

Deep Decomposition and Bilinear Pooling Network for Blind Night-Time Image Quality Evaluation

Qiuping Jiang, Jiawu Xu, Wei Zhou, Xiongkuo Min, Guangtao Zhai

arXiv:2205.05880v1 [cs.MM] 12 May 2022

Abstract—Blind image quality assessment (BIQA), which aims to accurately predict the image quality without any pristine reference information, has been highly concerned in the past decades. Especially, with the help of deep neural networks, great progress has been achieved so far. However, it remains less investigated on BIQA for night-time images (NTIs) which usually suffer from complicated authentic distortions such as reduced visibility, low contrast, additive noises, and color distortions. These diverse authentic degradations particularly challenges the design of effective deep neural network for blind NTI quality evaluation (NTIQE). In this paper, we propose a novel deep decomposition and bilinear pooling network (DDB-Net) to better address this issue. The DDB-Net contains three modules, i.e., an image decomposition module, a feature encoding module, and a bilinear pooling module. The image decomposition module is inspired by the Retinex theory and involves decoupling the input NTI into an illumination layer component responsible for illumination information and a reflectance layer component responsible for content information. Then, the feature encoding module involves learning multi-scale feature representations of degradations that are rooted in the two decoupled components separately. Finally, by modeling illumination-related and content-related degradations as two-factor variations, the two multi-scale feature sets are bilinearly pooled and concatenated together to form a unified representation for quality prediction. The superiority of the proposed DDB-Net is well validated by extensive experiments on two publicly available night-time image databases.

Index Terms—Night-time image, image quality assessment, blind/no-reference, Retinex decomposition.

I. INTRODUCTION

Due to the poor lighting condition in night-time, the captured night-time images (NTIs) are usually perceived with poor visibility and low visual quality. Given that high-quality NTIs are crucial for consumer photography and practical applications such as automated driving systems, many NTI quality/visibility enhancement algorithms have been proposed. However, the research efforts on designing objective quality metrics that can automatically quantify the visual quality of NTIs and compare the performance of different NTI enhancement algorithms remain limited, which hereby hinders the development of this field. Generally, objective image quality assessment (IQA) methods can be roughly divided into three categories, i.e., full-reference (FR), no-reference (NR),

Q. Jiang and J. Xu are with the School of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jiangqiuping@nbu.edu.cn).

W. Zhou is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada (e-mail: wei.zhou@uwaterloo.ca).

X. Min and G. Zhai are with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (minxiongkuo, zhaiguangtao@sjtu.edu.cn).

and reduced-reference (RR) [1]. Among them, FR and RR IQA methods require full and partial reference information, respectively. However, for the NTIs we concerned, there is usually no available pristine image to provide any reference information. Therefore, NR-IQA is more valuable for NTIs in this regard.

Early studies on NR-IQA mainly focus on specific distortion types, i.e., assuming that a particular distortion type is known and then specific distortion-related features are extracted to predict image quality [2–6]. Obviously, such the specificity limits their applications like the real-world night-time scenario. Although the rapid advances in the IQA community during the last decade push to produce general-purpose blind IQA (BIQA) methods [7–24] that can simultaneously work with a number of distortion types, their efficacies are still limited to synthetic distortions. This is evident by the fact that they usually validate their performance on legacy synthetic distortion benchmark databases where the distorted images are simulated from pristine corpus in laboratory. As a result, the existing general-purpose BIQA methods still cannot work well with the authentically distorted images like the NTIs captured in the real-world night-time scenario. Recently, inspired by the success of deep neural networks in many image processing and computer vision tasks, great progresses have also been achieved on deep learning-based BIQA. However, it remains less investigated on deep learning-based BIQA for NTIs which usually suffer from complicated authentic distortions such as reduced visibility, low contrast, additive noises, invisible details, and color distortions. The diverse authentic degradations in NTIs pose great challenges to the design of highly effective end-to-end deep network architectures for blind NTI quality evaluation (NTIQE).

To evaluate the visual quality of NTIs, Xiang et al. [25] first established a dedicated large-scale natural NTI database (NNID), which contains 2,240 NTIs with 448 different image contents captured by three different photographic equipments in real-world scenarios along with their corresponding subjective quality scores (obtained by conducting human subjective experiments). Then, a NR quality metric called BNBT is proposed by considering both brightness and texture features. The experimental results on NNID database have demonstrated an acceptable performance of BNBT, i.e., the predicted quality scores by BNBT are consistent with ground truth subjective quality scores. Despite its effectiveness, BNBT requires elaborately-designed handcrafted features, which enlightens us to adopt an end-to-end data-driven method by taking the advantage of deep learning.

However, designing tailored end-to-end deep neural net-

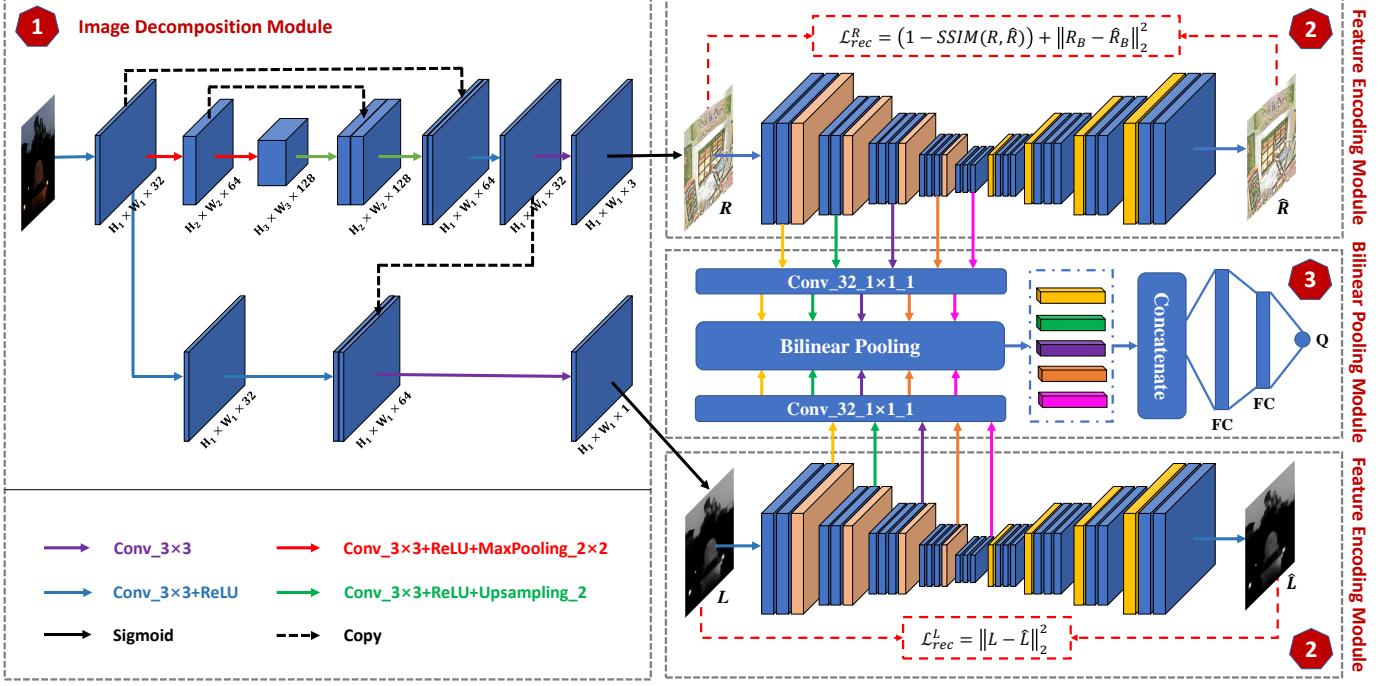


Fig. 1. The proposed deep decomposition and bilinear pooling network (DDB-Net) for blind NTIQE. It contains an image decomposition module, a feature encoding module, and a bilinear pooling module. The image decomposition module takes an NTI as input and decouples it into two layer components, i.e., illumination (L) and reflectance (R). Then, the feature encoding module involves learning multi-scale feature representations of degradations that are rooted in the illumination and reflectance separately. Finally, the two multi-scale feature sets are bilinearly pooled and concatenated together to form a unified representation for quality prediction.

works for blind NTIQE is non-trivial due to the diverse authentic degradations. The main challenge is that heterogeneous degradations in NTIs make it difficult to learn a unified mapping from input NTI to quality score. An important observation is that the commonly-encountered degradations in NTIs can have impacts on either illumination perception or content perception. For example, the color distortion and additive noise are only influential in content perception while the reduced visibility and low contrast are only influential in illumination perception. Thus, it is intuitive to consider decomposing the input NTI into two independent components with each component accounting for illumination information and content information, respectively. Assisted by such a tailored image decomposition process, the degradation features related to illumination perception and content perception can be better learned and then fused to facilitate blind NTIQE.

In this paper, we propose a novel deep decomposition and bilinear pooling network (DDB-Net) for blind NTIQE to better address the above issues. As shown in Fig. 1, our DDB-Net contains three modules namely image decomposition module, feature encoding module, and bilinear pooling module. Inspired by the Retinex theory [26], the image decomposition module involves decoupling the input NTI into two layer components, i.e., one layer component (illumination) is responsible for illumination information, while the other one (reflectance) for content information. Then, the feature encoding module involves learning multi-scale feature representations of degradations that are rooted in the two decoupled components separately. Finally, by modeling illumination-related

and content-related degradations as two-factor variations, the two multi-scale feature sets are bilinearly pooled and concatenated together to form a unified representation for quality prediction. Extensive experiments conducted on two publicly available night-time image databases well demonstrated the superiority of the proposed DDB-Net against state-of-the-art BIQA methods. In summary, this paper presents the following contributions:

1) We make the first attempt to perform Retinex decomposition to facilitate NTI QE by decoupling the input NTI into two independent layer components (i.e., illumination and reflectance) with each component accounting for illumination information and content information, respectively.

2) We introduce a self-reconstruction-based feature encoding module and design tailored loss functions to regularize the training process towards learning multi-scale illumination-related and content-related feature representations from the two decoupled components separately.

3) We model the illumination-related and content-related degradations as two-factor variations and perform bilinear pooling to fuse the two multi-scale feature sets into a unified representation for quality prediction of NTIs.

The rest of this paper is organized in the following manner. Section II introduces the related works. Sections III illustrates the proposed method with details. Section IV presents the experimental results. Section V concludes the paper.

II. RELATED WORKS

In this section, we will review the existing related works, including traditional blind image quality assessment, deep

learning-based blind image quality assessment, and blind image quality assessment in poor conditions.

A. Traditional Blind Image Quality Assessment

In the literature of traditional blind image quality assessment, natural scene statistics (NSS) and human visual system (HVS) are two main cues for designing objective BIQA models. As for NSS-based frameworks, Moorthy et al. [10] proposed the Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) index to evaluate perceptual image quality in a no-reference manner, which is composed of distortion identification and NSS-based quality regression. Likewise, the CurveletQA [9] also performs within a two-stage framework containing distortion classification and quality assessment. Different from DIIVINE, the quality assessment of CurveletQA is based on NSS features in curvelet domain. Except for the two-stage frameworks, other NSS-based BIQA algorithms have been developed. For example, Mittal et al. [8] presented the blind/referenceless image spatial quality evaluator (BRISQUE), which is operated in the spatial domain. Moreover, the BLIINDS-II [7] was proposed by exploiting the NSS model of discrete cosine transform (DCT) coefficients. In addition, some opinion-unaware BIQA methods based on NSS, i.e. so-called “completely blind” models, such as Natural Image Quality Evaluator (NIQE) [18] and ILNIQE [17], have shown competitive performance with the help of massive natural images.

Beyond the NSS features, many other statistical factors have been considered by researchers. In the family of two-stage framework, Liu et al. [15] proposed the Spatial-Spectral Entropy-based Quality (SSEQ) index, where local spatial and spectral entropy features are used to predict perceptual image quality. The GM-LOG method [13] extracts the joint statistics of local contrast features to assess image quality, including gradient magnitude and Laplacian of Gaussian response. Apart from statistical structural features, luminance histogram is used in the NRSL model [11]. The NSS features are combined with contrast, sharpness, brightness and colorfulness, together forming the BIQME framework [16].

For the HVS-based objective BIQA methods, some HVS-inspired features are applied to estimate perceptual image quality. Among these methods, Gu et al. [12] proposed the No-reference Free Energy-based Robust Metric (NFERM) on the basis of free energy principle. In [19], Li et al. used contrast masking to design the BIQA model based on structural degradation. Besides, according to the similar concept regarding the HVS properties, they proposed the GWH-GLBP by computing the gradient-weighted histogram of local binary pattern [14].

However, the above-mentioned conventional BIQA methods generally need to design elaborate handcrafted features with the pre-defined NSS or HVS mechanisms. Thus, resorting to data-driven methods based on deep learning is a promising alternative.

B. Deep Learning-based Blind Image Quality Assessment

Recently, deep learning has achieved great success in the field of blind image quality assessment. These methods can

be typically divided into two categories, consisting of those using pre-trained deep features and end-to-end learning ones. For the first category, Wu et al. [27] proposed the HFD-BIQA that integrates deep semantic features from ResNet [28] into local structure features. Moreover, a Network in Network (NIN) model [29] pre-trained on ImageNet [30] was utilized to make image quality prediction. For the second category, Kang et al. [31] proposed a relatively shallow convolutional neural network (CNN) structure for BIQA. Each patch is assigned a subjective quality of the corresponding image as the ground-truth targets for training. In this way, the visual quality of whole image is then calculated by averaging predicted patch quality values. Furthermore, a BIQA model was developed based on shearlet transform and stacked auto-encoders [32]. In [33], the RankIQA was designed to synthesize masses of ranked images for training a Siamese network. Ma et al. [34] proposed an end-to-end optimized deep neural Network for BIQA. Additionally, Bosse et al. [35] presented the end-to-end WaDIQaM that can blindly learn perceptual image quality. The deep bilinear convolutional neural network called DBCNN was proposed to bilinearly pool the feature representations to a single quality score [36].

Although the deep learning-based BIQA models can deliver good performance, they are not suitable for evaluating the perceptual quality of NTIs. This is mainly because these models usually neglect the specific characteristics of NTIs, e.g. reduced visibility, low contrast, additive noises, invisible details, and color distortions.

C. Quality Assessment for Images in Poor Conditions

In real-world applications, people may encounter many kinds of poor imaging environments, e.g. hazy, rainy, underwater, and so on. In such poor conditions, capturing images with high-quality is quite challenging and thus addressing the blind quality assessment issue is urgently needed.

In the quality evaluation of hazy images, Min et al. [37] proposed the haze-removing features, structure-preserving features, and over-enhancement features to construct the objective quality assessment index. They also used synthetic hazy images to build an effective quality assessment model for image dehazing [38]. For image deraining, an efficient objective quality assessment model to predict the human perception towards derained images was developed, which belongs to a bi-directional gated fusion network [39]. They further extended it to a bi-directional feature embedding network to further advance the performance [40]. As for underwater image quality evaluation, Yang et al. [41] proposed to linearly combine chroma, saturation and contrast factors for quantifying the perceptual quality of underwater images. In [42], colorfulness, sharpness and contrast measures were fused to predict the underwater image quality.

It should be noted that the degradation characteristics of NTIs are different from those of hazy, rainy and underwater images. Thus, these quality evaluation methods developed for hazy, rainy and underwater images cannot achieve good performance on NTIs. In this paper, we focus on designing efficient end-to-end deep network architectures for blind NTI QE.

III. PROPOSED DDB-NET FOR BLIND NTIQE

In this section, we first describe the architecture of the image decomposition module. We then introduce the self-reconstruction-based feature encoding module for hierarchical illumination-related and content-related feature learning. Finally, we present the bilinear pooling module for fusing the two hierarchical feature sets.

A. Image Decomposition Module

According to the Retinex theory [26], a single image I can be considered as a composition of two independent layer components, i.e., reflectance R and illumination L , in the fashion of $I = R \otimes L$, where \otimes denotes element-wise product. However, recovering two independent components from one single input is a typical ill-posed problem. Here, we resort to deep neural network to achieve this goal. In what follows, we first describe the detailed architecture of our image decomposition module and then present how to train it in advance. That is, the image decomposition module is pre-trained and kept fixed during the training of DDB-Net.

The architecture of image decomposition module is shown in the upper left of Fig. 1. It contains two streams corresponding to the reflectance (R) and illumination (L), respectively. The reflectance stream adopts a typical 5-layer U-Net, followed by two convolutional (conv) layers and a Sigmoid layer in the end, while the illumination stream is composed of two conv+ReLU layers and a conv layer on concatenated feature maps from the reflectance branch, finally followed by a Sigmoid layer in the end.

Since no/few ground-truth reflectance and illumination maps for real images are available, designing a well-defined *non-reference* loss function is the key to the success for training a robust deep Retinex decomposition network. Keeping in mind that different shots of a certain scene should share the same reflectance. Furthermore, while the illumination maps, though could be intensively varied, are of simple and mutually consistent structure. This inspires us to take a pair of images (describing the same scene) as input and impose both reflectance and illumination constraints between the image pair to train the image decomposition module. Specifically, during the training stage, the input to image decomposition module is an image pair of the same scene with different light/exposure configurations, as denoted by $[I_l, I_h]$. Similarly, the decomposed reflectance and illumination components are denoted by $[R_l, R_h]$ and $[L_l, L_h]$, respectively. The training of the proposed image decomposition module is guided by hybrid loss terms which are to be detailed subsequently.

Inter-consistency loss: The inter-consistency loss includes reflectance consistency loss and illumination mutual consistency loss. First, the reflectance consistency loss \mathcal{L}_{con}^R encourages the reflectance similarity, which is defined as follows:

$$\mathcal{L}_{con}^R = \|R_l - R_h\|_1, \quad (1)$$

where $\|\cdot\|_1$ means the ℓ_1 norm. Second, the illumination mutual consistency loss \mathcal{L}_{con}^L is defined as follows:

$$\mathcal{L}_{con}^L = f(M) = \left\| \frac{M}{c^2} \otimes \exp \left(-\frac{M^2}{2c^2} \right) \right\|_1, \quad (2)$$

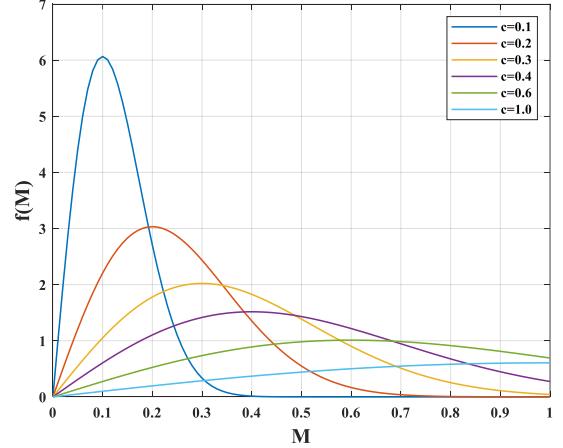


Fig. 2. Penalty curves with different values of c .

$$M = |\nabla L_l| + |\nabla L_h|, \quad (3)$$

where ∇ means the first order derivative operator along both horizontal and vertical directions, c is a parameter controlling the shape of the above penalty curve. To facilitate understanding, we draw the penalty curves with different values of c in Fig. 2. As we can see, the penalty value first increases and then decreases to zero as M increases. In our implementation, we set $c = 0.1$ empirically. By minimizing such an illumination mutual consistency loss, the mutual strong edges are encouraged to be well preserved and all weak edges are to be suppressed.

Individual smoothness loss: Besides the inter-consistency loss, we also consider isolate loss for each decomposed component separately by considering their own smoothness properties. On the one hand, the illumination maps should be piece-wise smooth, thus we introduce a structure-aware smoothness loss \mathcal{L}_S^L to constraint both L_l and L_h :

$$\mathcal{L}_{sm}^L = \left\| \frac{\nabla L_l}{\max\{(\nabla R_l)^2, \tau\}} \right\|_1 + \left\| \frac{\nabla L_h}{\max\{(\nabla R_h)^2, \tau\}} \right\|_1, \quad (4)$$

where τ denotes a small positive constant which is empirically set to $\tau = 0.01$ to avoid the denominator being zero. This loss measures the relative structure of the illumination with respect to the reflectance. Therefore, the illumination loss can be aware of image structure reflected by the reflectance. Specifically, for a strong edge point in the reflectance map, the penalty on the illumination will be small; for a point in the flat region of the reflectance map, the penalty on the illumination turns to be large.

On the other hand, different from the illumination maps that should be piece-wise smooth, the reflectance maps are usually tend to be piece-wise continuous. Thus, we directly introduce a classical total-variation loss \mathcal{L}_S^R to constraint both R and R_{he} :

$$\mathcal{L}_{sm}^R = \|\nabla R_l\|_1 + \|\nabla R_h\|_1. \quad (5)$$

Image reconstruction loss: The third consideration is that the decomposed two components should well reproduce

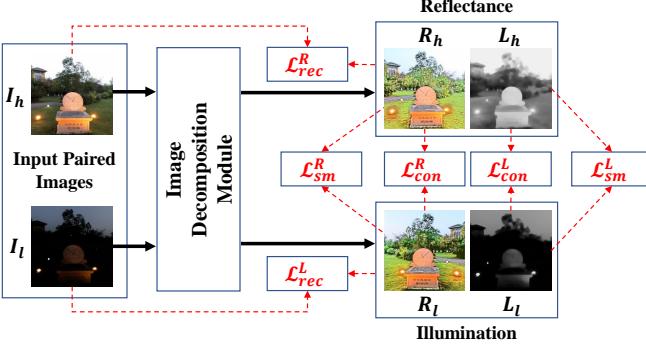


Fig. 3. Diagram for the training process of our image decomposition module.

the input in the fashion of element-wise product, which is constrained by an image reconstruction loss:

$$\mathcal{L}_{rec} = \|I_l - L_l \otimes R_l\|_1 + \|I_h - L_h \otimes R_h\|_1. \quad (6)$$

Total loss: The total loss for our layer decomposition module is defined as follows:

$$\mathcal{L}_{decomp} = \mathcal{L}_{con}^R + \mathcal{L}_{con}^L + \mathcal{L}_{sm}^R + \mathcal{L}_{sm}^L + \mathcal{L}_{rec}, \quad (7)$$

To facilitate understanding, as shown in Fig. 3, we draw a simple diagram to better illustrate the training process of our image decomposition module where all the involved loss terms are specified. It should be emphasized that our image decomposition module is pre-trained on a collection of images pairs of the same scenes with different light/exposure configurations. Such paired images are only required during the pre-training stage. Once it is pre-trained, all the parameters of this module will not be updated during the training of other modules involved in our DDB-Net.

B. Feature Encoding Module

After obtaining the reflectance and illumination components via the pre-trained image decomposition module, the next step is to build feature representations for each of these two components separately. In this work, we design a simple self-reconstruction-based encoder-decoder architecture to achieve this goal. Specifically, both the reflectance and illumination components share the same feature encoding network architecture. However, these two feature encoding networks are optimized with different loss terms, i.e., tailored loss terms are designed to regularize the feature encoding of reflectance and illumination components separately. In what follows, we first present the detailed network configurations and then introduce the tailored loss terms.

As shown in the upper right of Fig. 1, our proposed self-reconstruction-based encoder-decoder module involves two parts namely encoder and decoder. The encoder receives either the reflectance (R) or illumination (L) component as input and progressively forms a set of hierarchical feature representations C_1, C_2, C_3, C_4 , and C_5 . Then, the decoder takes the last feature representations C_5 as input and progressively reconstruct the input signal (\hat{R} or \hat{L}). The detailed architecture of this module is depicted in Fig. 4. The encoder contains several stacked convolutional blocks with each block

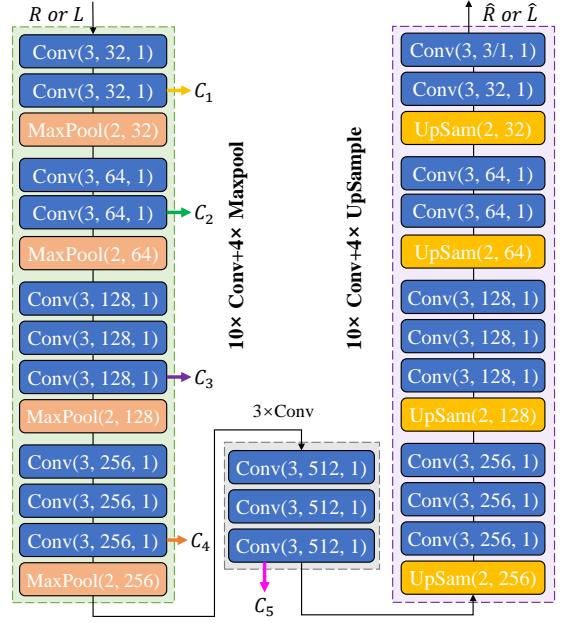


Fig. 4. Architecture of our self-reconstruction-based encoder-decoder network for hierarchical feature learning.

consisting of two or three 3×3 convolutional layers followed by one 2×2 max-pooling layer. The stride of all convolutional layers is set to 1. In addition, each convolutional layer is equipped with an activation layer ReLU. The numbers of feature channels are set to 32, 64, 128, 256, and 512, for C_1, C_2, C_3, C_4 , and C_5 , respectively.

Since the reflectance and illumination components contain NTI degradation information in different aspects, it is necessary to design customized loss terms to guide the reconstruction of each component. To be specific, the loss constraints imposed on the reflectance reconstruction include a structure loss \mathcal{L}_{str} and a color loss \mathcal{L}_{color} , while the loss constraints imposed on the illumination reconstruction include a mean square error (MSE) loss \mathcal{L}_{mse} . The learned reflectance feature representations will focus more on the structural and color information due to the joint guidance of \mathcal{L}_{str} and \mathcal{L}_{color} , while the learned illumination feature representations will focus more on the luminance information with the guidance of \mathcal{L}_{mse} . In the following, we will introduce the definitions and formulations of these loss terms one by one.

1) *Structure loss:* Previous works have reported that the HVS is highly sensitive to the structural information of images and low-quality NTIs will change the structural perception [43]. We adopt the widely-used structural similarity (SSIM) [43] loss between the input reflectance image R and its corresponding reconstructed version \hat{R} for encouraging the encoder to have the capacity of extracting informative structural features. The SSIM loss is defined as follows:

$$\mathcal{L}_{str} = 1 - SSIM(R, \hat{R}), \quad (8)$$

where $SSIM(A, B)$ computes the structural similarity score between image A and image B according to [43].

2) *Color loss:* It is a common sense that NTIs will introduce color distortions and the reflectance component contains

almost all the color information in an image. Thus, a simple yet effective color loss term between R and \hat{R} is desired, which will encourage the encoder to have the capability of extracting color features. Inspired by [44], the blurring operation can remove high frequencies of an image and promote color comparison. Thus, the following color loss is introduced:

$$\mathcal{L}_{color} = \left\| R_B - \hat{R}_B \right\|_2^2, \quad (9)$$

where R_B and \hat{R}_B are the blurred versions of R and \hat{R} , respectively:

$$R_B(i, j) = \sum_{\Omega(i, j)} R(i + \Delta_i, j + \Delta_j) G(\Delta_i, \Delta_j), \quad (10)$$

$$\hat{R}_B(i, j) = \sum_{\Omega(i, j)} \hat{R}(i + \Delta_i, j + \Delta_j) G(\Delta_i, \Delta_j), \quad (11)$$

where $\Omega(i, j)$ is an image patch centered by the pixel at (i, j) and $G(\Delta_i, \Delta_j)$ is the 2-D Gaussian blur kernel, which can be expressed as:

$$G(\Delta_i, \Delta_j) = T \cdot \exp \left(-\frac{(\Delta_i - \mu)^2 + (\Delta_j - \mu)^2}{2\sigma} \right), \quad (12)$$

where the parameters are set according to [44] and we set $T = 0.053$, $\mu = 0$, and $\sigma = 3$, respectively.

3) *MSE loss*: For the reconstruction of illumination component, we only apply a simple MSE loss which is defined by the Euclidean distance between the L and \hat{L} :

$$\mathcal{L}_{mse} = \left\| L - \hat{L} \right\|_2^2. \quad (13)$$

Constrained by the above loss terms, the content-related features and illumination-related features can be well extracted from the reflectance and illumination component, respectively.

C. Bilinear Pooling Module

We consider bilinear techniques to combine the reflectance and illumination feature representations into an unified one. Bilinear models have shown powerful capability in modeling two-factor variations, such as style and content of images [45], location and appearance for fine-grained recognition [46], temporal and spatial aspects for video analysis [47], etc. It also has been applied to address the BIQA problem where the synthetic and authentic distortions are modeled as the two-factor variations [48]. Here, we tackle the blind NTIQE problem with a similar philosophy, where the reflectance and illumination components are modeled as the two-factor variations.

Given an input NTI and its side output feature maps from the reflectance and illumination encoders, C_i^R and C_i^L are both with the size of $h_i \times w_i \times d_i$ since the reflectance and illumination encoder share the same architectures and configurations. Before performing bilinear pooling, C_i^R and C_i^L are separately fed into a 1×1 conv layer to obtain their corresponding compact version with 32 channels (\hat{C}_i^R and \hat{C}_i^L), i.e., $h_i \times w_i \times 32$. Then, bilinear pooling is performed on C_i^R and C_i^L as follows:

$$B_i = (\hat{C}_i^R)^T \hat{C}_i^L, \quad (14)$$

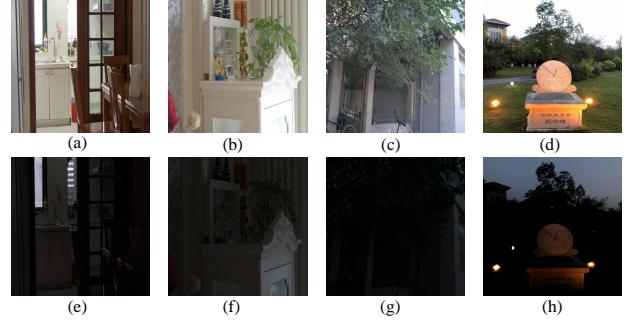


Fig. 5. Sample images pairs with different light/exposure configurations. (a)-(d) are four well-illuminated images. (e)-(h) are the corresponding ill-illuminated images.

where the outer product B_i is a vector of dimension 32×32 .

According to [49], bilinear representation is usually mapped from Riemannian manifold into an Euclidean space using signed square root and ℓ_2 normalization [50]:

$$\hat{B}_i = \frac{\text{sign}(B_i) \odot \sqrt{|B_i|}}{\left\| \text{sign}(B_i) \odot \sqrt{|B_i|} \right\|_2}, \quad (15)$$

where \odot means the element-wise product. Finally, the bilinear pooled feature representations over all scales are concatenated into a single vector:

$$\hat{B} = \text{concat}(\hat{B}_1, \hat{B}_2, \hat{B}_3, \hat{B}_4, \hat{B}_5), \quad (16)$$

Finally, \hat{B} is fed into two fully connected layers for quality prediction, which outputs a scalar indicating the overall quality score. Here, we consider the ℓ_2 norm as the empirical loss, which has been widely used in previous works:

$$\mathcal{L}_{quality} = \frac{1}{K} \sum_{k=1}^K \left\| Q_k - \hat{Q}_k \right\|_2^2, \quad (17)$$

where Q_k is the ground truth subjective quality score of the k -th image in a mini-batch and \hat{Q}_k is the predicted quality score by DDB-Net. It is noteworthy that bilinear pooling is a global strategy and therefore our DDB-Net can receive the input image with an arbitrary size. Consequently, we can directly feed the whole image instead of small patches cropped from it into network during both training and testing stages.

D. Network Training

The training of our proposed DDB-Net involves two stages: pre-training of image decomposition module and training of remaining modules. The pre-training of image decomposition module relies on 500 images pairs of the same scenes while with different light/exposure configurations ¹. Some examples are shown in Fig. 5. Once the image decomposition module is pre-trained, all the parameters of this module will not be updated during the training of remaining modules. Then, the training of remaining modules is based on the target NTI

¹Download: https://pan.baidu.com/s/1ne0gugLmo9_ZnkVEZ3iYSg, password: acsj

quality database by minimizing the following hybrid loss function:

$$\mathcal{L}_{total} = \mathcal{L}_{str} + \mathcal{L}_{color} + \mathcal{L}_{mse} + \mathcal{L}_{quality}. \quad (18)$$

All parameters of the image decomposition module are randomly initialized and we use the Adam optimization algorithm [51] with a mini-batch of 16. We run 100 epoches with a learning rate decaying in the interval $[3 \times 10^{-3}, 3 \times 10^{-4}]$. All the input images are resized into $512 \times 512 \times 3$ before feeding into the network. For the training of the whole DDB-Net, we also adopt the Adam with a learning rate of 3×10^{-5} for the target NTI database and use Batch normalization to stabilize the training process. The model is implemented by PyTorch [52] with a single NVIDIA GTX 2080Ti GPU card.

IV. EXPERIMENTAL RESULTS

In this section, we first describe the experimental setups, including benchmark databases, evaluation protocols, and performance criteria. Then, we compare the performance of DDB-Net with state-of-the-art BIQA models on each individual database. Finally, we conduct several ablation studies to justify the rationality of each critical component involved in DDB-Net and present an application test by applying the DDB-Net to automatic parameter tuning of an off-the-shelf NTI enhancement algorithm.

A. Experimental Setups

1) Benchmark Databases: The main experiments are conducted on the large-scale natural night-time image database (NNID) [25] which contains 2,240 NTIs with 448 different image contents captured by three different photographic equipments (i.e., a digital camera (Device I: Nikon D5300), a mobile phone (Device II: iPhone 8plus) and a tablet (Device III: iPad mini2)) in real-world night-time scenarios. For each image content, one device is used with five different settings to capture five images of different visual quality levels. The five settings are different for different image contents. In NNID, 1,400 images with 280 different image contents are captured by Nikon D5300, 640 images with 128 different image contents are captured by iPhone 8plus, and 200 images with 40 different image contents are captured by iPad mini2. The resolutions of the images in NNID include 512×512 , 1024×1024 , and 2048×2048 . The ground truth subjective quality score, i.e., mean opinion score (MOS), is also provided for each image in the database.

Besides the NNID database, we also use an additional enhanced night-time image database (EHND)² for further performance evaluation. Different from the NNID database which only contains raw NTIs, the EHND database contains both raw NTIs and their corresponding enhanced versions by different NTI enhancement algorithms. Specifically, EHND contains a total number of 1,500 images obtained by applying 15 off-the-shelf NTI enhancement algorithms on 100 raw NTIs. Similarly, the ground truth subjective quality score, i.e., mean opinion score (MOS), for each enhanced NTI is also available.

2) Evaluation Protocols and Performance Criteria: We conduct experiments by following the general evaluation protocol adopted in existing learning-based BIQA studies. Specifically, we divide all the images in each individual database into two splits with the 80% – 20% train-test ratio. The splitting is conducted according to source images to guarantee that there is no overlap of image content. The training and testing procedures are repeated five times on each database so that each image can be tested for once. For each time, we compute four criteria to measure the model performance. The four performance criteria include Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SRCC), Kendall rank order correlation coefficient (KRCC), and root mean square error (RMSE). Among these criteria, PLCC and RMSE measure the prediction precision while SRCC and KRCC measure the prediction monotonicity. These criteria results from the five sessions are calculated respectively and averaged to serve as the final model performance.

B. Performance Comparisons

Since there is always no available pristine reference for real-world NTIs, the quality evaluation of NTIs can only be performed in a no-reference/blind manner. Therefore, we compare the performance of the proposed DDB-Net against 15 existing BIQA methods, including 12 handcraft feature-based BIQA methods (i.e., BLIINDS-II [7], BRISQUE [8], CurveletQA [9], DIIVINE [10], NRSL [11], NFERM [12], GM-LOG [13], GWH-GLBP [14], SSEQ [15], BIQME [16], ILNIQE [17], NIQE [18], and BNBT [25]) and two popular deep learning-based BIQA methods (i.e., WaDIQaM [35] and DBCNN [36]). The handcraft feature-based BIQA methods include two types: training-based and training-free. The training-based ones commonly adopt elaborately designed features to characterize the level of deviations from statistical regularities of natural scenes, based on which a quality prediction function is learned via support vector regression (SVR) [53]. The training-free ones first build a pristine statistical model from a large collection of high-quality natural images and then measure the distance between this pristine statistical model and the statistical model of the distorted image as the estimated quality score. By contrast, the deep learning-based BIQA methods directly optimize an end-to-end function mapping from the input image to its quality score while without any effort on manual feature engineering.

1) Comparisons on NNID: The performance comparison results of different BIQA methods on the NNID database are shown in Table I. From the results, we can have the following observations. First, most training-based BIQA models perform better than the two training-free ones (i.e., NIQE and ILNIQE) while the two deep learning-based BIQA models (i.e., WaDIQaM and DBCNN) are superior to most handcraft feature-based BIQA methods. It is reasonable because BIQA is a challenging task where training is particularly useful to model the complex non-linear relationship between the extracted features and perceived quality score, and end-to-end deep learning technique further provides an effective solution to directly establish the explicit image-to-quality mapping

²<https://sites.google.com/site/xiangtaooo/>

TABLE I
PERFORMANCE RESULTS OF DIFFERENT BIQA METHODS ON THE NNID DATABASE.

Methods	Entire Database (2240 images)				Device I: Nikon D5300 (1400 images)				Device II: iPhone 8plus (640 images)				Device III: iPad mini2 (200 images)			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
BLLIINDS-II	0.7438	0.5403	0.7549	0.1119	0.7520	0.5461	0.7627	0.1108	0.6419	0.4564	0.6574	0.1103	0.6777	0.5048	0.7333	0.0892
BRISQUE	0.7365	0.5352	0.7452	0.1132	0.7315	0.5332	0.7420	0.1150	0.6445	0.4598	0.6652	0.1091	0.5704	0.4166	0.6431	0.0980
CurveletQA	0.8676	0.6762	0.8679	0.0924	0.8937	0.7115	0.8953	0.0844	0.8110	0.6147	0.8183	0.0916	0.7712	0.5881	0.8217	0.0889
DIIIVINE	0.7744	0.5675	0.7637	0.1092	0.7601	0.5545	0.7330	0.1178	0.6830	0.4793	0.5844	0.1187	0.6661	0.4698	0.6491	0.0998
NRSL	0.8291	0.6265	0.8327	0.0936	0.8165	0.6131	0.8192	0.0981	0.7417	0.5417	0.7325	0.1007	0.6625	0.4848	0.6903	0.0966
NFERM	0.8512	0.6572	0.8556	0.1099	0.8706	0.6803	0.8764	0.1110	0.8122	0.6146	0.8224	0.1257	0.7610	0.5727	0.7882	0.1231
GM-LOG	0.8114	0.6072	0.8125	0.0985	0.8135	0.6099	0.8171	0.0992	0.7338	0.5338	0.7313	0.0998	0.6996	0.5117	0.7107	0.0951
GWH-GLBP	0.7111	0.5108	0.7098	0.1350	0.6998	0.5020	0.6819	0.1382	0.6383	0.4731	0.6174	0.1614	0.6244	0.4547	0.7071	0.1343
SSEQ	0.7838	0.5894	0.7865	0.1144	0.7809	0.5878	0.7891	0.1258	0.6735	0.4919	0.6968	0.1451	0.6673	0.4617	0.6436	0.1689
BIQME	0.8255	0.6185	0.8273	0.0911	0.8189	0.6141	0.8245	0.0913	0.8140	0.6144	0.8027	0.0972	0.7935	0.6064	0.7905	0.1005
BNBT	0.8769	0.6822	0.8784	0.1061	0.8866	0.7066	0.8939	0.1020	0.8632	0.6737	0.8698	0.1157	0.8517	0.6890	0.8576	0.1137
ILNIQE	0.7115	0.5183	0.6335	0.1691	0.6712	0.4831	0.6766	0.1679	0.6949	0.5018	0.6809	0.1639	0.7983	0.6086	0.6721	0.1720
NIQE	0.5983	0.4220	0.5701	0.1803	0.6007	0.4240	0.5859	0.1847	0.5772	0.4017	0.5874	0.1811	0.6591	0.4694	0.6092	0.1842
WaDIQaM	0.8272	0.6213	0.8229	0.0954	0.8127	0.6258	0.8263	0.0895	0.8194	0.6017	0.8101	0.0952	0.8069	0.6048	0.8016	0.0937
DBCNN	0.8938	0.6953	0.8958	0.0849	0.8745	0.6779	0.8826	0.0843	0.8704	0.6738	0.8796	0.0852	0.8526	0.6539	0.8614	0.0893
DDB-Net	0.9275	0.7841	0.9284	0.0752	0.9146	0.7828	0.9183	0.0774	0.8901	0.7635	0.8928	0.0813	0.8857	0.7542	0.8796	0.0837

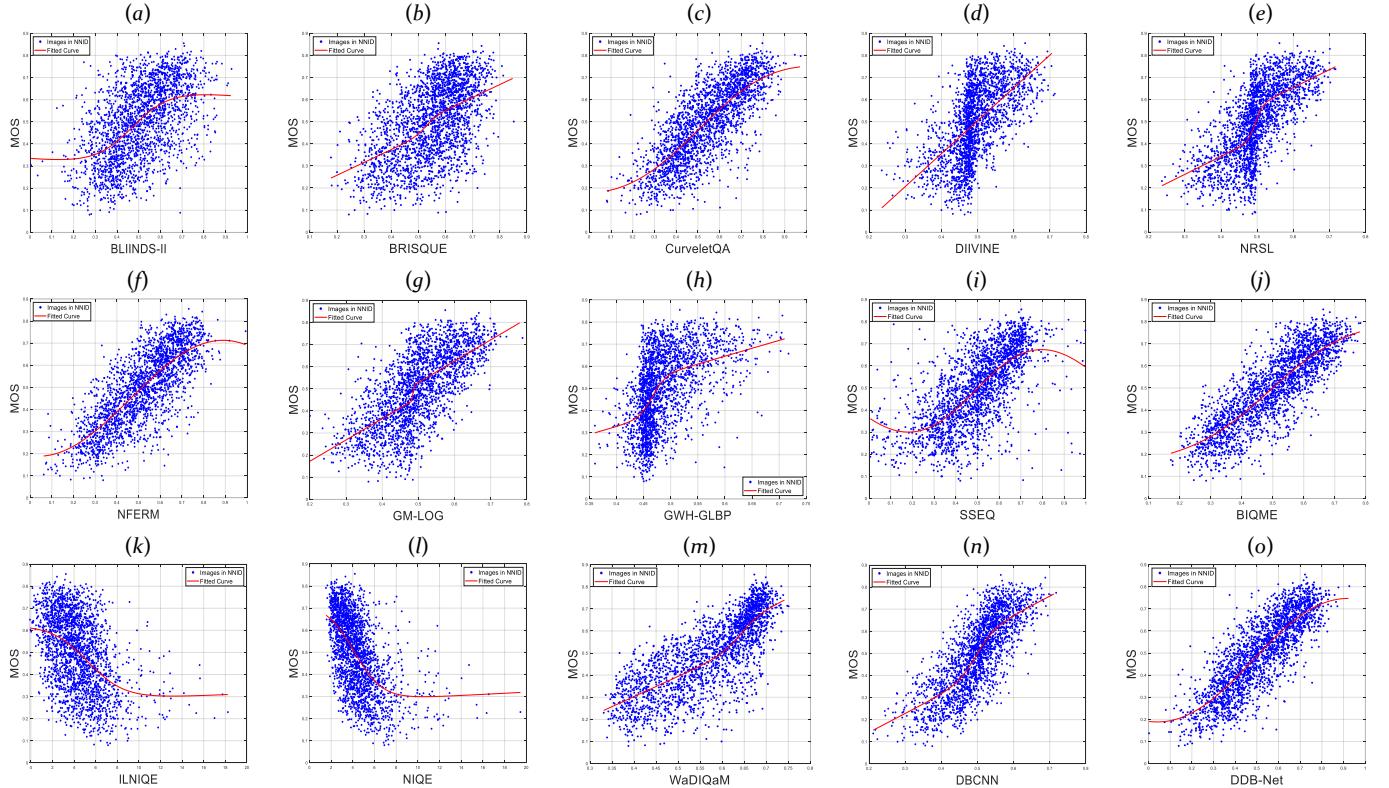


Fig. 6. Scatter plots between the objective scores (predicted by BIQA methods) and the subjective MOSs (provided in the NNID database). (a)-(o) correspond to BLLIINDS-II [7], BRISQUE [8], CurveletQA [9], DIIIVINE [10], NRSL [11], NFERM [12], GM-LOG [13], GWH-GLBP [14], SSEQ [15], BIQME [16], ILNIQE [17], NIQE [18], WaDIQaM [35], DBCNN [36], and DDB-Net, respectively.

function owing to its powerful capacity of automatic feature representation learning. Second, the existing NSS feature-based BIQA cannot obtain satisfactory results for evaluating NTIs as NSS is not that suitable to characterize the degradation

properties of in-the-wild NTIs. Third, the proposed DDB-Net delivers the best performance among all the competitors in terms of all performance criteria, i.e., the highest PLCC, SRCC, KRCC values and the lowest RMSE value. The reason

	BLIINDS-II	BRISQUE	CurveletQA	DIIVINE	NRSL	NFERM	GM-LOG	GWH-GLBP	SSEQ	BIOME	ILNIQE	NIQE	WaDIQaM	DBCNN	DDB-Net
BLIINDS-II	-	1	-1	-1	-1	-1	1	-1	-1	1	1	1	-1	-1	-1
BRISQUE	-1	-	1	-1	-1	-1	1	-1	-1	1	1	-1	-1	1	1
CurveletQA	1	1	-	1	1	1	1	1	1	1	1	1	1	-1	-1
DIIVINE	1	1	-1	-	-1	-1	1	-1	-1	1	1	-1	-1	-1	-1
NRSL	1	1	-1	1	-	-1	1	1	1	1	1	-1	-1	-1	-1
NFERM	1	1	-1	1	1	-	1	1	1	1	1	1	-1	-1	-1
GM-LOG	1	1	-1	1	-1	-1	-	1	1	-1	1	1	-1	-1	-1
GWH-GLBP	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1
SSEQ	1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	-1
BIQME	1	1	-1	1	-1	-1	1	1	1	-	1	-1	-1	-1	-1
ILNIQE	-1	-1	-1	-1	-1	-1	-1	-1	-1	-	1	-1	-1	-1	-1
NIQE	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-	1	-1	-1	-1
WaDIQaM	1	1	-1	1	1	1	-1	1	1	1	1	-	-1	-1	-1
DBCNN	1	1	1	1	1	1	1	1	1	1	1	1	-	-1	-1
DDB-Net	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-

Fig. 7. Significance t-test results on NNID. In the figure, 1/-1 indicates that row models perform statistically better/worse than the column models.

is that we have dedicated decomposing the sophisticated blind NTIQE task into two easier sub-tasks with each sub-task accounting for illumination perception and content perception, respectively. In such a way, the features related to the illumination perception and the content perception can be better learned and finally fused to facilitate blind NTIQE.

In addition to the numerical performance results, we also show the scatter plots between the objective scores (predicted by BIQA methods) and the subjective MOSs (provided in the database) in Fig. 6. In the scatter plot, each point corresponds to an image in the NNID database and the x -axis represents the prediction scores by BIQA methods while the y -axis represents the ground truth subjective MOSs. A good BIQA method is expected to produce scatter points that are close to the fitted curve. It can be easily found from Fig. 6 that the proposed DDB-Net produces the best fitting result on the NNID database.

Finally, we use a hypothesis testing approach based on t-statistics [54] to further demonstrate the superiority of our proposed DDB-Net. In our experiment, the two-sample t-test between the pair of PLCC values at the 5% significance level is conducted. Fig. 7 shows the results of t-test, where the value 1/-1 indicates that row models perform statistically better/worse than the column models. From the results, we find that our DDB-Net always performs better than all the other competitors, which further validates the superiority.

2) *Comparisons on EHND*: A well-performing NTIQE should also be able to measure the performance of different NTI quality enhancement algorithms, i.e., well evaluate different enhanced results. Actually, a certain enhancement algorithm may result in particularly bad enhanced result which may still suffer from unsatisfactory brightness and even more serious color distortions than the original raw NTI. Therefore, we also evaluate the performance of different BIQA methods on another nighttime image database EHND which contains 1,500 images obtained by applying 15 off-the-shelf NTI enhancement algorithms on 100 raw NTIs. The numerical

TABLE II
PERFORMANCE RESULTS OF DIFFERENT BIQA METHODS ON THE EHND DATABASE.

Methods	SRCC (\uparrow)	KRCC (\uparrow)	PLCC (\uparrow)	RMSE (\downarrow)
BLIINDS-II	0.7168	0.5016	0.7026	0.7383
BRISQUE	0.7021	0.5077	0.6907	0.7424
CurveletQA	0.7525	0.5743	0.7624	0.6931
DIIVINE	0.6868	0.4750	0.6216	0.7593
NRSL	0.7853	0.5845	0.7812	0.6241
NFERM	0.7546	0.5683	0.7532	0.6831
GM-LOG	0.7915	0.5947	0.7794	0.6325
GWH-GLBP	0.7235	0.5074	0.7196	0.7240
SSEQ	0.7012	0.5135	0.6981	0.7383
BIQME	0.6916	0.4847	0.7174	0.7063
ILNIQE	0.3815	0.2145	0.4637	0.8034
NIQE	0.2723	0.1839	0.3125	0.8279
WaDIQaM	0.7528	0.5843	0.7598	0.6865
DBCNN	0.7935	0.6442	0.8051	0.6234
DDB-Net	0.8647	0.6852	0.8748	0.6057

	BLIINDS-II	BRISQUE	CurveletQA	DIIVINE	NRSL	NFERM	GM-LOG	GWH-GLBP	SSEQ	BIOME	ILNIQE	NIQE	WaDIQaM	DBCNN	DDB-Net
BLIINDS-II	-	1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	-1	-1
BRISQUE	-1	-	1	1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1
CurveletQA	1	1	-	1	-1	1	1	1	1	1	1	1	1	-1	-1
DIIVINE	-1	-1	-1	-	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1
NRSL	1	1	1	1	-	1	1	1	1	1	1	1	1	-1	-1
NFERM	1	1	-1	1	-1	-	1	1	1	1	1	1	-1	-1	-1
GM-LOG	1	1	-1	1	-1	-1	-	1	1	1	1	1	-1	-1	-1
GWH-GLBP	1	1	-1	1	-1	-1	-1	-	1	1	1	1	-1	-1	-1
SSEQ	-1	1	-1	1	-1	-1	-1	-1	-	1	1	1	-1	-1	-1
BIQME	1	1	-1	1	-1	-1	-1	-1	-1	-	1	1	-1	-1	-1
ILNIQE	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-	1	-1	-1	-1
NIQE	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-	-1	-1	-1
WaDIQaM	1	1	-1	1	-1	1	1	1	1	1	1	1	-	-1	-1
DBCNN	1	1	1	1	1	-1	1	1	1	1	1	1	-	-1	-1
DDB-Net	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-

Fig. 8. Significance t-test results on EHND. In the figure, 1/-1 indicates that row models perform statistically better/worse than the column models.

performance results of different BIQA methods on the EHND database are shown in Table II and the significance t-test results are shown in Fig. 8. It is observed from these results that our proposed DDB-Net outperforms other competitors by a large margin in terms of all performance criteria on the EHND database.

In this case, the most important role of NTIQE is to automatically select the one with the highest visual quality from 15 enhanced results generated from the same NTI. Therefore, it is of great interests to conduct experiments to further compare such a kind of capability of different BIQA methods. Specifically, we measure the rank- n accuracy which is closely relevant with the capability of a certain objective quality metric in selecting the optimal enhanced result from a set of candidates. Given 15 different enhanced results associated with the same raw NTI, the rank- n accuracy is defined as the percentage of images whose top-1 result in terms of MOS

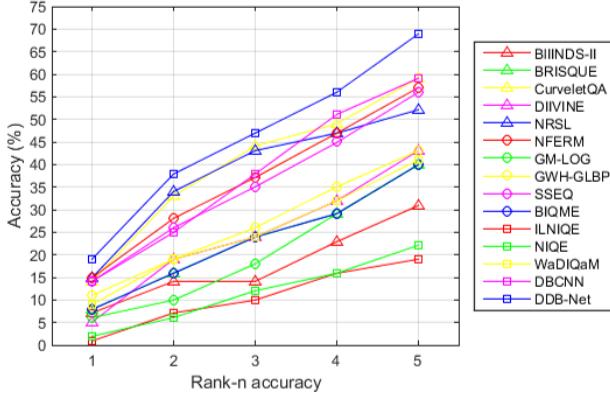


Fig. 9. Comparison of the rank-1, rank-2, rank-3, rank-4, and rank-5 accuracy values by different BIQA methods on the EHND database.

appears within the top- n results in terms of objective predicted score. Obviously, a higher rank- n accuracy value indicates a better performance of a certain NTIQE. In Fig. 9, we show the rank-1, rank-2, rank-3, rank-4, and rank-5 accuracy values by different BIQA methods on the EHND database. It is observed that our DDB-Net always delivers highest rank- n accuracy values, indicating the best capability in selecting the one with the highest visual quality from a set of candidates.

C. Application: Automatic Parameter Tuning of NTI Quality Enhancement Algorithm

An effective blind NTIQE should be able to well guide the optimization of NTI quality enhancement algorithms. In this section, we demonstrate this idea by applying the proposed DDB-Net to automatic parameter tuning of off-the-shelf NTI quality enhancement algorithms. There are always one or several parameters in NTI quality enhancement algorithms whose optimal values vary with contents. It is challenging and time-consuming to handpick a set of parameters that work well for all image contents. A well-performing blind NTIQE is able to replace the role of humans in this task, especially when the volume of images to be processed is particularly large.

Here, we use the LIME algorithm [55] as a representative example of NTI quality enhancement algorithm, which involves two tunable parameters g and l . The default values are: $g = 0.6$ and $l = 0.2$. However, the visual quality of the final enhanced image is highly sensitive to these two parameters. Fig. 10 shows example images generated with different g and l values. In the figure, warmer color indicates better predicted quality of the corresponding enhanced image. The corresponding scores predicted by our DDB-Net are also shown under each image. By varying g and l , we can obtain enhanced results with significantly different visual quality. For example, the two enhanced results in the left side of Fig. 10 still suffers from over-/under exposure problem while the two enhanced results in the right side exhibits much better visual quality with much more finer details and natural color appearance. It is found that our DDB-Net can evaluate their visual qualities consistently with human subjective perception. Furthermore, we also find that the visual quality of the upper right image is better than that of the bottom right one which is

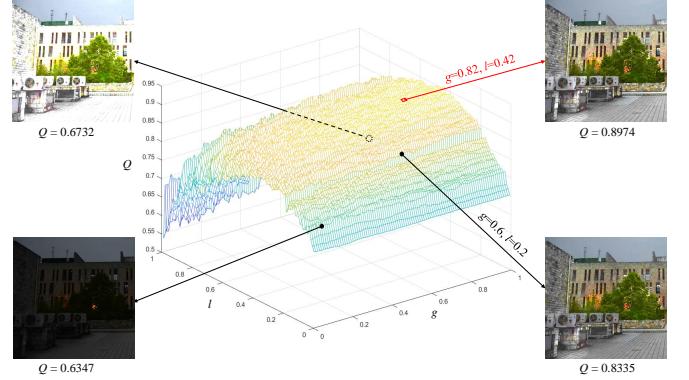


Fig. 10. Automatic parameter tuning of an off-the-shelf NTI quality enhancement algorithm using the proposed DDB-Net method. Warmer color in the surface plot represents better visual quality.

produced by using the default parameter values. It means that it is possible to adaptively determine the optimal parameter values under the guidance of our proposed DDB-Net.

V. CONCLUSION

This paper has presented a novel deep NTIQE called DDB-Net which consists of three modules namely image decomposition module, feature encoding module, and bilinear pooling module. With the help of decomposing the input NTI into two independent layer components (illumination and reflectance), the degradation features related to illumination perception and content perception are better learned and then fused with bilinear pooling to improve the performance of blind NTIQE. Experiments on two benchmark databases have demonstrated the superiority of our proposed DDB-Net.

Although our proposed DDB-Net is promising, future works towards further improving the performance may focus on the following directions: 1) designing more efficient unsupervised solutions for image layer decomposition; 2) designing more effective loss functions to facilitate learning degradation features from each component; 3) designing more powerful feature fusion schemes by considering other variants of bilinear pooling to further improve the performance.

REFERENCES

- [1] G. Zhai and X. Min, "Perceptual image quality assessment: A survey," *Science China: Information Sciences*, vol. 63, no. 11, pp. 76–127, 11 2020.
- [2] Y. Zhan and R. Zhang, "No-reference jpeg image quality assessment based on blockiness and luminance change," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 760–764, 2017.
- [3] S. A. Golestaneh and D. M. Chandler, "No-reference quality assessment of jpeg images via a quality relevance map," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 155–158, 2014.
- [4] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3218–3231, 2015.
- [5] T. Oh, J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, "No-reference sharpness assessment of camera-shaken images by analysis of spectral structure," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5428–5439, 2014.

- [6] C. Tang, X. Yang, and G. Zhai, "Noise estimation of natural images via statistical analysis and noise injection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1283–1294, 2015.
- [7] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.
- [8] 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.
- [9] L. Liu, H. Dong, H. Huang, and A. C. Bovik, "No-reference image quality assessment in curvelet domain," *Signal Processing: Image Communication*, vol. 29, no. 4, pp. 494–505, 2014.
- [10] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.
- [11] 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.
- [12] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.
- [13] 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.
- [14] 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, 2016.
- [15] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.
- [16] 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, 2017.
- [17] 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.
- [18] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.
- [19] Q. Li, W. Lin, and Y. Fang, "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing*, vol. 236, pp. 93–103, 2017.
- [20] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1098–1105.
- [21] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4444–4457, 2016.
- [22] Q. Jiang, F. Shao, G. Jiang, M. Yu, and Z. Peng, "Supervised dictionary learning for blind image quality assessment using quality-constraint sparse coding," *Journal of Visual Communication and Image Representation*, 2015.
- [23] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.
- [24] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 425–440, 2016.
- [25] 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.
- [26] J. McCann, *Retinex Theory*. New York, NY: Springer New York, 2016, pp. 1118–1125.
- [27] J. Wu, J. Zeng, Y. Liu, G. Shi, and W. Lin, "Hierarchical feature degradation based blind image quality assessment," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 510–517.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [29] Y. Li, L.-M. Po, L. Feng, and F. Yuan, "No-reference image quality assessment with deep convolutional neural networks," in *IEEE International Conference on Digital Signal Processing*, 2016, pp. 685–689.
- [30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [31] 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.
- [32] Y. Li, L.-M. Po, X. Xu, L. Feng, F. Yuan, C.-H. Cheung, and K.-W. Cheung, "No-reference image quality assessment with shearlet transform and deep neural networks," *Neurocomputing*, vol. 154, pp. 94–109, 2015.
- [33] X. Liu, J. van de Weijer, and A. D. Bagdanov, "Rankqa: Learning from rankings for no-reference image quality assessment," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1040–1049.
- [34] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2017.
- [35] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2017.
- [36] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2018.
- [37] 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.
- [38] X. Min, G. Zhai, K. Gu, Y. Zhu, J. Zhou, G. Guo, X. Yang, X. Guan, and W. Zhang, "Quality evaluation of image dehazing methods using synthetic hazy images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2319–2333, 2019.
- [39] Q. Wu, L. Wang, K. N. Ngan, H. Li, and F. Meng, "Beyond synthetic data: A blind deraining quality assessment metric towards authentic rain image," in *IEEE International Conference on Image Processing*, 2019, pp. 2364–2368.
- [40] 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.
- [41] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.
- [42] P. Guo, L. He, S. Liu, D. Zeng, and H. Liu, "Underwater image quality assessment: Subjective and objective methods," *IEEE Transactions on Multimedia*, 2021.

- [43] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [44] A. Ignatov, N. Kobyshev, R. Timofte, and K. Vanhoey, “Dslr-quality photos on mobile devices with deep convolutional networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 3297–3305.
- [45] J. B. Tenenbaum and W. T. Freeman, “Separating style and content,” in *NIPS*, 1996.
- [46] T.-Y. Lin, A. RoyChowdhury, and S. Maji, “Bilinear cnn models for fine-grained visual recognition,” *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1449–1457, 2015.
- [47] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” in *NIPS*, 2014.
- [48] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, “Blind image quality assessment using a deep bilinear convolutional neural network,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 36–47, 2020.
- [49] X. Pennec, P. Fillard, and N. Ayache, “A riemannian framework for tensor computing,” *International Journal of Computer Vision*, vol. 66, pp. 41–66, 2005.
- [50] F. Perronnin, J. Sánchez, and T. Mensink, “Improving the fisher kernel for large-scale image classification,” in *ECCV*, 2010.
- [51] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2015.
- [52] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *NeurIPS*, 2019.
- [53] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011.
- [54] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. Applied statistics and probability for engineers, 2014.
- [55] X. Guo, Y. Li, and H. Ling, “Lime: Low-light image enhancement via illumination map estimation,” *IEEE Transactions on Image Processing*, vol. 26, pp. 982–993, 2017.