_m$ maps of (d)-(f); and (m)-(r) histograms of (g)-(r).

where H is empirically set to 3. Finally, the spatial structural information is obtained by computing the histogram of HGM:

$$H_{gm} = hist(HGM(i,j)) \qquad (19)$$

where $hist$ is the histogram operator. In our paper, we empirically set the number of GM histogram bins to 10.

The second-order structure represents the minor structure (e.g., texture), which is also indispensable for visual perception. We employ the CLBP operator [51] to extract the second-order structure, which can be expressed as three maps: the intensity of the central pixel $CLBP_c$ and the sign $CLBP_s$ and the magnitude $CLBP_m$ of the local difference of neighboring pixels [51]. Likewise, we employ the histogram to capture the statistical information of these three maps for quality evaluation. Figs. 8(j)-8(l) show the $CLBP_m$ maps of Figs. 8(d)-8(f), and Figs. 8(p)-8(r) are the corresponding histograms of Figs. 8(d)-8(f). From these figures, we find that the $CLBP_m$ map (Fig. 8(k)) accurately portrays the over-enhancement area (Fig. 8(e)). Moreover, as the MOS value of the image increases, the histogram changes regularly (the first bin decreases while the last bin increases). Thus, it is reasonable to use the histogram to represent the statistical information of CLBP maps for over-enhancement evaluation.

Given a central pixel g, g_p ($p = 0,...,P - 1$) denotes its surrounding neighbors in a circular space. The central intensity can be expressed as

$$CLBP_c = t(g,\eta), \quad t(x,e) = \begin{cases} 1, & x \geq e \\ 0, & x < e \end{cases} \qquad (20)$$

where η denotes the average pixel value of the image. The sign of the local difference is given by

$$CLBP_s = \begin{cases} \sum_{p=0}^{P-1} t(g_p - g, 0), & \text{if } U_g \leq 2 \\ P + 1, & \text{otherwise} \end{cases} \qquad (21)$$

$$U_g = |t(g_{P-1} - g, 0) - t(g_0 - g, 0)| \\ + \sum_{p=1}^{P-1} |t(g_p - g, 0) - t(g_{p-1} - g, 0)| \qquad (22)$$

where U_g measures the discontinuity in the circular binary representation [52]. The magnitude of the local difference is given by

$$CLBP_m = \sum_{p=0}^{P-1} t(|g_p - g|, \eta_d) \cdot 2^p \qquad (23)$$

where η_d is the average value of all local differences $|g_p - g_c|$ in an image. Since $CLBP_c \in \{0, 1\}$, we use the $hist$ operator to map it into a histogram H_{CLBP_c} with 2 bins. Similarly, we map $CLBP_s$ and $CLBP_m$ into H_{CLBP_s} and H_{CLBP_m}, respectively, with $P + 2$ bins. Here, we set $P = 8$ empirically.

Finally, we combine 10 first-order and 22 second-order structural features as the structure-preserving feature vector:

$$F_s = [H_{gm}, H_{CLBP_c}, H_{CLBP_s}, H_{CLBP_m}]. \qquad (24)$$

D. Colorfulness

Since EHAs enlarge the dynamic range of original night-time images, their corresponding enhanced images usually

exhibit more abundant color information. In other words, the color information of enhanced night-time images changes. Moreover, different EHAs emphasize different regions, and depending on the complexity of the image contents, they may produce unpredictable visual effects. Thus, it is important to consider the colorfulness for quality evaluation. As shown in Fig. 8(b), image enhancement usually introduces extra unnatural colorfulness.

As the HVS perceives color information in an opponent color space [53], we first transform the RGB color space into the opponent space, which is represented as

$$O_1 = R - G \tag{25}$$

$$O_2 = (R + G)/2 - B \tag{26}$$

where O_1 and O_2 are the converted red-green and yellow-blue channels, respectively. Then we compute the following three colorfulness features. The first colorfulness feature is computed from the opponent space:

$$C_1 = log(\frac{\sigma_{O_1}^2}{|\mu_{O_1}|^{0.2}}) \cdot log(\frac{\sigma_{O_2}^2}{|\mu_{O_2}|^{0.2}}) \tag{27}$$

where μ_{O_1}, μ_{O_2}, σ_{O_1} and σ_{O_2} are the mean values and the variances of O_1 and O_2. Since LMN color space is highly consistent with the perception of HVS, we compute the second colorfulness feature according to the chrominance components (M, N) of LMN defined in Formula (3), which is defined as

$$C_2 = log(\frac{\sigma_M^2}{|\mu_M|^{0.2}}) \cdot log(\frac{\sigma_N^2}{|\mu_N|^{0.2}}) \tag{28}$$

where μ_M, μ_N, σ_M and σ_N are the mean values and the variances of M and N. The dynamic range of YCbCr color space can also capture the colorfulness feature [54]. Thus, we define the third colorfulness feature as

$$C_3^c = \frac{DR(I^c)}{\mu(I^c)} \tag{29}$$

where $DR(I^c)$ and $\mu(I^c)$ denotes the dynamic range and the global mean of the image I on channel $c \in \{Y, Cb, Cr\}$.

Combining these three colorfulness features, we obtain 5 components as the colorfulness feature vector:

$$F_c = [C_1, C_2, C_3^Y, C_3^{Cb}, C_3^{Cr}]. \tag{30}$$

E. Ensemble Model and Quality Prediction

After feature extraction, a regression model is required to learn a mapping function from the feature space to the quality score. In this paper, we present an ensemble training model AdaBoost-RF based on the AdaBoost algorithm [55] and RF [56], which has better generalization ability than a single base learner, such as SVR.

There are two critical issues in designing an ensemble model: the accuracy of the base learners and the differences between the base learners. Here, we choose RF as the base learner for two reasons. First, RF combines bagging theory and the random selection of features, making it robust to noises and outliers. Second, the decision trees in RF are independent of each other, so there is a big difference between different RFs. AdaBoost is used to ensemble the base learners, which is
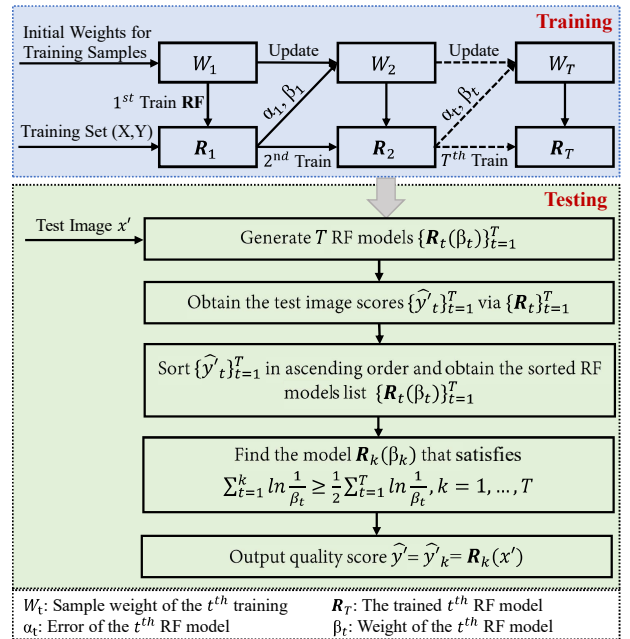

Fig. 9. The procedure of the proposed ensemble algorithm AdaBoost-RF.

achieved by changing the training data distribution and adaptively adjusting the errors of the base learners. AdaBoost-RF combines the advantages of the boosting and bagging methods, which enables RF to fully use the extracted features to improve the accuracy and robustness of the quality prediction model.

The core principle of the proposed AdaBoost-RF algorithm is that if a training sample produces a greater prediction error, it will be assigned a greater weight in the next iteration during the training procedure. Let $(X, Y) = \{(x_1, y_1), \cdots, (x_q, y_q), \cdots, (x_Q, y_Q)|q = 1, 2, \cdots, Q\}$ be the training set, where X and Y denote the extracted features and MOS values of the training data, Q is the total number of training samples. The detailed operation steps of AdaBoost-RF are described below. First, we assign an initial weight to each sample in the training set:

$$W_1 = (w_{1,1}, w_{1,2}, \cdots, w_{1,Q}), w_{1,q} = \frac{1}{Q}. \tag{31}$$

where $w_{1,q}$ denotes the initial weight of the qth sample. Then, we use the training set with initial weight to train an RF model (base learner) $\mathbf{R_1} = f_{RF}((X, Y), W_1)$, f_{RF} denotes the training operation of RF. Thereafter, we can obtain the predicted scores $\widehat{Y}$ of the training set via $\mathbf{R_1}$. Based on the error between the MOS values (Y) and predicted scores $(\widehat{Y})$ of the training samples, we update the weight of each training sample and train the next RF model. The error of each sample can be defined as:

$$e_q = \frac{|y_q - \widehat{y}_q|}{\max\{y_q - \widehat{y}_q\}_{q=1}^Q} \tag{32}$$

where $\widehat{y}_q$ is the predicted score of the qth sample. Thus, the error α and the weight β of $\mathbf{R_1}$ can be defined as

$$\alpha = \sum_{q=1}^Q w_{1,q} e_q; \beta = \frac{\alpha}{1 - \alpha}. \tag{33}$$

Finally, we can update the weight of each training sample by

$$w_{1+1,q} = w_{2,q} = \frac{w_{1,q}\beta^{1-e_q}}{\sum_{q=1}^{Q} w_{1,q}\beta^{1-e_q}}. \quad (34)$$

Based on the updated weights of the training samples, we can train the next model. Totally, we repeat the above operations T times, and we can obtain T models $\mathbf{R}_t, t = 1, \cdots, T$.

Given a test sample x', we can obtain the corresponding T predicted scores $\{\hat{y}'_t\}_{t=1}^{T}$ via the trained T RF models. Then we sort $\{\hat{y}'_t\}_{t=1}^{T}$ in ascending order and obtain the corresponding weight list of these models as $\{\beta_1, \beta_2, \cdots, \beta_T\}$. Based on the sorted RF models, we define the kth RF model $\mathbf{R}_k$ as the final model to predict the quality of the test image, which should satisfy the following condition:

$$\sum_{t=1}^{k} \ln\frac{1}{\beta_t} \geq \frac{1}{2}\sum_{t=1}^{T} \ln\frac{1}{\beta_t} \quad (35)$$

where β_t denotes the weight of the tth RF model. In this paper, the base learner RF contains 256 decision trees, the depth of each decision tree is 6, and the ensemble scale of AdaBoost-RF is $T = 24$. Fig. 9 illustrates the training and testing procedures of the proposed AdaBoost-RF method.

V. EXPERIMENTS

In this section, we conduct a performance evaluation and compare the propsoed BEHN metric with the exsiting representative IQAs on two databases.

A. Experimental Protocol

1) IQAs: In this paper, we compare our proposed BEHN with four types of IQAs. The IQAs of the first type are general-purpose IQAs, including BRISQUE [10], GM-LOG [11], IL-NIQE [57], NRSL [13] and SNP-NIQE [14]. These methods are blind IQAs and achieve promising performance on popular IQA databases. The IQAs of the second type are contrast-related IQAs, including two RR IQAs, namely, RIQMC [19] and CPCQI [18], and three blind/no-reference (NR) IQAs, namely, NR-CDIQA [37], NIQMC [17] and BIQME [18]. The IQA of the third type is a night-time IQA, i.e., BNBT [22], which is a blind IQA for natural night-time images. Moreover, to further demonstrate the practicality of the proposed subjective database, we also train two NR CNN-based IQA models (CNNIQA [58] and ResNet18-IQA [59]) on the EHNQ database for comparison. In our study, RR IQAs use enhanced images and their corresponding night-time images, and blind IQAs only use enhanced images.

2) Databases: Besides the constructed EHNQ database, there exists another relevant database, namely, the EAQA database [36]. The EAQA database consists of three subsets, and each subset contains 500 images created by enhancing underwater, haze, and low-light images using five different EHAs. In our experiments, we only consider 500 enhanced low-light images. The EAQA database only contains the performance ranks of five EHAs, specifically, the rank range is from 1 to 5, where 1 represents the best and 5 the worst. Therefore, we use these ranks to represent quality scores. If the rank of an image is 1, we set its subjective score to 1. A higher score indicates the worse visual quality of the image.
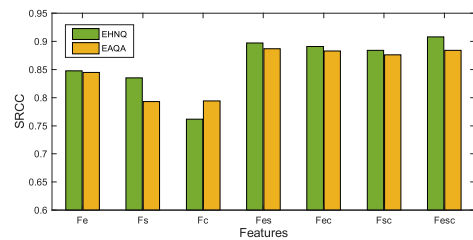

Fig. 10. SRCC values of different groups of features on the EHNQ and EAQA databases. F_e: enhancement-aware feature; F_s: structure-preserving feature; F_c: colorfulness feature; F_{es}: $F_e + F_s$; F_{ec}: $F_e + F_c$; F_{sc}: $F_s + F_c$; F_{esc}: and $F_e + F_s + F_c$.
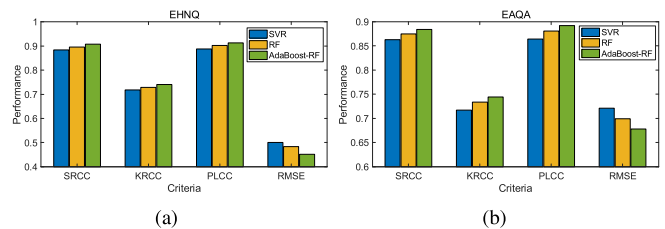

Fig. 11. Performance of different regression models. (a) EHNQ; (b) EAQA.

3) Criteria: Four representative evaluation criteria are used to measure the performance of these IQAs. Specifically, the Spearman rank-order correlation coefficient (SRCC) and Kendall's rank-order correlation coefficient (KRCC) are employed to measure the prediction monotonicity. The Pearson linear correlation coefficient (PLCC) and root-mean-squared error (RMSE) are used to measure the prediction accuracy. A better objective algorithm should obtain higher SRCC, KRCC, and PLCC values but lower RMSE, indicating a high degree of consistency between the objective and subjective scores.

B. Ablation Study

1) Effectiveness of Features: The proposed BEHN extracts three groups of features, namely, enhancement-aware features, structure-preserving features, and colorfulness. To verify the effectiveness of each group of features, we analyze their contribution to the performance of BEHN. The performance of each group of features is determined by training the model individually with the corresponding features. For instance, only the extracted enhancement-aware features are used to train the quality prediction model when measuring the contribution of the enhancement-aware features.

Fig. 10 shows how well different groups of features correlate with the perceptual quality of subjective judgment. From Fig. 10, we make the following observations. First, by using only enhancement-aware or structure-preserve features, considerable performance can be achieved, which demonstrates that recovering brightness and contrast and avoiding structural artifacts are crucial for enhancing night-time images. Second, colorfulness also contributes to the proposed BEHN, and this is not surprising since color is of great importance for visual perception. Last, each group of features contributes to the overall performance, and the combination of these three groups of features achieves the best results. This verifies the effectiveness of these three groups of features in evaluating

TABLE II
GENERALIZATION ABILITY EVALUATION

Test Database	Metric	SVR	RF	AdaBoost-RF
EAQA	SRCC	0.565	0.591	**0.682**
	PLCC	0.582	0.588	**0.673**
TID2008	SRCC	0.435	0.449	**0.512**
	PLCC	0.458	0.462	**0.496**
TID2013	SRCC	0.324	0.339	**0.418**
	PLCC	0.341	0.357	**0.434**
CSIQ	SRCC	0.347	0.361	**0.454**
	PLCC	0.365	0.386	**0.476**

the quality of enhanced night-time images. All the features perform better on the EHNQ database than on the EAQA database except the colorfulness (F_c), which is due to the differences between the EHAs used in the two databases. The EHAs in the EAQA database bring heavy color distortion, and the colorfulness features designed in our paper are conductive to capturing such color distortions.

2) Effectiveness of the Regression Model: To demonstrate the effectiveness of our proposed ensemble training algorithm AdaBoost-RF, we conduct an ablation test under different regression schemes on the two databases. In our experiments, SVR, RF, and AdaBoost-RF are separately employed to map the features into the quality score, and all the other experimental conditions remain unchanged.

Fig. 11 illustrates the performance of different regression models on the EHNQ and EAQA databases. We can find that our BEHN using RF to train the regression model outperforms BEHN using SVR in terms of all the evaluation criteria. This is because RF utilizes bagging theory and the random selection of features, resulting in better robustness to noise than SVR. Thus, we choose RF as the base learner of the proposed AdaBoost-RF algorithm. In addition, BEHN trained via AdaBoost-RF outperforms the other two models on both databases, which demonstrates the effectiveness of the proposed AdaBoost-RF in training the regression model for our task.

3) Generalization Ability: The ensemble learning model has better generalization ability than a single base learner. We conduct a cross-database test to verify the generalization ability of AdaBoost-RF. We train a model on the EHNQ database and test it on the EAQA database and the contrast enhancement-related subsets of TID2008 [60] (200 images), TID2013 [61] (160 images), and CSIQ [62] (250 images) databases. Images in the EAQA database are generated by enhancing low-light images, while images in the remaining test subsets are generated by changing the contrast of the undistorted natural images. Thus, the test datasets are mainly distorted by contrast change. Besides the contrast distortion, the enhanced images in our constructed EHNQ database are also corrupted by blur, noise, and other complex distortions created by the enhancing procedure, such as structural deformation and unnatural colorfulness. Table II presents the performance results, it is observed that the proposed AdaBoost-RF shows better generalization ability than a single regression model SVR and RF on all the test datasets.

C. Performance

1) Performance on Individual Databases: We compare BEHN with 11 state-of-the-art IQAs and 2 CNN-based IQA models (CNNIQA [58] and ResNet18 [59]) in terms of overall performance. Among them, BRISQUE, GM-LOG, NRSL, NR-CDIQA, BNBT, and BEHN are training-based methods, which require a training stage to adjust regression models. For the training-based methods, the database is randomly divided into two non-overlapping sets: a training set containing 80% of the enhanced night-time images and a testing set with the remaining 20% of the images. Enhanced images corresponding to the same night-time image are divided into the same set to avoid content overlap between the training set and the testing set. We retrain these training-based methods on the training set and test them on the testing set. The training-testing process is repeated 1,000 times, and the median performance criteria are reported. We test the remaining methods without the training procedure on the same testing set for a fair comparison.

Table III shows the comparison results on the EHNQ and EAQA databases. We make the following observations. First, BEHN achieves better performance on the two databases. Second, the general-purpose training-based IQAs (BRISQUE, GM-LOG, and NRSL) outperforms the non-training IQAs (IL-NIQE and SNP-NIQE). This is understandable because the training-based models are conducive to capturing the distortions of enhanced night-time images. Third, although BIQME and NIQMC are designed for contrast-enhanced images, their performance is not impressive. This is because, in addition to contrast distortion, the enhanced night-time images still suffer from more complicated corruptions, such as structural deformation and unnatural colorfulness. Last, the results of the two CNN-based IQA models demonstrate the practicality of the constructed subjective database and the effectiveness of the proposed objective quality evaluation methods.

2) Performance on Individual EHAs: In addition to the overall performance on each database, to more comprehensively assess the performance of all involved methods, we also conduct comparison experiments on individual EHAs. The result for each EHA is obtained using the model trained with all the enhanced images on the EHNQ or EAQA database. For brevity, we only present the SRCC value here, and similar results are obtained for KRCC, PLCC, and RMSE.

Table IV reports the SRCC values of different EHAs on the EHNQ and EAQA databases. First, none of these methods always performs the best on all EHAs. Second, the proposed BEHN achieves much better and more robust SRCC performance across different EHAs than other IQAs on the EHNQ database. Specifically, BEHN obtains the highest SRCC values on all CEHAs. For LEHAs and NEHAs, our proposed BEHN occupies the top place 9 times. In contrast, NIQMC, GM-LOG, and BNBT exhibit significant changes on different EHAs. In addition, the performance of BEHN on the EAQA database stands out from those of the other methods. These results indicate the effectiveness and stability of BEHN in evaluating the quality of enhanced night-time images.

3) Running Time Measurement: A good IQA metric should simultaneous achieve efficacy and efficiency. Since the running

TABLE III
OVERALL PERFORMANCE OF ALL INVOLVED METHODS ON THE EHNQ AND EAQA DATABASES

Method	Type	Subset-512 of EHNQ				Subset-1024 of EHNQ				EHNQ				EAQA			
		SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
RIQMC [19]	RR	0.428	0.301	0.539	0.791	0.592	0.423	0.669	0.789	0.504	0.355	0.597	0.806	0.396	0.285	0.429	1.278
CPCQI [18]	RR	0.184	0.122	0.315	0.891	0.250	0.167	0.345	0.996	0.217	0.145	0.323	0.951	0.207	0.142	0.243	1.372
IL-NIQE [57]	NR	0.575	0.399	0.587	0.760	0.469	0.324	0.512	0.912	0.518	0.360	0.542	0.844	0.499	0.370	0.540	1.190
SNP-NIQE [14]	NR	0.211	0.137	0.280	0.901	0.152	0.100	0.214	1.037	0.181	0.118	0.239	0.975	0.072	0.053	0.075	1.410
NIQMC [17]	NR	0.591	0.422	0.659	0.706	0.607	0.429	0.665	0.792	0.597	0.423	0.660	0.755	0.453	0.330	0.594	1.137
BIQME [18]	NR	0.698	0.511	0.755	0.615	0.729	0.546	0.775	0.670	0.715	0.530	0.767	0.645	0.533	0.399	0.560	1.172
BRISQUE [10]	NR	0.659	0.471	0.681	0.715	0.640	0.457	0.641	0.630	0.640	0.447	0.655	0.754	0.747	0.592	0.775	1.278
GM-LOG [11]	NR	0.641	0.456	0.663	0.689	0.658	0.468	0.689	0.760	0.725	0.530	0.739	0.669	0.645	0.495	0.686	1.410
NR-CDIQA [37]	NR	0.508	0.353	0.579	0.761	0.464	0.320	0.536	0.876	0.518	0.359	0.580	0.821	0.792	0.676	0.809	1.198
NRSL [13]	NR	0.744	0.549	0.768	0.597	0.693	0.499	0.739	0.715	0.721	0.530	0.728	0.855	0.831	0.681	0.860	1.025
BNBT [22]	NR	0.825	0.629	0.821	0.508	0.823	0.627	0.818	0.507	0.817	0.665	0.823	0.519	0.803	0.639	0.863	1.130
CNNIQA [58]	NR	0.834	0.671	0.841	0.497	0.827	0.665	0.830	0.501	0.828	0.692	0.848	0.489	0.841	0.719	0.836	0.797
ResNet18-IQA [59]	NR	0.849	0.687	0.862	0.475	0.846	0.698	0.858	0.483	0.852	0.715	0.856	0.480	0.856	0.727	0.845	0.742
BEHN	NR	**0.887**	**0.711**	**0.894**	**0.441**	**0.877**	**0.698**	**0.897**	**0.428**	**0.908**	**0.741**	**0.913**	**0.452**	**0.884**	**0.744**	**0.892**	**0.678**

TABLE IV
SRCC VALUES OF DIFFERENT EHAs ON THE EHNQ AND EAQA DATABASES

Method	Type	EHNQ															EAQA				
		CEHAs			LEHAs							NEHAs					EHAs of Low-Light Subset				
		LDR	FSBE	HE	BCP	NPE	REIA	PEM	FWIE	LIME	WVEM	IMSRE	PNIE	CNEFS	RGF	IMSR	HE	CHA	CGC	ELLV	CNEFS
RIQMC [19]	RR	0.501	0.124	0.271	0.188	0.264	0.217	0.033	0.149	0.324	0.116	0.325	0.355	0.340	0.033	0.405	0.100	0.017	0.326	0.069	0.121
CPCQI [18]	RR	0.247	0.480	0.468	0.650	0.230	0.315	0.140	0.213	0.456	0.164	0.259	0.103	0.289	0.133	0.350	0.261	0.109	0.324	0.063	0.025
IL-NIQE [57]	NR	0.466	0.479	0.265	0.103	0.312	0.228	0.317	0.284	0.069	0.235	0.428	0.179	0.336	0.228	0.060	0.206	0.042	0.279	0.135	0.074
SNP-NIQE [14]	NR	0.375	0.300	0.089	0.205	0.089	0.006	0.172	0.056	0.077	0.187	0.451	0.202	0.591	0.289	0.124	0.367	0.126	0.141	0.187	0.267
NIQMC [17]	NR	0.544	0.275	0.025	0.258	0.178	0.245	0.130	0.107	0.314	0.302	0.279	0.333	0.730	0.135	0.370	0.220	0.228	0.123	0.337	0.155
BIQME [18]	NR	0.710	0.315	0.154	0.320	0.111	0.370	0.355	0.365	0.454	0.385	0.293	0.483	0.800	0.319	0.506	0.233	0.182	0.121	0.243	0.230
BRISQUE [10]	NR	0.749	0.649	0.685	0.745	0.574	0.609	0.686	0.703	0.737	0.640	0.557	0.694	0.745	0.650	0.553	0.748	0.463	0.672	0.802	0.603
GM-LOG [11]	NR	0.736	0.622	0.562	0.811	0.730	0.538	0.677	0.741	0.842	0.661	0.452	0.778	0.758	**0.846**	0.781	0.619	0.358	0.570	0.601	0.621
NR-CDIQA [37]	NR	0.612	0.618	0.526	0.652	0.696	0.682	0.624	0.633	0.619	0.517	0.544	0.708	0.760	0.588	0.778	0.611	0.577	0.631	0.597	0.620
NRSL [13]	NR	0.881	0.874	0.803	0.774	**0.783**	0.819	0.780	0.783	0.811	**0.866**	0.824	0.789	0.851	0.652	0.817	0.619	0.595	0.564	0.785	0.827
BNBT [22]	NR	0.813	0.744	0.703	0.826	0.676	0.795	0.714	0.734	0.775	0.836	0.580	0.597	0.850	0.769	0.784	0.564	0.491	0.555	0.708	0.658
CNNIQA [58]	NR	0.854	0.806	0.731	0.796	0.743	0.766	0.702	0.779	0.823	0.854	0.646	0.683	0.820	0.775	0.807	0.681	0.504	0.613	0.772	0.825
ResNet18-IQA [59]	NR	0.873	0.845	0.792	0.851	0.654	0.804	0.699	0.773	0.848	0.812	0.769	0.791	0.847	0.755	0.816	0.723	**0.618**	0.586	0.748	0.805
BEHN	NR	**0.922**	**0.890**	**0.882**	**0.891**	0.715	**0.824**	**0.796**	**0.802**	**0.853**	0.838	**0.839**	**0.812**	**0.855**	0.797	**0.831**	**0.747**	0.533	**0.664**	**0.795**	**0.843**

TABLE V
COMPUTATIONAL COST

Non-Learning IQAs	RIQMC [19]	CPCQI [18]	IL-NIQE [57]	SNP-NIQE [14]	NIQMC [17]	BIQME [18]	BRISQUE [10]
Time (sec./image)	0.721	0.048	3.473	1.822	1.174	0.480	0.057
Learning-Based IQAs	GM-LOG [11]	NR-CDIQA [37]	NRSL [13]	BNBT [22]	CNNIQA [58]	ResNet18-IQA [59]	BEHN
Time (sec./image)	0.053	0.036	0.086	1.882	**0.011**	0.027	0.424

time is crucial in real-time applications, we test the computational complexity of our BEHN and all compared IQAs, and we report the results in Table V. One hundred enhanced night-time images with a fixed resolution of 512×512 are selected from the EHNQ database, and the average running time of feature extraction and quality prediction is regarded as the computation cost. Experiments are conducted on a personal computer equipped with a 3.20GHz Intel Core i5 CPU and 4 GB RAM. The development platform is MATLAB R2014b running on Windows 10 Home Premium.

From Table V, we can make the following four observations. First, the non-learning methods IL-NIQE, SNP-NIQE, and NIQMC require relatively long computing times, since both the feature extraction and quality regression of these methods require complex calculations. Second, the learning-based methods BRISQUE, GM-LOG, NR-CDIQA, and NRSL, have lower computational overhead. The reason is that they extract features using statistical histograms, and they are easy to compute. The computational cost of the CNN-based models depends on the network depth. Since CNNIQA has 2 convolutional layers while ResNet18-IQA has 18 convolutional layers,

the runing time of CNNIQA is lower than that of ResNet18-IQA. Last, the proposed BEHN has a low computational cost, though not the lowest among all compared methods.

VI. CONCLUSION

In this paper, we conduct subjective and objective quality evaluations of enhanced night-time images. First, we construct a large-scale enhanced night-time image quality database, namely, the EHNQ database, with subjective quality scores rated by human observers, which can be employed as the ground truth when conducting an objective quality evaluation of enhanced night-time images. Second, we present BEHN, a blind quality evaluation index for enhanced night-time images, by analyzing three groups of features: enhancement-aware features, structure-preserving features, and colorfulness. After extracting these visual features, we propose an ensemble training algorithm, AdaBoost-RF, for mapping the extracted features into a quality score using AdaBoost and RF. Extensive experiments are conducted on the constructed EHNQ database and the EAQA database to evaluate and analyze the performance of our proposed BEHN. The results show that

the proposed method outperforms the state-of-the-art IQAs in terms of both efficacy and efficiency.

References

[1] X. Fu, Y. Liao, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[2] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2016.

[3] H. Kuang, L. Chen, F. Gu, J. Chen, L. Chan, and H. Yan, "Combining region-of-interest extraction and image enhancement for nighttime vehicle detection," *IEEE Intelligent Systems*, vol. 31, no. 3, pp. 57–65, 2016.

[4] Z. Shi, M. Zhu, B. Guo, and M. Zhao, "A photographic negative imaging inspired method for low illumination night-time image enhancement," *Multimedia Tools and Applications*, vol. 76, no. 13, pp. 15 027–15 048, 2017.

[5] S. Hao, X. Han, Y. Guo, and M. Wang, "Decoupled low-light image enhancement," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 4, pp. 1–19, 2022.

[6] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 9, pp. 1480–1494, 2014.

[7] S.-Y. Yu and H. Zhu, "Low-illumination image enhancement algorithm based on a physical lighting model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 28–37, 2019.

[8] R. Kumar and A. K. Bhandari, "Fuzzified contrast enhancement for nearly invisible images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2802–2813, 2021.

[9] K. Xu, H. Chen, C. Xu, Y. Jin, and C. Zhu, "Structure-texture aware network for low-light image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[10] 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.

[11] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.

[12] C. Yan, T. Teng, Y. Liu, Y. Zhang, H. Wang, and X. Ji, "Precise no-reference image quality evaluation based on distortion identification," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 3s, pp. 1–21, 2021.

[13] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016.

[14] 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.

[15] Y. Liu, K. Gu, X. Li, and Y. Zhang, "Blind image quality assessment by natural scene statistics and perceptual characteristics," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 3, pp. 1–91, 2020.

[16] W. Xia, Y. Yang, J.-H. Xue, and J. Xiao, "Domain fingerprints for no-reference image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1332–1341, 2021.

[17] 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 Transactions on Cybernetics*, vol. 47, no. 12, pp. 4559–4565, 2016.

[18] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 1301–1313, 2018.

[19] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 284–297, 2016.

[20] G. Zhai, W. Sun, X. Min, and J. Zhou, "Perceptual quality assessment of low-light image enhancement," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 4, pp. 1–24, 2021.

[21] M. H. Khosravi and H. Hassanpour, "Blind quality metric for contrast-distorted images based on eigendecomposition of color histograms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 48–58, 2020.

[22] T. Xiang, Y. Yang, and S. Guo, "Blind night-time image quality assessment: subjective and objective approaches," *IEEE Transactions on Multimedia*, vol. 22, no. 5, pp. 1259–1272, 2020.

[23] R. ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2012.

[24] J. S. Lim, "Two-dimensional signal and image processing," *Englewood Cliffs, NJ, Prentice Hall*, 1990.

[25] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[26] Q. Wang, X. Fu, X.-P. Zhang, and X. Ding, "A fusion-based method for single backlit image enhancement," in *Proceedings of ICIP*, 2016, pp. 4077–4081.

[27] X. Fu, D. Zeng, Y. Huang, X. Ding, and X.-P. Zhang, "A variational framework for single low light image enhancement using bright channel prior," in *Proceedings of GlobalSIP*, 2013, pp. 1085–1088.

[28] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[29] X. Fu, Y. Sun, M. LiWang, Y. Huang, X.-P. Zhang, and X. Ding, "A novel retinex based approach for image enhancement with illumination adjustment," in *Proceedings of ICASSP*, 2014, pp. 1190–1194.

[30] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of CVPR*, 2016, pp. 2782–2790.

[31] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[32] L. Cai and J. Qian, "Night color image enhancement using fuzzy set," in *Proceedings of CISP*, 2009, pp. 1–4.

[33] H. Lin and Z. Shi, "Multi-scale retinex improvement for nighttime image enhancement," *Optik*, vol. 125, no. 24, pp. 7143–7148, 2014.

[34] J. Wei, Q. Zhijie, X. Bo, and Z. Dean, "A nighttime image enhancement method based on retinex and guided filter for object recognition of apple harvesting robot," *International Journal of Advanced Robotic Systems*, vol. 15, no. 1, pp. 1–12, 2018.

[35] X. Dong, G. Wang, Y. Pang, W. Li, J. Wen, W. Meng, and Y. Lu, "Fast efficient algorithm for enhancement of low lighting video," in *Proceedings of ICME*, 2011, pp. 1–6.

[36] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proceedings of CVPR*, 2014, pp. 3003–3010.

[37] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2015.

[38] 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, 2018.

[39] 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 Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3883–3897, 2020.

[40] R. ITU-T P, "Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union*, 1999.

[41] A. M. Van Dijk, J.-B. Martens, and A. B. Watson, "Quality asessment of coded images using numerical category scaling," in *Advanced Image and Video Communications and Storage Technologies*, vol. 2451, 1995, pp. 90–101.

[42] 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.

[43] J.-M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338–1350, 2001.

[44] C. Li and T. Chen, "Aesthetic visual quality assessment of paintings," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 236–252, 2009.

[45] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," *Nature*, vol. 447, no. 7141, pp. 206–209, 2007.

[46] Y. Zhou, L. Li, J. Wu, K. Gu, W. Dong, and G. Shi, "Blind quality index for multiply distorted images using biorder structure degradation and nonlocal statistics," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3019–3032, 2018.

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

[48] C. Chen, H. Zhao, H. Yang, T. Yu, C. Peng, and H. Qin, "Full-reference screen content image quality assessment by fusing multilevel structure similarity," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 3, pp. 1–21, 2021.

[49] J. Larsson, M. S. Landy, and D. J. Heeger, "Orientation-selective adaptation to first- and second-order patterns in human visual cortex," *Journal of Neurophysiology*, vol. 95, no. 2, pp. 862–881, 2006.

[50] T. Xiang, Y. Yang, H. Liu, and S. Guo, "Visual security evaluation of perceptually encrypted images based on image importance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4129–4142, 2020.

[51] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[52] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[53] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness, and structure," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3784–3793, 2018.

[54] T.-J. Liu and K.-H. Liu, "No-reference image quality assessment by wide-perceptual-domain scorer ensemble method," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1138–1151, 2017.

[55] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[56] T. Windeatt, "Accuracy/diversity and ensemble MLP classifier design," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1194–1211, 2006.

[57] 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.

[58] 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.

[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of CVPR*, 2016, pp. 770–778.

[60] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[61] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proceedings of EUVIP*, 2013, pp. 106–111.

[62] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006–1–011 006–21, 2010.

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 90 papers on international journals and conferences. He also served as a referee for numerous international journals and conferences.

Shangwei Guo received the PhD degree in computer science from Chongqing University, China, in 2017. He worked at Hong Kong Baptist University as a postdoctoral research fellow from 2018-2019. He is currently a postdoctoral research fellow at Nanyang Technological University, Singapore. His research interests include multimedia security, cloud security and data privacy.
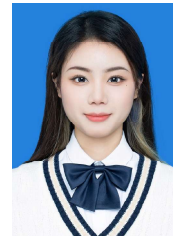
Xiao Lv received the B.Sc. degree in computer science from Chongqing University, China, in 2020, where she is currently pursuing the Ph.D. degree with the College of Computer Science. Her current research interests include image quality assessment and multimedia security.

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.

Xiaofeng Liao (Fellow, IEEE) received the BSc and MS degrees in mathematics from Sichuan University, Chengdu, China, in 1986 and 1992, respectively, and the PhD degree in circuits and systems from the University of Electronic Science and Technology of China, Chengdu, in 1997. From 1999 to 2012, he was a Professor with Chongqing University, Chongqing, China. From July 2012 to July 2018, he was a Professor and the Dean of the College of Electronic and Information Engineering, Southwest University, Chongqing. He is currently a Professor and the Dean of the College of Computer Science, Chongqing University. He is also a Yangtze River Scholar of the Ministry of Education of China, Beijing, China. From November 1997 to April 1998, he was a Research Associate with the Chinese University of Hong Kong, Hong Kong. From October 1999 to October 2000, he was a Research Associate with the City University of Hong Kong, Hong Kong. From March 2001 to June 2001 and March 2002 to June 2002, he was a Senior Research Associate at the City University of Hong Kong. From March 2006 to April 2007, he was a Research Fellow at the City University of Hong Kong. He holds five patents, and published four books and over 300 international journal and conference papers. His current research interests include optimization and control, machine learning, neural networks, bifurcation and chaos, and cryptography. Prof. Liao currently serves as an Associate Editor of the IEEE Transactions on Cybernetics and IEEE Transactions on Neural Networks and Learning Systems.

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.