

Blind Quality Assessment of Night-Time Photos: A Region Selective Approach

Zongxi Han , Guangtao Zhai and Rong Xie *

Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China

* Correspondence: xierong@sjtu.edu.cn

Abstract: Despite the emergence of low-light enhancement algorithms and the associated quality assessment metrics in literature, there are rare works considering the quality assessment of real night-time photos captured by mobile cameras. In this paper, we handle this task by first constructing a night-time photo database (NPHD), which consists of 510 photos captured by 30 mobile devices in 17 scenes. Their mean opinion scores are rated by 10 people using the anchor ruler method. Furthermore, we propose a region selective approach for the objective image quality assessment (RSIQA), based on which different feature sets are extracted. Specifically, the center and around regions are partitioned for the brightness, contrast, vignetting, saturation and shading. The brightest areas are located as the region where the highlight suppressing capability is qualified. Finally, we select the foreground and sharpest regions for the assessment of preserving details, naturalness, noises, and image structure. To map different/multiple quality attributes of the night-time photo into a single quality score, four regressors: support vector regression, decision tree, random forest or AdaBoost.R2 are chosen and compared. Experiments on NPHD demonstrate that the proposed RSIQA achieves superior result compared to 17 state-of-the-art, 4 types of quality metrics, including conventionally general-purpose, deep learning based, contrast oriented and night specific ones.

Keywords: low-light enhancement; night-time photos; anchor ruler method; region selective approach; multiple quality attributes; quality regressors

1. Introduction

With rapid development of mobile imaging techniques [1,2], pictures taken by smartphones have dominated the social media and our lives. However, compared to daytime shots, night-time photography places a more challenging task, especially for a mobile camera. The reasons are as follows: 1) the night environment is characterized by low-light conditions, which often lead to under-exposure appearance; 2) due to the tiny sensor and high digital ISO on mobile cameras, dark areas and skies in night-time photos are prone to white and color noises; 3) in night-time cityscape scenes, light sources such as the neon lamps and billboards can raise the dynamic range of the illuminance as high as 100000:1 [3], making it difficult to balance the tone of the whole picture.

Traditional nightography overcomes these difficulties by relying on large sensors and fast apertures on DSLRs, long exposure time on the tripod, and post-processing of the raw format through software such as Adobe Camera Raw. However, such pipeline requires much time, as well as expertises at photo editing. In contrast, low-light image enhancement algorithms (LIEAs) mainly handle this task by four types of approaches: **heuristic method**, **Retinex-theory based decomposition**, **fusion method**, and **deep learning based ones**.

Among the heuristic methods, gamma correction or power function, logarithmic transform, and contrast stretching are commonly used to increase the brightness or contrast of the image [4]. Another widely used technique is the histogram equalization (HE), which aims to produce an uniform histogram distribution. However, HE always results in over-enhancement artifacts and washed-out appearances. To improve HE, different histogram modification strategies such as partitioning, weighting and smoothing have been proposed

Citation: Han, Z.; Xie, R.; Zhai, G.; Blind Quality Assessment of Night-Time Photos: A Region Selective Approach. *Journal Not Specified* **2023**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2024 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

to restrict the contrast or brightness enhancement degree. Representative works can be found in [5–10].

According to the Retinex theory [11], an image I is regarded as the product: $I = \mathbf{R} \circ \mathbf{L}$, where \mathbf{R} and \mathbf{L} refers to the scene reflectance and illumination map, respectively. Arguing that logarithmic operation of the single/multi-scale, or variational retinex [12–15] would amplify irrelevant details due to $\nabla(\log(x)) = \frac{1}{x}\nabla x$, Fu *et al.* [16] used a weighted total variation model to recover fine textures in the reflectance map. In [17], the same author formulated the joint estimation of \mathbf{R} and \mathbf{L} as a maximum *a posteriori* (MAP) problem in the linear domain. To alleviate halo artifacts and low-contrast problem in the enhanced image, Li *et al.* [18,19] penalized the ℓ_1 -norm of the illumination gradients instead of ℓ_2 -norm, and explicitly regularized the reflectance map by its ℓ_2 -norm fidelity with a scaled image gradient. Moreover, a noise penalty term, along with the reflectance gradient thresholding, are introduced to suppress noise. Instead of solving \mathbf{R} and \mathbf{L} simultaneously, NPE [20] directly made a coarse estimation of the \mathbf{L} through the maximum value of all three channels in a small neighborhood. Then a brightness-pass filter (i.e. a range filter after morphological closing) and bi-log histogram specification (i.e. a JND-inspired [21] tone mapping function) are leveraged to refine and enhance the illuminance map, respectively. Improving on NPE, Guo *et al.* [22] applied the first iteration of the relative total variation filter [23] on the coarse \mathbf{L} in order to remove textures and obtain a piecewise smooth illumination map. Gamma correction and an adaptive BM3D [24] are performed afterwards to increase the illumination and suppress noise. Realizing the potential of beta-gamma brightness transfer function (BTF) for enlightening the image, Ying *et al.* [25] applied a spatially varying BTF for each pixel. The pixel-wise enhanced level is determined by the inverse of the estimated illuminance map. After that, the same authors discussed various forms of BTFs and compared their enhancement performances in [26].

The basic idea of fusion-based methods is to first generate a set of enhanced images (or \mathbf{L}) from the low-light input, and then merge them together. In [27], the author decomposed the original image into multi-layer of \mathbf{L} using a cascade of brightness-pass filter. Each layer of \mathbf{L} is enhanced based on the statistical lightness prior of daytime photos, and synthesized back with \mathbf{R} into the final result. In [28], the intermediate one is obtained by applying a global BTF on the low-light image, where the single enhanced level is determined by maximizing the entropy (i.e. gray-level distribution) of the enhanced dark areas. The merging weight is calculated by the power function of \mathbf{L} . In [29], instead of enhancing \mathbf{L} via the common gamma function, the author merged three maps: the original \mathbf{L} , arc tangent function-boosted \mathbf{L} and the CLAHE-enhanced [5] \mathbf{L} into the final illumination map. The classic Laplacian pyramid [30,31] is utilized for fusion, where the merging weight considers both the brightness and color contrast factor [32].

Among the deep learning methods, LLnet [33] was the first to use sparse denoising autoencoder (SDA) to simultaneously brighten and denoise the low-light image. After that, the convolutional neural networks (CNNs) have been widely utilized for the LIEA. Taking low and normal-light image pairs as training inputs, the Retinex-Net [34] used several 3×3 convolutional layers and ReLus to decompose images into \mathbf{R} and \mathbf{L} under the constraint of reconstruction loss, reflectance invariance, and illumination smoothness. An enhancement-subnet was learned to adjust the \mathbf{L} by minimizing the ℓ_2 -norm between the enhanced image and normal-light image. Slightly different from Retinex-Net, KinD [35] adopted a 5-layer U-Net [36] and a 3-layer CNN for the decomposition of \mathbf{R} and \mathbf{L} . A restoration-subnet and an adjustment-subnet were used for denoising and lightening. Instead of performing the decomposition process, the MBLLEN [37] extracted multi-level features at each stage of a 10-layer CNN, which are enhanced via an encoder-decoder module. Thereafter, these feature maps are merged into the final image through a 1×1 convolutional kernel. In this way, MBLLEN can also be deemed as fusion-based method. Inspired by [38], EnlightenGAN [39] adopted an attention guided U-Net as the image generator using unpaired training sets, and included a global relativistic discriminator, in addition to a local one to handle spatially varying light conditions inside the image. The loss function combines local and global

GAN losses with the VGG perceptual loss [40]. To reduce the search domain, Zero-DCE [41] solved the LIEA in the curve parameter space. The network output is a sequence of quadratic curve parameter maps to be applied on each pixel, and the referenceless training losses consist of spatial consistency, exposure control, color constancy and illumination smoothness. In [42], Liu *et al.* designed a two-component network to sequentially refine **L** and denoise the **R**. Neural architecture search and feature distillation techniques are utilized for speeding. Since the enhanced image quality would largely depend on the training losses, Chen *et al.* [43] integrated image quality assessment (IQA) with LIEA into a loop game. The quality prediction loss and enhancement quality loss are optimized alternatively until convergence. For a complete overview of deep learning based LIEAs, readers can refer to [44].

Aside from the above LIEAs, nearly all phone manufacturers have developed their own night-mode algorithms to improve the image quality of night-time photos. One typical example is the *night sight* algorithm [45,46]. In Figure 1, we compare 13 state-of-the-art LIEA images with a modern mobile camera. Images enhanced by histogram based methods and RUAS [42] are not presented, since they are *less effective* compared to others in our experiments. From Figure 1, we have the following observations: the NPE amplifies too many fine details (or noises) in the low-frequency component, such as the roofs. The SRIE suffers from halo artifacts (see black borders around the two background buildings). Although PIE alleviates the halo problem, the darkest areas aren't enhanced very effectively (see the roofs). Both CRF and EFF preserve image naturalness well and produce minimal artifacts, but their saturation are sacrificed, resulting a slightly pale image (although EFF has increased contrast of CRF by fusing with the original image). As for the two methods with default denoising, the LIME tends to over-saturate and over-expose the image (see the street lamps). Its perceived contrast is also not as normal as the SRIE, PIE, EFF and FEM. In contrast, the SRLLIE image has a flatter look. Among the deep-learning based methods, the Retinex-net image is characterized by oil-painting-like appearance. Due to the weakly learned **R** and **L**, curly, fake details are generated in the KinD image (see the enlarged sky and trees). The KinD also avoids severe red color cast by operating in individual R, G, B channels, but its grayed image doesn't reproduce the true night tone, either. Both Zero-DCE and EnlightenGAN preserve details better than KinD, but Zero-DCE suffers from halo effects (see slight haloing of the tall building) while the EnlightenGAN presents noticeably yellowish shifts (see the yellowish building and trees). Lastly, benefiting from the encoder-decoder module, the MBLLEN has impressive denoising performance, significantly better than that of KinD, LIME and SRLLIE (we have fine-tuned the noise variance of BM3D in LIME, the threshold and noise penalty parameter in SRLLIE for this single scene). For the figure formatting, we don't show the result of LLnet. In fact, the denoising quality of LLnet is similar to MBLLEN, but LLnet suffers from over-brightening, pale color and slight haloing.

Overall, EFF, Zero-DCE achieve the best result in terms of luminance tone mapping and white balancing (EFF mimics the operation of increasing digital ISO in shooting; ZeroDCE mimics pulling curves in the photoshop; the other metrics working in HSV space run the risk of rendering the dark sky too red), while the MBLLEN is the top denoising method. However, it can be seen from Figure 1 (b) that, there exists a quality gap between LIEA images and the night-mode one by a mobile camera, in terms of highlight suppressing (see the gas station), detail preserving, noise cancelation, artifacts reduction such as banding, blocking, etc. One reason is that the night-mode leverages the multi-frame stacking technique. Another is that different network architectures such as denoising autoencoder (DAE) [33,37] and GAN [47] should be further integrated for the noise & texture tradeoff.

In this article, instead of assessing or improving current low-light enhancement algorithms, we focus on the quality assessment of real night-time photos taken by smartphones. Despite the emergence of IQA metrics for LIEAs, there are rare works considering the quality assessment of night images taken by **mobile cameras** [48]. To achieve this goal, we

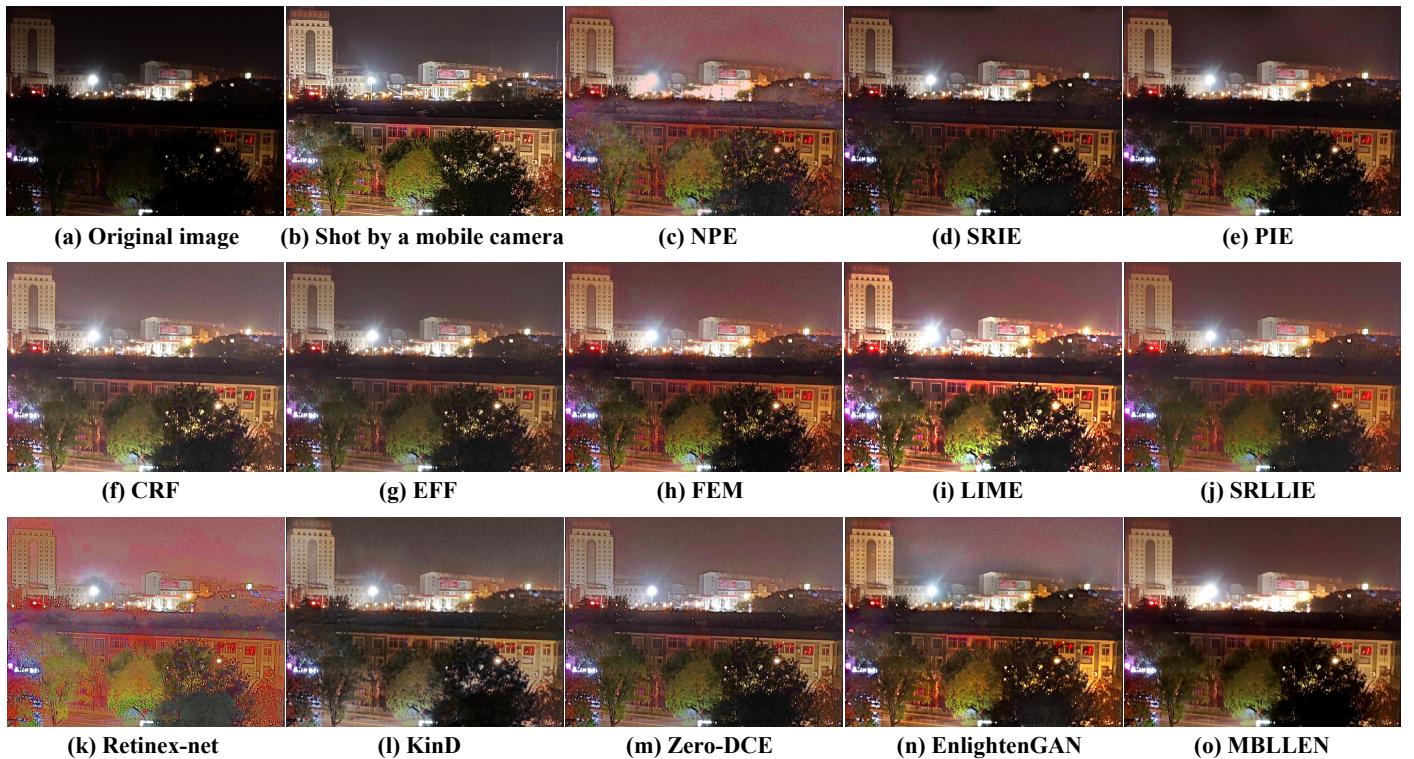


Figure 1. Visual comparison of state-of-the-art LIEAs with a mobile camera. (a) original low-light image (ISO=400, T=0.5s, EV=-3). (b)-(o) images enhanced by a camera, NPE [20], SRIE[16], PIE[17], CRF[25], EFF[28], FEM[29], LIME[22], SRLLIE[19], Retinex-net[34], KindD[35], EnlightenGAN[39], Zero-DCE[41], and MBLLEN[37]. Please zoom in the high-resolution images for a clear view.

first build a night-time photo database called NPHD, where 510 photos are taken by 30 devices in 17 scenes. Their mean opinion score are obtained using anchor ruler method. Then we propose a region selective approach for the objective quality assessment of night-time photos where three different sets of quality features are extracted in distinct region proposals. Finally, four types of quality regressors map features into a quality score.

In summary, our contributions are as follows:

- Our Night-time Photo Database (NPHD) extends from our previous work [49]. Not only are 7 more test scenes added, but also mean opinion scores are provided using the anchor ruler method. Compared to [48], our database contains more (30 vs.3), relatively newer devices, and more importantly, is closer to practical usage.
- Compared to IQA metrics [50–54] where either all patches or randomly cropped patches are assessed, our quality metric explicitly finds region proposals where quality assessment is probably performed. Such region selective strategy is close to how people rate night-time photos subjectively (e.g. subjects pay more attention to certain areas when assessing night-time photos).
- Based on the region proposals, distinct feature sets are extracted in different regions. This differentiates our metric from existing IQA algorithms [50–54] where the same feature sets are computed for every patch.
- To determine the weights of different features/quality aspects, four regressors: SVR, decision tree random forest, AdaBoost are chosen and compared. Experiments show that the proposed region selective feature extraction process, with the random forest regressor achieves the best result in the night-time photo database, outperforming general-purpose, contrast-oriented, and night-time specific algorithms.

2. Related Work

In this section, we review four types of IQA metrics that are closely related to our work: conventionally general purpose, deep learning based, contrast oriented and night-time specific ones. Existing night-time IQA databases will also be mentioned.

2.1. Conventionally general purpose IQA metrics

In [55], BRISQUE used the mean subtracted contrast normalized (MSCN) coefficients in the spatial domain as the natural scene statistics (NSS), which are regressed into quality score by SVR. With the same MSCN features, NIQE [50] extracted them in a patchwise manner, and fitted them through a MVG model. The quality score is computed by the symmetric JS-divergence between MVGs of test image and the learned one from pristine images. In [51], IL-NIQE included a more comprehensive feature set for NSS. Also, a local MVG fitting and patch-wise pooling strategy replaced the global MVG by NIQE. In [56], Liu *et al.* incorporated both structural and free energy based features [57] into the NIQE framework. Later, the same author developed several metrics for either natural images or underwater pictures in [58–60]. In [61], Xue *et al.* extracted the gradient magnitude (GM) and laplacian of gaussian (LoG) as image structural features. Then, taking the effective divisive normalization and joint distribution techniques [55,62], the author regressed the marginal and conditional probabilities of the normalized GM and LoG into quality.

In contrast with previous works which utilized pristine anchors for quality prediction, Min *et al.* [63,64] leveraged the degraded version of the test image as pseudo reference. In [63], three types of distorted pseudo references: JPEG, noise and blur are generated with one distortion level. Then their quality distances are measured based on the detected corner and LBP similarity, respectively. In [64], the author added JP2K as a new distortion type and four more distortion levels. The LBP similarities are computed for such 20 anchors and regressed into quality by SVR.

2.2. Deep learning based IQA metrics

In the deep learning based IQA metrics, Bosse *et al.* first proposed a FR-IQA called WaDIQaM-FR [52] which used the Siamese network to extract features from the distorted and the reference image patches simultaneously. The two extracted feature vectors, along with their differences, are concatenated and fed into a fully connected (FC) network for quality estimation. Another FC-CNN is optional for the calculation of patch weights. WaDIQaM-FR can also be modified into no-reference and the average pooling type by removing certain modules. It achieves about 0.97 in the LIVE database, but only mediocre result in the LIVE challenge.

To handle authentic distortions, Zhang *et al.* proposed a two stream convolutional network named DB-CNN [65]. For simulated distortions, they used a S-CNN which is pretrained on a synthesized databases containing 9 types of distortions, 39 degradation levels with pseudo scores. To deal with authentic distortions, another pretrained, tailored VGG-16 on the Imagenet is adopted. Then the two stream outputs are bilinearly convolved [66] into a vector, and regressed into the quality score. The training process is performed by fine-tuning on target database (*i.e. transfer learning*).

In [53], Su *et al.* proposed a HyperIQA network leveraging the Resnet50 as the backbone structure. Multiscale features are extracted from each stage of the backbone network to capture both local and global image quality attributes. Moreover, the last layer, which encodes semantic features, are passed into a hyper network to produce different weights for diverse image contents. Finally, a simple FC-CNN convolves the hyper weights and biases with multiscale features for quality prediction.

Instead of concatenating multi-level features, and then weighting them as in the HyperIQA, Sun *et al.* [54] hierarchically fused features at four stages of the backbone network through a staircase design. Moreover, an iterative mixed database training (IMDT) strategy is adopted for the feature extraction and fusion part of the network, while the

quality regressor is trained individually for each database. The proposed StairIQA achieves promising results on all six authentic databases.

2.3. Contrast oriented IQA metrics

In [67], Fang *et al.* solved the quality assessment of contrast distorted images based on NSS. Five image statistics: the mean, standard deviation, skewness, kurtosis and entropy are computed from the test image and regressed into the quality. Gu *et al.* [68] considered the contrast distortion problem from the perspective of information loss. The local information is defined as the entropy of unpredicted residuals after joint autoregressive (AR) and bilateral filtering, while the global information loss is measured by the symmetric JS-divergence with a uniform histogram. In [69], Gu *et al.* extracted more contrast relevant features than [67]. Moreover, leveraging the full-reference metric [70] as pseudo scores, a robust regression model is trained against big data.

2.4. Night-time specific IQA (NIQA) metrics

As one of the first attempt in NIQA, Zhai *et al.* [71] addressed this problem by constructing a database called LIEQ, which consists of 1000 low-light enhanced images by applying 10 LIEAs on 100 under-exposed images. In their proposed FR-IQA algorithm, either the multi-exposure fusion or the HDR tone mapping image serves as the pseudo pristine reference. The overall quality is the multiplication of four image quality factors: brightness improvement, color rendition, noise estimation and structural preservation. However, the FR nature limits its application in real world scenarios.

In [72], Lin *et al.* built a larger Low-light Enhanced Image Subjective Dataset (LEISD), where 255 low-light images are enhanced by 8 LIEAs including 6 deep learning based ones, generating 2040 low-light enhanced images in total. In their proposed NIQA algorithm, brightness, color, structure, and naturalness features are extracted and regressed into the quality. However, we note that the original low-light image is involved in the extraction of color and structure features. Thus, it should belong to a variant of RR-IQA, rather than the claimed NR one. Similar misclassification occurred in the NLIEE [73], where the original low-light image is involved in the computation of color and structure features.

In [48], Yang *et al.* constructed a natural night-time database (NNID) consisting of 2240 images captured by three real cameras. An objective quality measure BNBT was also proposed which extracted the weighted mean of the LMN components as brightness and color features; mean, standard deviation of the gray-level co-occurrence matrix (GLCM)'s energy, contrast, homogeneity as texture features at two scales. The SVR is to map features into quality score. In [74], the same author designed three types of features: enhancing-aware, structure-preserving, and colorfulness features for the task of low-light enhancement quality assessment.

In [75], Hu *et.al* proposed a metric called BNIQM which extracted brightness, sharpness, noisiness, etc. as low-level features, and applied a squeeze net [76] on the center of the image for semantic features. By combing low and high-level features, BNIQM outperforms BNBT on the NNID database[48]. However, the cropped 227×227 center may not reflect the quality degradation outside, and thereby the global image appearance.

Recently, Jiang *et al.* [77] proposed a three-component CNN called DDB-Net for the NIQA. Specifically, the first module separates the night-image into **R** and **L**. Content and illuminance related features are extracted from **R** and **L** in the second encoder-decoder module. The third convolves such two features bilinearly [66], and feeds them into a FC-CNN for quality prediction.

3. Database construction and subjective experiments

In this section, we detail how to extend our previous work [49], from the construction process of the night-time photo quality database, to the subjective experiments that help to obtain mean opinion scores of these photos.

3.1. Database construction

The construction of our night-time photo database is an extension from our previous work [49]. Specifically, we select 30 flagship mobile camera units from the market, which cover nearly all kinds of brands and span from year 2015 to 2020. About 15 of them are equipped with the night-mode. Before shooting, we clean the lens of each camera to avoid lens flare. During the shooting process, we switch these cameras to the night-mode if available and hold them as steady as we can. We also open the framing grid and align the perspective of each camera as close as possible. For each test scene, we shoot three photos consecutively and pick the best one for our database (in reality, most three pictures are duplicates of each other, so any of them is feasible).

Based on the previous work [49], we exclude one test scene that failed to trigger night-mode for some cameras. But instead, we add 7 more test scenes. Thus, totally 17 scenes are included in the current database. To ensure that each newly added scene satisfy the truly night conditions, the start shooting time is put off from 5:00 pm to 8:00 pm and lasts until 11:00 pm. Figure 2 shows the 17 scenes in our database. It can be seen that: 1) they all belong to the *truly night* scenes. This requirement is essential for triggering the night-mode correctly, and can test the lightness enhancement ability 2) they also contain highlights such as lamps, neon boards, LED panels, and lightening objects, which helps to discriminate the highlight suppressing quality 3) rich details such as trees and textured walls are included in these scenes. These are for differentiating the detail preserving quality.

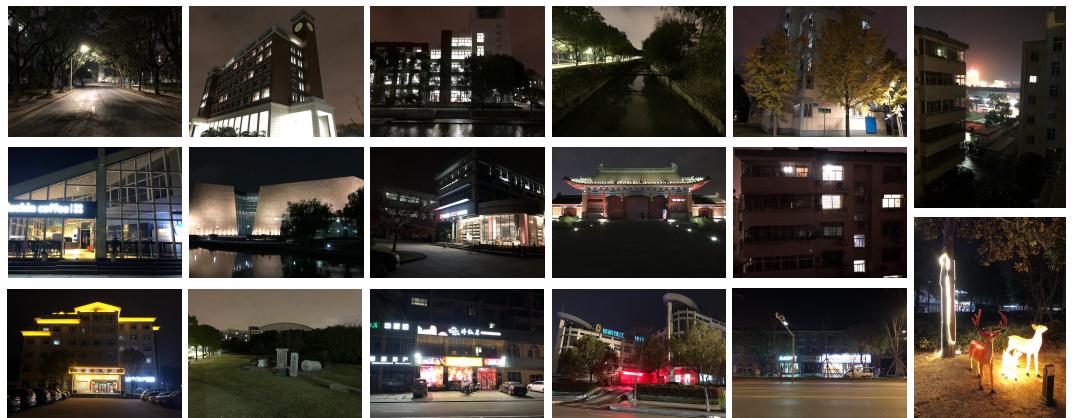


Figure 2. The 17 test scenes in our database.

Table 1 gives a comparison of our Night-time PHoto Database (NPHD) to the existing low-light enhancement and night-time photo quality database. We can see that both the LIEQ, LEISD and EHNQ generate the database by simulating. Although NNID is created by shooting, the amount of involved device is much smaller than ours (3 vs. 30); the simultaneous compared devices per scene are even less (1 vs. 30). Moreover, it adds the number of photos by manually adjusting the exposure level, which doesn't accord with the practical usage scenarios, either (most people shoot in the auto or night mode, both of which expose automatically). By contrast, our NPHD rightly match the real-world usage. As to the number of images, if we ignore such manually exposed images, our database has a larger number of realistic photos than NNID (510 vs. 448). The image resolution is also the highest.

3.2. Subjective experiments

Our subjective experiment is also an extension from our previous work [49], where the quality rank orders have been obtained through the anchor ruler method by three experts. To acquire mean opinion scores of these photos, the most intuitive way is to rate photos from 0 to 100 for every single scene. But there emerges a problem: different scenes have different lighting conditions (e.g. the rank #20 photos in different scene are of different quality), meaning that the quality scores may be inconsistent across different scenes. To

Table 1. Comparison of the NPHD with existing night-time image quality databases

| Database | No.of Scenes | Devices/LIEAs per Scene | Creating method | Resolution | No. of Images |
|------------|--------------|-------------------------|-----------------|---|---------------|
| LIEQ [71] | 100 | 10 | simulating | most 810×540 | 1000 |
| LEISD [72] | 255 | 8 | simulating | 600×400 | 2040 |
| EHNQ [74] | 100 | 15 | simulating | max 1024×1024 | 1500 |
| NNID [48] | 448 | 1 | shooting | max 2048×2048 | 448/2240 |
| NPHD | 17 | 30 | shooting | most 3000×4000 | 510 |

The boldface denotes the preferable items.

solve this problem, we uses the anchor ruler method as in [49]: the expert first chooses one scene that covers the best and worst quality pictures in the database, then picks 4 pictures which are in the middle of the quality range. These 6 photos serve as quality anchors in our database. The only difference from our previous work is that their MOSs are set to 100, 80, 60, 40, 20, 0, respectively. Figure 3 shows the scene and the six quality anchors. 310
311
312
313
314



Figure 3. (a)-(f) serve as six quality anchors in our database. Their MOSs are set to 100, 80, 60, 40, 20, 0, respectively.

Having prepared these anchors, we invite 10 people to rate the rest photos of our database. Since the resolution of the photo is too high to be displayed in a pixel-to-pixel manner, we scale it to fit in the screen initially, but allow the observer to zoom in to scrutinize the details of these photos by themselves. Thus, the overall quality score is the combination of global appearance and local preserving details. During the scoring process, the subject first skim through the 30 photos. Then he/she compares each tested photo to the 6 anchors, and gives their opinion score accordingly. With the six quality anchors as reference, and the assistance of rank orders from our previous work, we believe that people can score with more of certainty. Also, compared to the single stimulus method in [78], our two-stage rating approach (expert/rank orders + people/quality scores) has the dual advantage of making use of expert ranked orders and obtaining the MOSs with ease. The final mean opinion score is the average of 10 individual results. 315
316
317
318
319
320
321
322
323
324
325
326

Figure 4 (a) shows the MOS distribution over the entire database. It can be seen that the MOS covers the whole quality scale, but it neither follows an uniform nor a Gaussian distribution. Instead, it resembles more of a bimodal distribution that has peaks at [30, 40] 327
328
329

and [80, 90]. To explain this phenomenon, Figure 4 (b) shows two typical images falling in such two quality intervals. The left image is taken by a camera with night-mode while the right one is shot in normal mode. Therefore, the two peaks of the MOS distribution reflect the most frequent quality range of cameras equipping with night-mode or not. Lastly, we release the night-time photo database, the quality rank orders, and the MOSs, all available at <https://github.com/hanzongxi/RSIQA>.

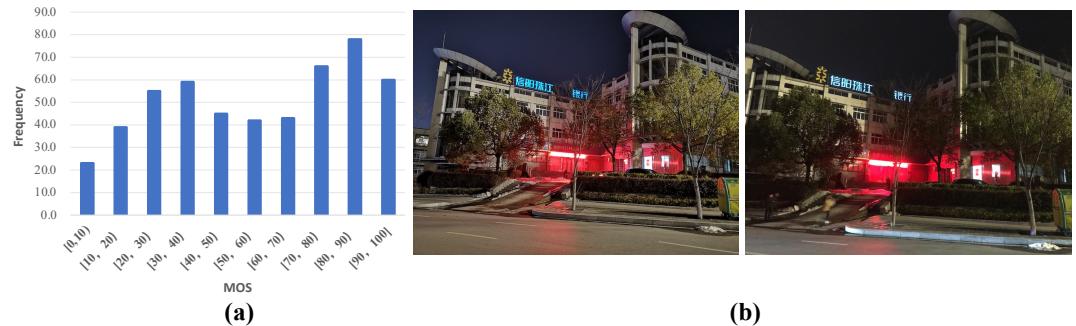


Figure 4. (a) MOS distribution of the database. (b) two sample images whose MOSs lie in the intervals [30, 40], [80, 90], respectively.

4. The proposed metric

In this section, we detail how to perform the region selection process, and which set of features would be extracted from these regions. Four kinds of quality regressors are also described at the end of this section.

4.1. Center and Around Region: Brightness, Contrast, Vignetting, Saturation, and Shading

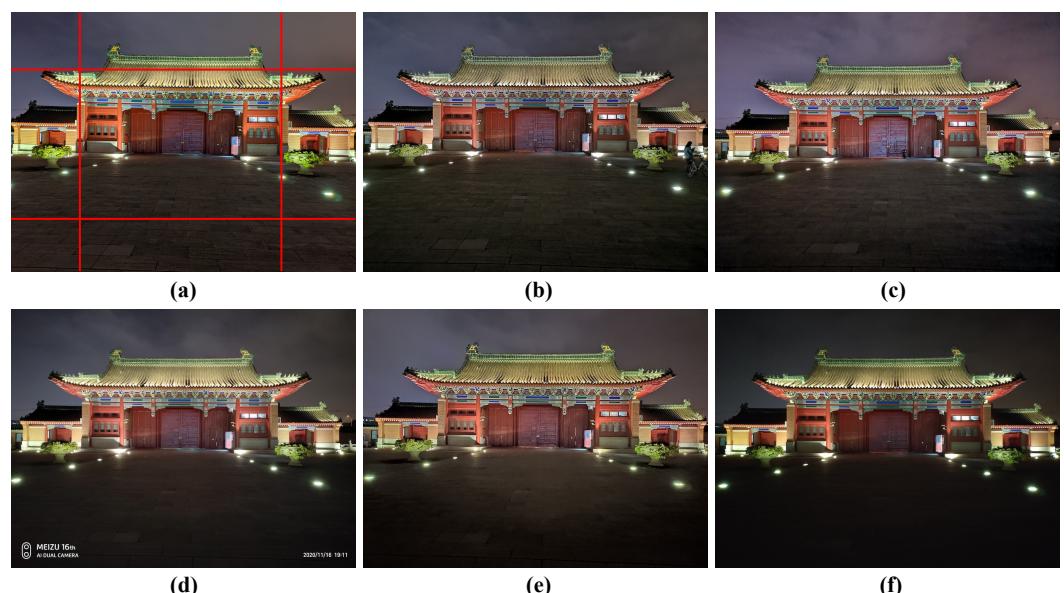


Figure 5. The overlay of the center and around region partition strategy on six night-time photos, where the first five images are taken by night-mode while the last one not. From (a)-(f), their MOSs are: 95, 88, 81, 72, 43, 36; the results of Fang[67] are: 2.83, 2.89, 3.35, 2.86, 3.18, 2.10, respectively.

Figure 5 shows six images overlaid by our center and around partition scheme. The top, bottom, left, right red lines split the image into a center region and four corner regions. Empirically, we choose 1/5 as the partition fraction to adapt to more scenes and photo compositions. To see why such region selective scheme is necessary, the output results of Fang[67] are given. It can be seen that, although the contrast metric can differentiate

between (a-e) and (f), i.e., images taken with and without night-mode, it contradicts the MOSs between (a-b) and (c); (d) and (e). The physical reason is that, the corner of imaging sensor receives less light than the center because of its larger angle with the main optical axis. As a result, the contrast metric mistaken vignetting as an indicator of high contrast/quality.

Therefore, only on the center region do we calculate the contrast related features. Slightly from the metric in [67], we convert the color image into the HSV color space instead of the grayscale, because the V component is closer to the perceptually perceived brightness [4], and has been widely used in LIEAs for illuminance estimation [16,20]. The HSV color space is defined as:

$$V = \max\{R, G, B\}, S = \begin{cases} \frac{\max\{R, G, B\} - \min\{R, G, B\}}{\max\{R, G, B\}} & \text{if } \max\{R, G, B\} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

In the below, we use the symbol V_{center} , $S_{topleft}$, etc to represent either center or corner regions of the V, S components. As in the [67], the mean, variance, skewness, kurtosis of the center image are used to calculate the center brightness, and three contrast indicators:

$$BR_{ce} = \text{mean}(V_{center}) \quad (2)$$

$$C1 = \text{variance}(V_{center}) \quad (3)$$

$$C2 = \text{skewness}(V_{center}) \quad (4)$$

$$C3 = \text{kurtosis}(V_{center}) \quad (5)$$

where *mean*, *variance*, *skewness*, *kurtosis* denotes the sample mean, variance, skewness operator, respectively. Moreover, we calculate the symmetric Jensen-Shannon divergence between the center image and its histogram equalized (HE) version in V channel as another contrast indicator:

$$C4 = D_{JS}(u, v) = \frac{D_{KL}(u \parallel \frac{u+v}{2}) + D_{KL}(v \parallel \frac{u+v}{2})}{2} \quad (6)$$

where D_{KL} refers to the KL-divergence defined by $D_{KL}(u \parallel v) = \sum_i u(i) \log \frac{u(i)}{v(i)}$; $u(i)$ and $v(i)$ are histogram frequencies of the original and HE enhanced image in V channel on the center region. In our experiments, after center cropping, the contrast of the Figure 5 (a)-(e) becomes comparable to each other and much higher than (f). However, it is the corner brightness and the center-corner difference of brightness that determine the quality in Figure 5. Thus, we calculate the corner brightness and luminance vignetting as:

$$BR_{co} = \text{mean}(V_{topleft} + V_{topright} + V_{bottomleft} + V_{bottomright}) \quad (7)$$

$$\text{Vignetting} = \frac{|BR_{co} - BR_{ce}|}{BR_{ce}} \quad (8)$$

Note that although vignetting is barely visible in daylight images for modern cameras [49], it plays a vital role in the quality of night-time photos. Besides brightness and contrast, color information are also easily lost in night scenes. This can be derived from Equation 1 that when $V = 0, S = 0$. Similar to the brightness definition, we calculate the saturation on center and corner regions, and define the color shading as:

$$SA_{ce} = \text{mean}(S_{center}) \quad (9)$$

$$SA_{co} = \text{mean}(S_{topleft} + S_{topright} + S_{bottomleft} + S_{bottomright}) \quad (10)$$

$$\text{Shading} = \frac{|SA_{co} - SA_{ce}|}{SA_{ce}} \quad (11)$$

Notice that our shading definition slightly differs from [79], where *Shading* is calculated in opponent R/G, B/G space. In our experiment, the *Shading* calculated in Figure 5 (a)-(e) are significantly lower than (f). Figure 6 further shows four photos where the first three are taken with night-mode while the last one not. It can be seen the saturation decreases from (a) to (b), and (c) to (d), which demonstrate the effectiveness of *SA* feature. Although the saturation of (b) is the lowest, it obtains the second highest MOS because of well preserved details. We will handle this issue in the third subsection.

Overall, 10 features are extracted under our center and around partition strategy:

$$f_{center_around} = [BR_{ce}, BR_{co}, SA_{ce}, SA_{co}, C1, C2, C3, C4, Vignetting, Shading]$$

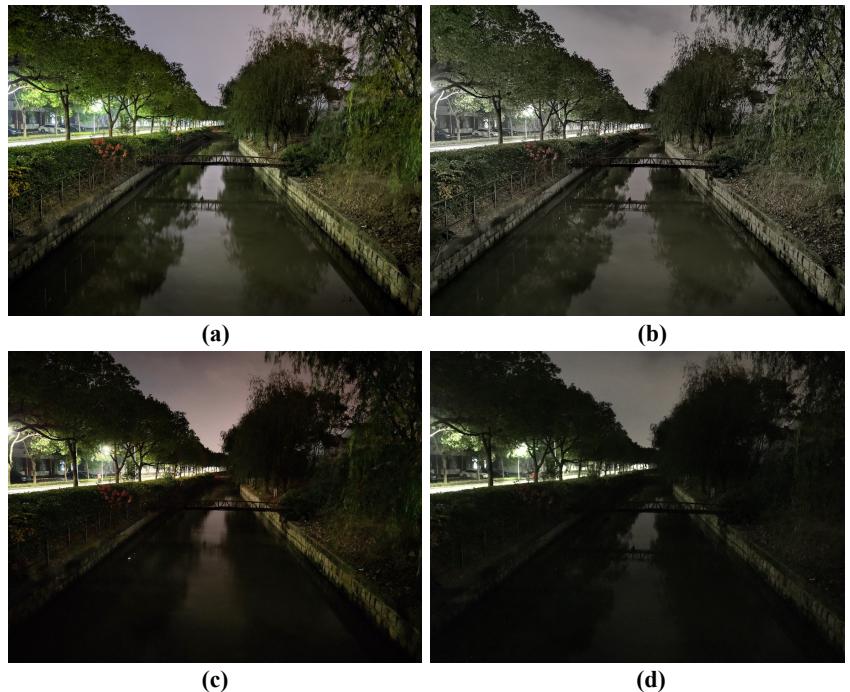


Figure 6. Influence of saturation on quality. From (a)-(d), their MOSs are 94, 83, 32, 14; mean saturation values are 0.4558, 0.3161, 0.4172, 0.3308, respectively.

4.2. Brightest Region: Highlight Suppressing

In addition to the center and around partition, to illustrate why we need to locate the brightest region and extract features that qualify the highlight suppressing capability, Figure 7 shows six images where the mean brightness are close, but their quality degrade from (a) to (c), (d) to (f), due to different highlight suppressing outcomes.

To detect the brightest regions, the RGB image is first converted to the grayscale:

$$I_g = 0.3 \times R + 0.59 \times G + 0.11 \times B \quad (12)$$

where I_g is the grayscale image, and R, G, B represent the red, green, blue channels, respectively. Then we could move a template through the entire image, and search for the area with the largest sum of pixel values. However, since the shapes, sizes, and numbers of highlights in each image are different, it is difficult to determine parameters of templates to detect all highlights at once. Therefore, we resort to the morphological methods.

For preprocessing, a median filter is applied to the I_g to remove obtrusive pixels and smooth the image:

$$I_m(x, y) = median\{I_g(x', y') | (x', y') \in N(x, y)\} \quad (13)$$

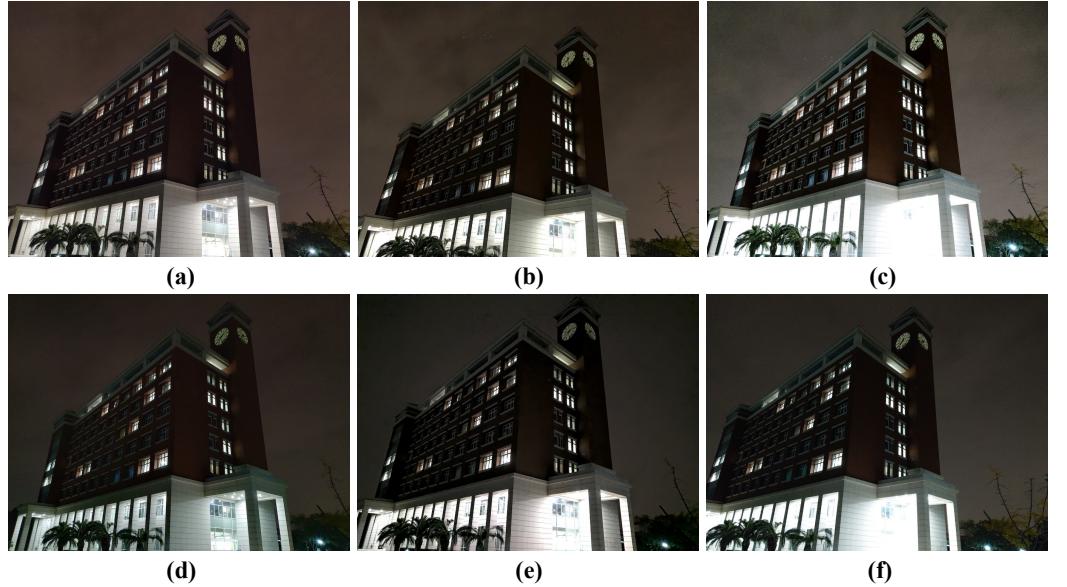


Figure 7. Effect of highlight suppressing on quality. From (a)-(c), (d)-(f), the mean brightness values are close to each other, but their MOSSs decrease due to attenuated highlight suppressing capability.

where I_m is the median filtered image, and $N(x, y)$ is a [5, 3] neighborhood centered on (x, y) . Then a simple thresholding function transforms the grayscale into the binary image:

$$I_b(x, y) = (I_m(x, y) \geq T) \quad (14)$$

where the threshold value T is manually tuned instead of using Ostu's method since the latter always sets the T too low for highlights detection in images such as the four of Figure 8. Different choices of T will be discussed in the implementation details. After obtaining the black and white image, the morphological dilation is applied to thicken skeleton of the white pixels:

$$I_d = I_b \oplus B \quad (15)$$

where \oplus is defined by: $A \oplus B = \{(x, y) | (A)_{(x,y)} \cap B \neq \emptyset\}$. For brevity, we use the same symbol I_b to represent either the set of white pixels or the image I_b itself. B is a rectangle structuring element of size [15, 15] in our experiments.

After dilation, we can obtain a series of connected components $R = \{R_1, R_2, \dots, R_n\}$ of the highlights (white pixels) by the 8-connectivity:

$$(p_1, p_n) \in R_i \quad (16)$$

if there exist a path $(p_1, p_2, \dots, p_{n-1}, p_n)$ that satisfies $(p_k, p_{k+1}) \in N_8$, where $p_n \in I_d$, R_i is the i -th region proposal and N_8 refers to the 8-connected neighborhood.

Figure 8 shows four images with their binary image I_d and the boundary mask of R . It can be seen that the thresholding and morphological segmentation successfully detect various shapes and numbers of highlights, both between the same scene and across different scenes. To quantitatively measure the highlight suppressing capability, we calculate five features: the ratio of over-exposed pixels, the weighted mean brightness in R , the mean saturation in R , the average total variation in R , and the entropy of pixel values in R . The ratio of over-exposed pixels is given by:

$$Ra = \frac{N_R}{N} \quad (17)$$



Figure 8. Top row: original images. Middle row: binary images I_d . Bottom row: boundary masks R .

where N_R, N refers to pixel numbers in R and I_g . To calculate the weighted mean brightness, we first sort $V(x, y)$ in R by ascending order: $i = (1, 2, \dots, N_R)$. Then the weight for the i -th order of $V(x, y)$ is:

$$w_i(x, y) = \log(1 + \frac{i(x, y)}{N_R}) \quad (18)$$

The log function is chosen due to its relationship between subjective brightness and objective lightness [4]. The weighted mean brightness and saturation in R are computed as:

$$Br = \frac{\sum_{(x,y) \in R} V(x, y) w_i(x, y)}{\sum_{(x,y) \in R} w_i(x, y)}, \quad Sa = \frac{\sum_{(x,y) \in R} S(x, y)}{N_R} \quad (19)$$

where (x, y) is the 2D pixel indexes in R ; V, S are channels of I_g predefined in Equation 1.

The mean total variation (TV) is calculated as:

$$TV = \frac{\sum_{(x,y) \in R} \sum_{(x',y') \in N_4(x,y)} |I_g(x, y) - I_g(x', y')|}{N_R} \quad (20)$$

where (x', y') belongs to the 4-connected neighbor of (x, y) . The discrete entropy of pixel values in R is defined by:

$$H = - \sum_{k=0}^{255} P(I_g(x, y) = k) \log P(I_g(x, y) = k), \quad (x, y) \in R \quad (21)$$

In practice, the histogram and sample frequency are used to approximate the probability $P(I_g(x, y) = k)$. Figure 9 shows three images whose highlight suppressing capability declines from (a) to (c). Their five computed features are also given. We can find that the Br, Ra increase, while Sa, TV, H decrease from (a) to (c). It is worth noting that if the logarithm weighting strategy were omitted, the Br of Figure 9 (a) would probably be higher than (b), due to the slightly brighter deer of the left (indicated by red arrows). Thus, the logarithm weighting is necessary for calculating the perceived over-exposure. Overall, 5 measures consist of the highlight suppressing feature:

$$f_{highlight_suppressing} = [Ra, Br, Sa, TV, H] \quad (22)$$

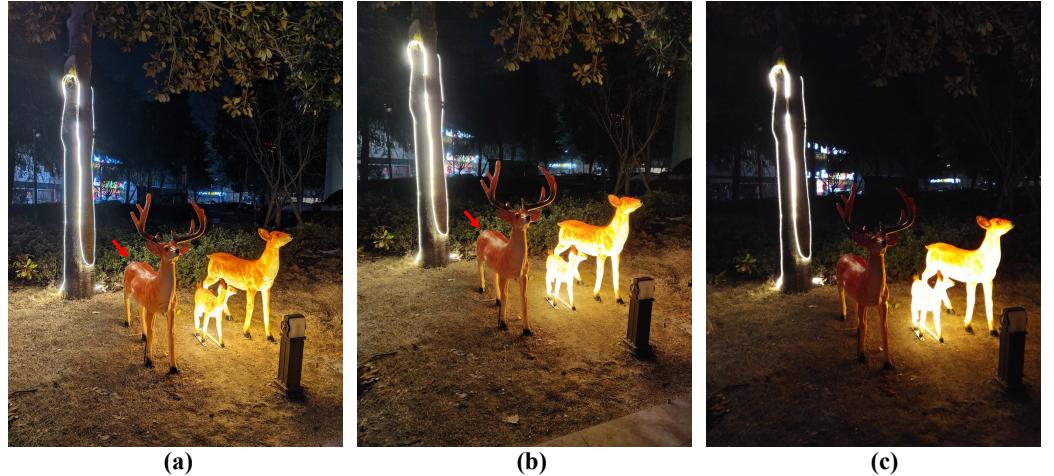


Figure 9. Measures of the highlight suppressing capability. (a) Ra, Br, Sa, TV, H=0.0257, 0.9794, 0.3566, 0.072, 3.4169; (b) Ra, Br, Sa, TV, H=0.0343, 0.9821, 0.2079, 0.0317, 2.8394; (c) Ra, Br, Sa, TV, H=0.0449, 0.9921, 0.1555, 0.0198, 2.0243.

4.3. Foreground and Sharpest Region: Preserving Details, Naturalness, Noise, and Image Structure

As stated in [49], the quality difference between photos with night-mode mainly lies in the details. In search of regions where detail quality should be accessed, we consider two factors: the foreground and the texture prior. The former has been widely applied in saliency detection; the latter is used because people prefer to determine the sharpness level according to the sharpest region of the image [80], and image structure should be compared in the same textured regions to ensure consistency. In our implementation, the center bias is to represent the foreground prior, which is defined as:

$$G_\sigma(x, y) = \frac{\exp\left\{-\frac{1}{2}[(x - x_0^2) + (y - y_0^2)]\right\}}{2\pi\sigma^2} \quad (23)$$

where (x_0, y_0) is the center coordinate, the σ is set so that 3σ equals half height of the image. The local phase map (LPC) is used for the texture prior, which is calculated based on [80]:

$$LPC(x, y) = \frac{\sum_{l=1}^L |c_{i,l}| \cos(\Phi(\{c_{i,l}\}) - \hat{\Phi}(\{c_{i,l}\}))}{\sum_{l=1}^L |c_{i,l}| + K} \quad (24)$$

where L is the number of orientations in the steerable pyramid decomposition of I_g , $\Phi(\{c_{i,l}\})$ is the phase of the i -th coefficient in the finest subband of the l -th orientation, $\hat{\Phi}(\{c_{i,l}\})$ is the corresponding predicted phase, and K is a small positive constant. Our importance map is the multiplication of $G_\sigma(x, y)$ and $LPC(x, y)$:

$$M(x, y) = LPC(x, y) * G_\sigma(x, y) \quad (25)$$

To obtain the foreground and sharpest region, we crop $M(x, y)$ by convolving it with a kernel whose values are all 1s, and finding the maximum crop location. The convolution can be efficient done in the frequency domain. The kernel size is set to 0.3 of the width and height of the image to adapt more scenes. Figure 10 shows three images with their $M(x, y)$, and the red rectangles indicate the final region proposals. Notice that our subsequent feature extraction processes will only be performed on such regions, which can be seen as a re-weighting strategy, and also reduces the computational time[81].

Technically speaking, the quality of details consists of two factors: edge acutance (EA), and texture resolution (TR)[82,83]. Following the work in [83], we use the normalized

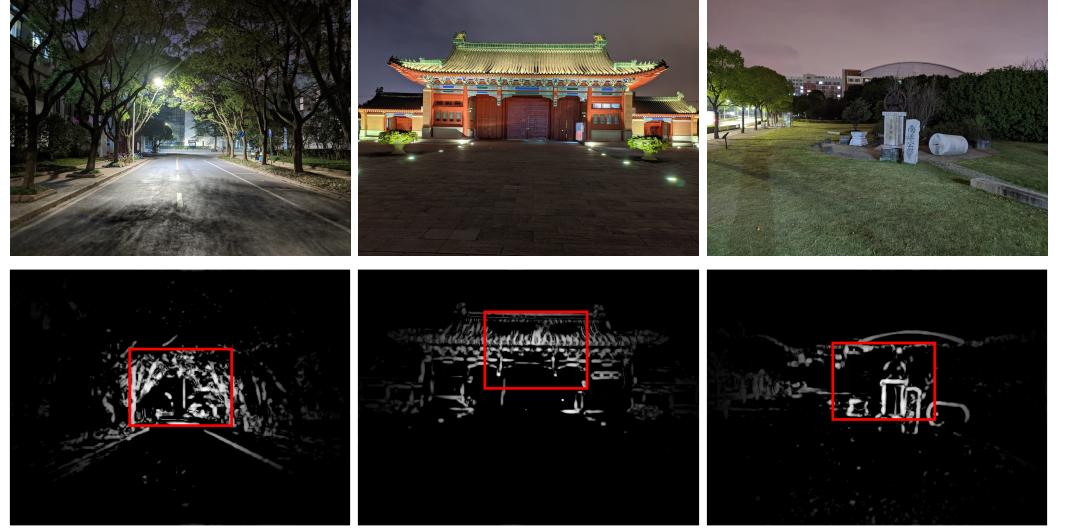


Figure 10. Top row: original images. Bottom row: Importance maps $M(x, y)$ and the region proposals.

energy of predicted/reconstructed signals to represent the edge acutance (EA). Let I_r be the
region proposal chosen from I_g , a patch $\mathbf{x}_k \in \mathbb{R}^B$ of size $\sqrt{B} \times \sqrt{B}$ is extracted from I_r by:

$$\mathbf{x}_k = \mathbf{R}_k(I) \quad (26)$$

where $\mathbf{R}_k(\cdot)$ copies the pixel values from image I at location k into \mathbf{x}_k , $k = 1, 2, 3 \dots N$. N
is the total number of image patches. The reconstructed version of I_r can be obtained by
sparse coding:

$$\boldsymbol{\alpha}_k^* = \underset{\boldsymbol{\alpha}_k}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x}_k - \mathbf{D}\boldsymbol{\alpha}_k\|_2 + \lambda \|\boldsymbol{\alpha}_k\|_0 \quad (27)$$

$$I'_r = \sum_{k=1}^N \mathbf{R}_k^T(\mathbf{D}\boldsymbol{\alpha}_k^*) ./ \sum_{k=1}^N \mathbf{R}_k^T(\mathbf{1}_B) \quad (28)$$

where \mathbf{D} refers to off-the-shelf 2-D orthonormal DCT bases, \mathbf{R}_k^T is the reverse operator of \mathbf{R}_k ,
 $\mathbf{1}_B$ is the all $\mathbf{1}$ vector of size B . Since the divisive normalization technique has been proved
effective in [55,61] to handle heterogeneous image contents, we normalize the energy of the
predicted signal by its variance, which defines the edge acutance(EA):

$$EA = \sum_{k=1}^N \frac{(\mathbf{x}'_k)^T \mathbf{x}'_k}{\sigma_k^2} = \sum_{k=1}^N \frac{(\mathbf{D}\boldsymbol{\alpha}_k^*)^T \mathbf{D}\boldsymbol{\alpha}_k^*}{\sigma_k^2} = \sum_{k=1}^N \frac{(\boldsymbol{\alpha}_k^*)^T \boldsymbol{\alpha}_k^*}{\sigma_k^2} \quad (29)$$

where σ_k^2 refers to the variance of \mathbf{x}_k . In practice, we operate in the gradient domain instead
of image domain to calculate EA. For calculating the texture resolution (TR), we first remove
 I'_r from I_r for texture components. Then the entropy of the residuals is computed to reflect
texture density/resolution:

$$I_{res} = |I_r - I'_r| \quad (30)$$

$$TR = - \sum_{i=0}^{255} p(i) \log_2 p(i) \quad (31)$$

where $p(i)$ gives the probability density of the i -th gray-scale in I_{res} . Figure 11 shows three
images where the quality degrade from (a) to (c). The EA and TR calculated on their region
proposals are also given. It can be seen that both EA and TR are sensitive to the detail
losses from (d) to (f). However, the multi-frame stacking technique in night-mode can lead

to over-sharpening details, for which both EA and TR may over-estimate. To alleviate this problem, we leverage the NIQE metric as a deputy of image naturalness, since it has been found that the kurtosis feature in NIQE emphasizes image smoothness/softness, rather than pure sharpness [83]. In fact, this phenomenon is also implicitly demonstrated in [44]. Let \mathbf{y} and \mathbf{z} be the 36-D NIQE feature vectors extracted from the \mathbf{x}_k and a corpus of natural images, we first fit them by the MVG function:

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{k/2}|\Sigma_y|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_y)^T \Sigma_y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right\} \quad (32)$$

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{k/2}|\Sigma_z|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_z)^T \Sigma_z^{-1} (\mathbf{z} - \boldsymbol{\mu}_z) \right\} \quad (33)$$

Then naturalness (NS) is defined as the symmetric JS-divergence between $f(\mathbf{y})$ and $f(\mathbf{z})$:

$$NS = D_{JS}(f(\mathbf{y}), f(\mathbf{z})) = D_{KL}(f(\mathbf{y}) \parallel \frac{f(\mathbf{y}) + f(\mathbf{z})}{2}) + D_{KL}(f(\mathbf{z}) \parallel \frac{f(\mathbf{y}) + f(\mathbf{z})}{2}) \quad (34)$$

$$= \sqrt{(\boldsymbol{\mu}_y - \boldsymbol{\mu}_z)^T \left(\frac{\Sigma_y + \Sigma_z}{2} \right)^{-1} (\boldsymbol{\mu}_y - \boldsymbol{\mu}_z)} + const \quad (35)$$

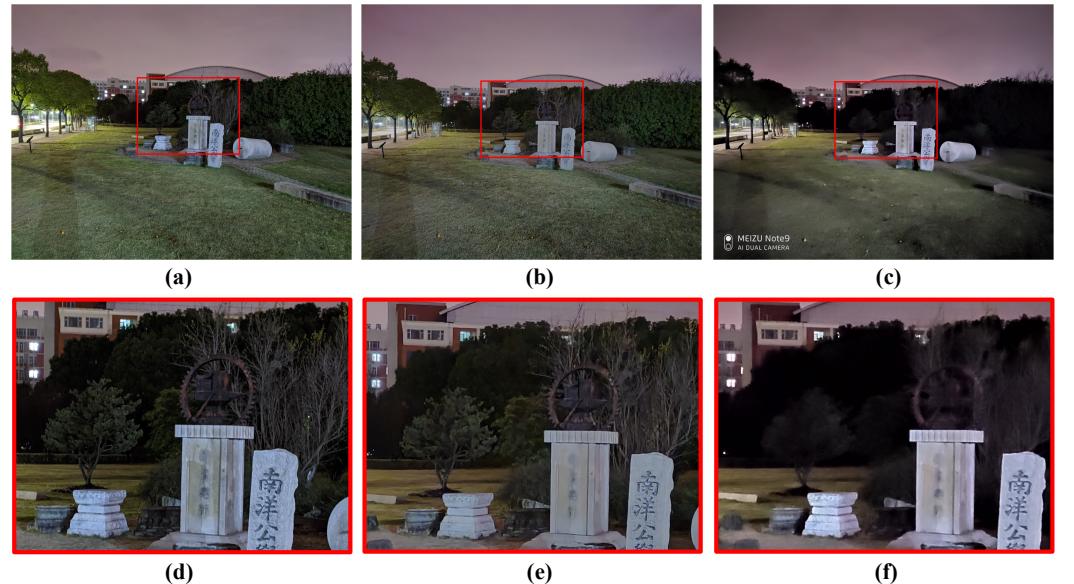


Figure 11. Quality difference of preserving details between night-time photos. (a)-(c) are original images. (d)-(f) are close-ups of the region proposals in (a)-(c). Please enlarge (d) and (e) to see more details. The EA and TR calculated on (d)-(f) are: (d) 11.6267, 4.3754; (e) 11.2271, 3.4063; (f) 7.5099, 2.5515. The energy, contrast, homogeneity of GLCM calculated on (d)-(f) are: (d) 0.1846, 0.2033, 0.9058; (e) 0.1916, 0.1292, 0.9428; (f) 0.2718, 0.0816, 0.9650;

Although the noise can be reduced by night mode algorithm, the noise still appears in photos without night-mode. To estimate the variance of noise (VN), we resort to the approach utilizing scale invariance of the marginal filter response[84]:

$$\hat{\kappa}_x, \hat{\sigma}_n^2 = \underset{\kappa_x, \sigma_n^2}{\operatorname{argmin}} \sum_{i=2}^{N^2} \left| \frac{\kappa_x - 3}{(1 + \frac{\sigma_n^2}{\hat{\sigma}_{y_i}^2 - \sigma_n^2})^2} + 3 - \hat{\kappa}_{y_i} \right| \quad (36)$$

where $\hat{\sigma}_{y_i}^2$ and $\hat{\kappa}_{y_i}$ refers to the variance and kurtosis of $N \times N$ DCT filtered response. The estimated $\hat{\sigma}_n^2$ is used as the VN feature.

For image structure, we leverage the classic gray-level co-occurrence matrix (GLCM) [4,85]. Let $p(i,j)$ be the joint probability of the 8×8 GLCM for 8-bit I_r . The three important statistics of GLCM are defined as [86]:

$$Energy = \sum_{i,j} p(i,j)^2 \quad (37)$$

$$Contrast = \sum_{i,j} |i - j|^2 p(i,j) \quad (38)$$

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (39)$$

Note that the *Contrast* defined here differs from the aforementioned C1 – C4 in that the *Contrast* emphasizes more of textural variation [85], rather than brightness change. Figure 11 gives *Energy*, *Contrast*, *Homogeneity* of (d)-(f). It can be found that the *Energy*, *Homogeneity* increase, while the *Contrast* decrease as the image structure becomes flat from (d) to (f).

As shown in Figure 3, 6, and 9, the corner acutance (CA) also affects the quality of night-time photos. Thus, we simply average four corners of the LPC map to define CA:

$$CA = mean(LPC_{topleft} + LPC_{topright} + LPC_{bottomleft} + LPC_{bottomright}) \quad (40)$$

where $LPC_{topleft}$ etc, refers to the corresponding region of the LPC map computed by Equation 24. As mentioned in section 3, people magnify the image to scrutinize details in the subjective experiment. Similar to this behavior, we downsample the I_g and extract the above features in two scales. Thus, the $f_{sharpness}$ computed in this sub-section consists of 16 features, consisting of 8 features: [$EA, TR, NS, VN, Energy, Contrast, Homogeneity, CA$] in each scale. Finally, our 31-D feature vector is the concatenation of three sets of features computed in three genres of region proposals, qualifying the brightness (contrast, saturation, vignetting, shading), highlight suppressing, and detail preserving capability:

$$f_{overall} = [f_{center_around}, f_{highlight_suppressing}, f_{sharpness}] \quad (41)$$

4.4. Quality Regressors: SVR, Decision Tree, Random Forest, AdaBoost.R2

In the IQA literature, the SVR has become the *de facto* tool for quality regression, despite that other regressors such as random forest and neural network are also leveraged in [87,88]. In this paper, four regressors are chosen as our candidates: SVR, decision tree, random forest, and AdaBoost.R2. Specifically, the objective of SVR is to find a hyperplane which creates the maximum margin. The decision tree greedily chooses each splitting feature and the splitting point based on the mean squared error until the minimum leaf node sizes are reached. Both random forest and AdaBoost pertain to the ensemble learning [89]. However, their aims are different: the random forest [90] tries to reduce the inherent instability, or the high variance of the decision tree by bagging and random feature selection, while the AdaBoost family (M1, M2, MH, MR, SAMME, R2) [91–94] improve the prediction precision of weak learners by iteratively training the learner on reweighted samples (giving more weight to large error samples), and aggregating them according to the prediction accuracy. In the below, the pseudo codes of linear SVR, random forest, AdaBoost.R2 are shown in Algorithm 1, 2, 3.

5. Experimental results

In this section, we first give the implementation details, then successively validate our proposed region selective model through comparison of four quality regressors, ablation study, and comparison with 17 state-of-the-art NR algorithms. The SROCC, KROCC, PLCC, RMSE are used as default criteria.

Algorithm 1 Linear SVR

Input the training set T of size n : $\{(x_1, y_1), \dots, (x_n, y_n)\}$, a cost parameter C to balance the model weights and the soft margin.

Solve the following minimization problem by introducing two slack variables ξ_i and ξ_i^* :

$$\begin{aligned} & \min \frac{1}{2} \|\omega\|^2 + C \left(\sum_{i=1}^n \xi_i + \sum_{i=1}^n \xi_i^* \right) \\ & \text{s.t. } \langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i \\ & \quad y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \xi_i^* \\ & \quad \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n, \end{aligned}$$

Output the hyperplane: $f_{svr}(x) = \langle \omega, x \rangle + b$.

Algorithm 2 Random Forest

Input the training set T of size n : $\{(x_1, y_1), \dots, (x_n, y_n)\}$, the number of trees B , the number of variables randomly sampled at each split m , and the minimum node size n_{min} .

For $b = 1, \dots, B$:

- 1: Draw a bootstrap sample Z^* of size n from the training set without replacement;
- 2: Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached:
 - 2.a: Select m variables at random from the p variables/features.
 - 2.b: Pick the best variable/split-point among the m .
 - 2.c: Split the mode into two daughter nodes based on mean squared error.

return the ensemble of trees $\{T_b\}_1^B$.

Output the average of all trees: $f_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$.

Algorithm 3 AdaBoost.R2

Input the training set T of size n : $\{(x_1, y_1), \dots, (x_n, y_n)\}$, maximum number of iterations N , and a base learning algorithm \mathcal{L} ; Unless otherwise specified, set the initial weight vector \mathbf{w}^1 such that $w_i^1 = 1/n$ for $1 \leq i \leq n$

For $t = 1, \dots, N$:

- 1: Call Learner \mathcal{L} with the training set T and the distribution \mathbf{w}^t , and get a hypothesis $h_t : X \rightarrow \mathbb{R}$;
- 2: Calculate the adjusted error e_i^t for each instance: let $D_t = \max_{j=1}^n |y_j - h_t(x_j)|$, then $e_i^t = |y_j - h_t(x_j)| / D_t$;
- 3: Calculate the adjusted error of h_t : $\epsilon_t = \sum_{i=1}^n e_i^t w_i^t$; if $\epsilon_t^t \geq 0.5$, stop and set $N = t - 1$;
- 4: Let $\beta_t = \epsilon_t / (1 - \epsilon_t)$;
- 5: Update the weighted vector: $w_i^{t+1} = w_i^t \beta_t^{1-e_i^t} / Z_t$, where Z_t is a normalizing constant.

Output the hypothesis: $h_f(x) = \text{the weighted median of } h_t(x) \text{ for } 1 \leq t \leq N$, using $\ln(1/\beta_t)$ as the weight for hypothesis h_t .

5.1. Implementation details

For 8-bit image, the reasonable range of T is 170 to 230. Lower T helps to detect more highlight areas while higher T distinguishes B_r better. In this paper, we set $T = 200$. To implement the four regressors, we use the *scikit-learn* package in Python [95]. For the SVR, we choose the radial basis function (RBF) $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ with $\gamma = 2^{-6}$ and $C = 128$ instead of the linear kernel. For the decision tree, the classification and regression tree (CART) with no cost-complexity pruning is adopted [96] since it can receive qualitative inputs and adapt to our regression task. The minimum leaf node size is constrained to 10 to avoid overfitting. The minimum leaf size of each tree in random forest is set to 5. The number of selected features m equals $p/3 = 10$. The number of trees in random forest is set to 500. For the weak base learner of the AdaBoost.R2, we choose the decision tree with the minimum leaf size set to 15. The number of iterations of AdaBoost.R2 is 100. To ensure complete reproducibility and benefit academia, we share all source & test codes, scripts, with detailed test data & results at <https://github.com/hanzongxi/RSIQA>.

5.2. Comparison of SVR, Decision Tree, Random Forest, AdaBoost.R2

To compare the performance of different quality regressors, we partition the database into a training set and testing set. This process is repeated 10 times. Table 2 shows the average SROCC, KROCC, PLCC. It can be seen that the performance of decision tree is at the bottom of the four regressors. More splits can lead to a boost in the training set (when setting minimum leaf sizes to 1, the training accuracy is 100%), but a drop in the testing set. In contrast, the random forest achieves the highest SROCC because it allows each decision tree to grow deeper, and avoids overfitting by bagging and random feature selection. The AdaBoost.R2 also shows better generalization capability than a single decision tree, but is inferior to the random forest. Considering the superiority of random forest, it is leveraged for regression for all algorithms in the remaining experiments.

Table 2. Performance comparison of four regressors on NPHD.

| Criteria | SVR | Decision tree | Random forest | AdaBoost.R2 |
|----------|---------|---------------|----------------|-------------|
| SROCC | 0.7481 | 0.6663 | 0.8053 | 0.7578 |
| KROCC | 0.5540 | 0.5033 | 0.6124 | 0.5631 |
| PLCC | 0.7953 | 0.6981 | 0.8345 | 0.8057 |
| RMSE | 16.1353 | 19.2989 | 14.6970 | 15.9243 |

5.3. Ablation test of individual feature set

To demonstrate the complementarity of each feature set, Table 3 shows results of regressing single f_{center_around} , $f_{highlight_suppressing}$, or $f_{sharpness}$. It can be seen the sharpness feature set performs slightly better than brightness related feature set, which is somewhat unexpected. The highlight suppressing feature set performs worst among the three, but still outperform all traditionally general purpose metrics, which will be reported in Table 5. By combining all three feature sets, the $f_{overall}$ performs significantly better.

Table 3. Prediction performance of three individual feature sets on NPHD.

| Features | SROCC | KROCC | PLCC | RMSE |
|------------------------------|---------------|---------------|---------------|----------------|
| f_{center_around} | 0.6799 | 0.5072 | 0.7027 | 19.1622 |
| $f_{highlight_suppressing}$ | 0.4965 | 0.3456 | 0.5395 | 23.0446 |
| $f_{sharpness}$ | 0.6913 | 0.5025 | 0.7334 | 18.2345 |
| $f_{overall}$ | 0.8053 | 0.6124 | 0.8345 | 14.6970 |

5.4. Computational time and performance comparison between algorithms with and without region selective scheme

To demonstrate the superiority of our region selective strategy, we experiment with the same feature sets but without region selective module. Table 4 shows computation time and performances of algorithms using region selective approach or not. It can be seen that through our region selective approach, the performance boost by a decent margin, while the spent time reduces by a huge amount. The former is due to the effectiveness of local brightness, saturation and local highlight features for NIQA; the latter is because of the efficient local sharpness extraction for high-resolution photo.

Table 4. Computational time and performance comparison between algorithms with and without region selective scheme.

| Strategy | Time(sec) | SROCC | KROCC | PLCC | RMSE |
|-----------------------|--------------|---------------|---------------|---------------|----------------|
| w.o region selection | 1425.14 | 0.7759 | 0.5954 | 0.8145 | 15.7802 |
| with region selection | 89.07 | 0.8053 | 0.6124 | 0.8345 | 14.6970 |

5.5. Performance comparison of RSIQA with 17 state-of-the-art IQA metrics

To demonstrate the advantage of RSIQA, we compare it with 17 NR-IQA metrics, including NIQE, BRISQUE, IL-NIQE, SNP-NIQE, M3, BPRI, BMPRI (conventionally general purpose type), WaDIQaM-NR, DB-CNN, HyperIQA, StairIQA (deep learning type), Fang, NIQMC, BIQME (contrast oriented type), and BNBT, BNIQM, BEHN (night specific type). Source codes of first three types of IQA can be found on the github except the night-specific one. So we implement the BNBT, BNIQM, BEHN by ourselves. To ensure fairness, we partition the NPHD into a training set and test set, and retrain them. For deep learning based IQA metrics, we use the LIVE-NR-weighted model for WaDIQaM, the LIVE model for DB-CNN, the Koniq10k model for HyperIQA, and the ensemble model for StairIQA. We don't retrain these metrics on NPHD because the database size is relatively small; some don't provide training code (e.g. WaDIQaM), or it is time-consuming to incorporate our database into the training process (e.g. StairIQA-ensemble).

Table 5. Performance comparison of RSIQA with 17 state-of-the-art algorithms on NPHD.

| Metrics | SROCC | KROCC | PLCC | RMSE |
|-----------------|---------------|---------------|---------------|----------------|
| BRISQUE [55] | 0.1224 | 0.0810 | 0.3008 | 26.6465 |
| NIQE [50] | 0.1527 | 0.1031 | 0.1810 | 27.4807 |
| ILNIQE [51] | 0.3334 | 0.2298 | 0.3822 | 25.8192 |
| SNP-NIQE [56] | 0.3961 | 0.2767 | 0.4184 | 25.3772 |
| M3 [61] | 0.1252 | 0.0858 | 0.2024 | 27.3617 |
| BPRI [63] | 0.2628 | 0.1790 | 0.3330 | 26.3460 |
| BMPRI [64] | 0.1592 | 0.1067 | 0.2242 | 27.2290 |
| Fang [67] | 0.6039 | 0.4348 | 0.6405 | 21.4575 |
| NIQMC [68] | 0.4115 | 0.2853 | 0.4909 | 24.3420 |
| BIQME [69] | 0.5665 | 0.4007 | 0.6199 | 21.9248 |
| WaDIQaM [52] | 0.0115 | 0.0109 | 0.0339 | 27.9242 |
| DB-CNN [65] | 0.5118 | 0.3613 | 0.5285 | 23.7200 |
| HyperIQA [53] | 0.6961 | 0.5082 | 0.7087 | 19.7126 |
| StairIQA [54] | 0.6778 | 0.4855 | 0.6954 | 20.0781 |
| BNBT [48] | 0.7337 | 0.5455 | 0.7561 | 17.6824 |
| BNIQM [75] | 0.7698 | 0.5805 | 0.8085 | 15.9520 |
| BEHN [74] | 0.7405 | 0.5518 | 0.7623 | 17.0744 |
| RSIQA(proposed) | 0.8053 | 0.6124 | 0.8345 | 14.6970 |

Table 5 gives comparative results of the proposed RSIQA with 17 state-of-the-arts. The best-performing metrics in each category are highlighted in boldface. Several meaningful observations can be derived:

1) The general-purpose IQA metrics produce consistently poor results in realistic NPHD. By adding more features, the SNP-NIQE [56] and IL-NIQE [51] improves original NIQE [50] in NPHD. The BPRI and BMPRI also perform not effectively because the anchors used only contain traditional distortion types, which don't include the brightness, contrast distortion type.

2) The contrast oriented metric achieve consistently decent results on NPHD, among which the Fang's method [67] achieves the best. This demonstrates the effectiveness of leveraging contrast related feature into the NIQA task. However, as mentioned in section 4.1, the drawbacks of applying contrast metrics in NIQA lie in their global operating manner, and lack of other measures such as saturation, highlight suppressing, and sharpness related features.

3) In contrast with its exceptional performance on LIVE (SROCC=0.97), the outcome of WaDIQaM is almost uncorrelated with MOS (SROCC=0.0115). The main reason is that the trained LIVE database lacks brightness-altered images; another is that WaDIQaM uses shallower neural network and extracts lower-level features. By including brightness-altered images in its pretraining set, DB-CNN achieved mediocre result in NPHD. Both HyperIQA and StairIQA exhibit significantly better results than generally-purpose and contrast-oriented type. This demonstrates the more adaptability of deep learning based metrics when training on realistic database such as Koniq10k or the ensemble.

4) Not surprisingly, all four NIQA show better results than the former three types of IQA. Within them, the BNBT metric performed worst, probably because it doesn't include contrast features, which are essential in differentiating between images with and without night-mode. The BNIQM performed better than BEHN because the former additionally considers the sharpness features, which play a vital role in differentiating high-quality and mid-range night-mode photos, as shown in Table 3. Our proposed RSIQA surpasses BNIQM by three main reasons. First, random forest helps to improve SROCC from 0.4654 (not reported in this paper, please check our github) to 0.7698 for BNIQM, from 0.7481 to 0.8053 for RSIQA, as shown in Table 2. Second, the combination of three different feature sets (31-D) are more effective than the cumbersome 1006-D feature in BNIQM (0.7698 to 0.7759 in SROCC, as shown in Table 4 and 5). Third, the effective region selective strategy helps to improve the SROCC of RSIQA from 0.7759 to 0.8053, as shown in Table 4.

5.6. Performance validation and computational time of RSIQA on NNID

To further validate RSIQA, we choose the class-leading BNIQM as competitor and NNID 2048 subset as the target database, since it is also realistic and of high resolution. Table 6 gives the feature extraction time, and performance of RSIQA with BNIQM on NNID using either SVR, SVR after feature normalization, or random forest as regressor, which is listed from top to the bottom in each column of the performance criteria. From Table 6, we have following observations:

1) Our RSIQA consistently exceeds the BNIQM in NNID by each regressor, which demonstrates the superiority of our feature design; 2) Our 31-D features save more time than the 1006-D feature extracted by BNIQM; 3) In the case of SVR, RSIQA outperforms the BNIQM by the largest margin. However, the gap diminishes when levering normalization or random forest. This demonstrates that either normalization or random forest is necessary in regressing high dimensional features (1006-D) on a small dataset (450 images). 4) Using the same random forest, both BNIQM and RSIQA achieves a boost on NNID over our NPHD (0.9 vs. 0.8). This demonstrates the very challenging nature of the NPHD since it is created by different devices, not by just tuning EV.

Table 6. Computational time and performance comparison between RSIQA and BNIQM on NNID .

| Metrics | Time(sec) | SROCC | KROCC | PLCC | RMSE |
|---------|-----------|--------|--------|--------|--------|
| BNIQM | 17.30 | 0.6570 | 0.4875 | 0.4750 | 0.1594 |
| | | 0.8632 | 0.6869 | 0.8562 | 0.0874 |
| | | 0.9090 | 0.7457 | 0.9173 | 0.0716 |
| RSIQA | 6.06 | 0.8817 | 0.7051 | 0.8939 | 0.0802 |
| | | 0.9052 | 0.7083 | 0.9245 | 0.0698 |
| | | 0.9233 | 0.7650 | 0.9339 | 0.0650 |

5.7. Future work

Despite the success of the proposed algorithm on the NPHD, we could extend our database and improve the metric in three ways: (1) with advances of both optics and ISPs, previously unusable night-time portraits/zoom photos have become mature in some mobile units. Since portrait quality heavily relies on face exposure, skin tone and facial details, we may include the night-time zoom photos and the face detection & assessment next; (2) aside from under-exposure, vignetting, detail loss, etc, real night-time photos also suffer from artifacts such as lens flare and multi-frame stacking induced ghosting. How to detect and assess these artifacts will also be our research interest. (3) although the bilinear convolution [66] of DB-CNN [65] and DDB-Net [77] shows decent results on night-time photos. In essence, the DB-CNN bilinearly convolves two effective features of the night-time photos: brightness (training with brightness-altered images) and texture (training with noisy, blurry, JPEG compressed images of LIVE). Thus, we could design/train a network that incorporates the region proposal and more effective features such as highlight suppressing in the future. The transfer learning may also be used to solve our small training data problem.

6. Conclusions

At the beginning of article, the quality gap between LIEA images and the real night-time photos is introduced. To solve the problem of quality assessment of night-time photos, we first build a database which contains 510 images shot by 30 devices in 17 scenes. In the proposed objective quality metric, We show that traditional contrast metric should be applied in the center and corner regions to adapt to the quality assessment of night-time photos. Five quantitative measures are further calculated in brightest area to access the highlight suppressing capability. Moreover, the texture related features are extracted in foreground and sharpest region to qualify the detail preserving ability. We also show that compared to commonly used SVR quality regressor, the random forest presents competitive results. Experiments demonstrated that the region selective approach in the feature extraction stage, with random forest in the quality regression stage, achieve superior result in the task of night-time photo quality assessment than state-of-the-art methods.

Funding: This work was supported by personal savings of the first author accumulated during his PhD in SJTU from 2017 to 2021 (I'd like to thank everyone who worked to make such opportunity available to me), and partly the National Science Foundation of China (62132006, 61831015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All source & test codes, scripts, with detailed test data & results are available at <https://github.com/hanzongxi/RSIQA>.

Conflicts of Interest: The authors declare no conflict of interests.

References

1. Nakazawa, K.; Yamamoto, J.; Mori, S.; Okamoto, S.; Shimizu, A.; Baba, K.; Fujii, N.; Uehara, M.; Hiramatsu, K.; Kumano, H.; et al. 3D Sequential Process Integration for CMOS Image Sensor. In Proceedings of the IEEE International Electron Devices Meeting (IEDM), 2021, pp. 30.4.1–30.4.4. 663
2. Yun, J.; Lee, S.; Cha, S.; et.al. A $0.6\text{ }\mu\text{m}$ Small Pixel for High Resolution CMOS Image Sensor with Full Well Capacity of 10,000e- by Dual Vertical Transfer Gate Technology. In Proceedings of the 2022 VLSI Technology and Circuits, 2022, pp. 351–352. 664
3. Reinhard, E.; Heidrich, W.; Debevec, P.; Pattanaik, S.; Ward, G.; Myszkowski, K. *High dynamic range imaging: acquisition, display, and image-based lighting*; Morgan Kaufmann, 2010. 665
4. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*; Pearson Prentice Hall, New Jersey, 2010. 670
5. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* **1987**, *39*, 355–368. 671
6. Ibrahim, H.; Kong, N.S.P. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics* **2007**, *53*, 1752–1758. 673
7. Arici, T.; Dikbas, S.; Altunbasak, Y. A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on image processing* **2009**, *18*, 1921–1935. 675
8. Wu, X. A linear programming approach for optimal contrast-tone mapping. *IEEE transactions on image processing* **2010**, *20*, 1262–1272. 677
9. Celik, T.; Tjahjadi, T. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing* **2011**, *20*, 3431–3441. 679
10. Lee, C.; Lee, C.; Kim, C.S. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Transactions on Image Processing* **2013**, *22*, 5372–5384. 680
11. Land, E.H. The retinex theory of color vision. *Scientific american* **1977**, *237*, 108–129. 682
12. Jobson, D.; Rahman, Z.; Woodell, G. Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing* **1997**, *6*, 451–462. 683
13. Jobson, D.; Rahman, Z.; Woodell, G. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing* **1997**, *6*, 965–976. 685
14. Petro, A.B.; Sbert, C.; Morel, J.M. Multiscale retinex. *Image Processing On Line* **2014**, pp. 71–88. 687
15. Kimmel, R.; Elad, M.; Shaked, D.; Keshet, R.; Sobel, I. A variational framework for retinex. *International Journal of computer vision* **2003**, *52*, 7–23. 688
16. Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016. 690
17. Fu, X.; Liao, Y.; Zeng, D.; Huang, Y.; Zhang, X.P.; Ding, X. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Transactions on Image Processing* **2015**, *24*, 4965–4977. 692
18. Li, M.; Liu, J.; Yang, W.; Guo, Z. Joint denoising and enhancement for low-light images via retinex model. In Proceedings of the International Forum on Digital TV and Wireless Multimedia Communications. Springer, 2017, pp. 91–99. 694
19. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing* **2018**, *27*, 2828–2841. 696
20. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing* **2013**, *22*, 3538–3548. 699
21. Yang, X.; Ling, W.; Lu, Z.; Ong, E.; Yao, S. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication* **2005**, *20*, 662–680. 700
22. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* **2016**, *26*, 982–993. 702
23. Xu, L.; Yan, Q.; Xia, Y.; Jia, J. Structure extraction from texture via relative total variation. *ACM transactions on graphics (TOG)* **2012**, *31*, 1–10. 704
24. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing* **2007**, *16*, 2080–2095. 706
25. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new low-light image enhancement algorithm using camera response model. In Proceedings of the IEEE international conference on computer vision workshops, 2017, pp. 3015–3022. 708
26. Ren, Y.; Ying, Z.; Li, T.H.; Li, G. LECARM: Low-Light Image Enhancement Using the Camera Response Model. *IEEE Transactions on Circuits and Systems for Video Technology* **2019**, *29*, 968–981. 710
27. Wang, S.; Luo, G. Naturalness Preserved Image Enhancement Using a Priori Multi-Layer Lightness Statistics. *IEEE Transactions on Image Processing* **2018**, *27*, 938–948. 712
28. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new image contrast enhancement algorithm using exposure fusion framework. In Proceedings of the Computer Analysis of Images and Patterns: 17th International Conference, CAIP 2017, Ystad, Sweden, August 22–24, 2017, Proceedings, Part II 17. Springer, 2017, pp. 36–46. 714
29. Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; Paisley, J. A fusion-based enhancing method for weakly illuminated images. *Signal Processing* **2016**, *129*, 82–96. 717
30. Burt, P.; Adelson, E. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications* **1983**, *31*, 532–540. 719

31. Shen, J.; Zhao, Y.; Yan, S.; Li, X. Exposure Fusion Using Boosting Laplacian Pyramid. *IEEE Transactions on Cybernetics* **2014**, *44*, 1579–1590. 720
721
32. Ancuti, C.O.; Ancuti, C.; Bekaert, P. Enhancing by saliency-guided decolorization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 257–264. 722
723
33. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* **2017**, *61*, 650–662. 724
725
34. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. In Proceedings of the British Machine Vision Conference. British Machine Vision Association, 2018. 726
727
35. Zhang, Y.; Zhang, J.; Guo, X. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM international conference on multimedia, 2019, pp. 1632–1640. 728
729
36. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2015, pp. 234–241. 730
731
37. Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In Proceedings of the BMVC, 2018, Vol. 220, p. 4. 732
733
38. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-To-Image Translation With Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017. 734
735
39. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing* **2021**, *30*, 2340–2349. 736
737
40. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the ECCV. Springer, 2016, pp. 694–711. 738
739
41. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *44*, 4225–4238. 740
741
42. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-Inspired Unrolling With Cooperative Prior Architecture Search for Low-Light Image Enhancement. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 10561–10570. 742
743
744
43. Chen, B.; Zhu, L.; Zhu, H.; Yang, W.; Lu, F.; Wang, S. The loop game: quality assessment and optimization for low-light image enhancement. *arXiv preprint arXiv:2202.09738* **2022**. 745
746
44. Al Sobhah, R.; Tekli, J. Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges. *Signal Processing: Image Communication* **2022**, p. 116848. 747
748
45. Liba, O.; Murthy, K.; Tsai, Y.T.; Brooks, T.; Xue, T.; Karnad, N.; He, Q.; Barron, J.T.; Sharlet, D.; Geiss, R.; et al. Handheld mobile photography in very low light. *ACM transactions on graphics (TOG)* **2019**, *38*, 164–1. 750
751
46. Monod, A.; Delon, J.; Veit, T. An Analysis and Implementation of the HDR+ Burst Denoising Method. *Image Processing On Line* **2021**, *11*, 142–169. 752
753
47. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems* **2014**, *27*. 754
755
48. Xiang, T.; Yang, Y.; Guo, S. Blind night-time image quality assessment: Subjective and objective approaches. *IEEE Transactions on Multimedia* **2019**, *22*, 1259–1272. 756
757
49. Han, Z.; Liu, Y.; Xie, R. A large-scale image database for benchmarking mobile camera quality and NR-IQA algorithms. *Displays* **2023**, p. 102366. 758
759
50. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a Completely Blind Image Quality Analyzer. *IEEE Signal Processing Letters* **2013**, *20*, 209–212. 760
761
51. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing* **2015**, *24*, 2579–2591. 762
763
52. Bosse, S.; Maniry, D.; Müller, K.R.; Wiegand, T.; Samek, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing* **2017**, *27*, 206–219. 764
765
53. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blblindly assess image quality in the wild guided by a self-adaptive hyper network. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3667–3676. 766
767
768
54. Sun, W.; Min, X.; Tu, D.; Ma, S.; Zhai, G. Blind quality assessment for in-the-wild images via hierarchical feature fusion and iterative mixed database training. *IEEE Journal of Selected Topics in Signal Processing* **2023**. 769
770
55. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE transactions on image processing* **2012**, *21*, 4695–4708. 771
772
56. Liu, Y.; Gu, K.; Zhang, Y.; Li, X.; Zhai, G.; Zhao, D.; Gao, W. Unsupervised Blind Image Quality Evaluation via Statistical Measurements of Structure, Naturalness, and Perception. *IEEE Transactions on Circuits and Systems for Video Technology* **2020**, *30*, 929–943. 773
774
57. Zhai, G.; Wu, X.; Yang, X.; Lin, W.; Zhang, W. A psychovisual quality metric in free-energy principle. *IEEE Transactions on Image Processing* **2012**, *21*, 41–52. 775
776
777

58. Liu, Y.; Gu, K.; Li, X.; Zhang, Y. Blind image quality assessment by natural scene statistics and perceptual characteristics. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* **2020**, *16*, 1–91. 778
779
59. Liu, Y.; Zhang, B.; Hu, R.; Gu, K.; Zhai, G.; Dong, J. Underwater Image Quality Assessment: Benchmark Database and Objective Method. *IEEE Transactions on Multimedia* **2024**, *26*, 7734–7747. 780
781
60. Liu, Y.; Gu, K.; Cao, J.; Wang, S.; Zhai, G.; Dong, J.; Kwong, S. UIQI: A Comprehensive Quality Evaluation Index for Underwater Images. *IEEE Transactions on Multimedia* **2024**, *26*, 2560–2573. 782
783
61. Xue, W.; Mou, X.; Zhang, L.; Bovik, A.C.; Feng, X. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Transactions on Image Processing* **2014**, *23*, 4850–4862. 784
785
62. Ruderman, D.L. The statistics of natural images. *Network: computation in neural systems* **1994**, *5*, 517. 786
63. Min, X.; Gu, K.; Zhai, G.; Liu, J.; Yang, X.; Chen, C.W. Blind Quality Assessment Based on Pseudo-Reference Image. *IEEE Transactions on Multimedia* **2018**, *20*, 2049–2062. 787
788
64. Min, X.; Zhai, G.; Gu, K.; Liu, Y.; Yang, X. Blind Image Quality Estimation via Distortion Aggravation. *IEEE Transactions on Broadcasting* **2018**, *64*, 508–517. 789
790
65. Zhang, W.; Ma, K.; Yan, J.; Deng, D.; Wang, Z. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology* **2018**, *30*, 36–47. 791
792
66. Lin, T.Y.; RoyChowdhury, A.; Maji, S. Bilinear CNN Models for Fine-Grained Visual Recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), December 2015. 793
794
67. Fang, Y.; Ma, K.; Wang, Z.; Lin, W.; Fang, Z.; Zhai, G. No-reference quality assessment of contrast-distorted images based on natural scene statistics. *IEEE Signal Processing Letters* **2014**, *22*, 838–842. 795
796
68. Gu, K.; Lin, W.; Zhai, G.; Yang, X.; Zhang, W.; Chen, C.W. No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization. *IEEE Trans. Cybernetics* **2016**, *PP*, 1–7. 797
798
69. Gu, K.; Tao, D.; Qiao, J.F.; Lin, W. Learning a no-reference quality assessment model of enhanced images with big data. *IEEE transactions on neural networks and learning systems* **2017**, *29*, 1301–1313. 799
800
70. Wang, S.; Ma, K.; Yeganeh, H.; Wang, Z.; Lin, W. A patch-structure representation method for quality assessment of contrast changed images. *IEEE Signal Processing Letters* **2015**, *22*, 2387–2390. 801
802
71. Zhai, G.; Sun, W.; Min, X.; Zhou, J. Perceptual quality assessment of low-light image enhancement. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* **2021**, *17*, 1–24. 803
804
72. Lin, W.; Wu, Y.; Xu, L.; Chen, W.; Zhao, T.; Wei, H. No-reference quality assessment for low-light image enhancement: Subjective and objective methods. *Displays* **2023**, *78*, 102432. 805
806
73. Zhang, Z.; Sun, W.; Min, X.; Zhu, W.; Wang, T.; Lu, W.; Zhai, G. A No-Reference Evaluation Metric for Low-Light Image Enhancement. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021, pp. 1–6. 807
808
74. Yang, Y.; Xiang, T.; Guo, S.; Lv, X.; Liu, H.; Liao, X. EHNQ: Subjective and objective quality evaluation of enhanced night-time images. *IEEE Transactions on Circuits and Systems for Video Technology* **2023**. 809
810
75. Hu, R.; Liu, Y.; Wang, Z.; Li, X. Blind quality assessment of night-time image. *Displays* **2021**, *69*, 102045. 811
76. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360* **2016**. 812
813
77. Jiang, Q.; Xu, J.; Mao, Y.; Zhou, W.; Min, X.; Zhai, G. Deep decomposition and bilinear pooling network for blind night-time image quality evaluation. *arXiv preprint arXiv:2205.05880* **2022**. 814
815
78. Sheikh, H.R.; Wang, Z.; Cormack, L.; Bovik, A.C. LIVE image quality assessment database release 2, 2006. 816
79. Silva, V.D.; Chesnokov, V.; Larkin, D. A novel adaptive shading correction algorithm for camera systems. *Electronic Imaging* **2016**, *2016*, 1–5. 817
818
80. Hassen, R.; Wang, Z.; Salama, M.M. Image sharpness assessment based on local phase coherence. *IEEE Transactions on Image Processing* **2013**, *22*, 2798–2810. 819
820
81. Lu, Q.; Zhai, G.; Zhu, W.; Zhu, Y.; Min, X.; Zhang, X.P.; Yang, H. Automatic region selection for objective sharpness assessment of mobile device photos. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020, pp. 106–110. 821
822
823
82. Vu, C.T.; Phan, T.D.; Chandler, D.M. S3: A spectral and spatial measure of local perceived sharpness in natural images. *IEEE Transactions on Image Processing* **2012**, *21*, 934–945. 824
825
83. Han, Z.; Liu, Y.; Xie, R.; Zhai, G. Image Quality Assessment for Realistic Zoom Photos. *Sensors* **2023**, *23*, 4724. 826
84. Zoran, D.; Weiss, Y. Scale invariance and noise in natural images. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009, pp. 2209–2216. 827
828
85. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* **1973**, pp. 610–621. 829
830
86. Raut, M.A.; Patil, M.M.A.; Dhondrikar, M.C.P.; Kamble, M.S.D. Texture parameters extraction of satellite image. *IJSTE-International Journal of Science Technology & Engineering* **2016**, *2*. 831
832
87. Pei, S.C.; Chen, L.H. Image quality assessment using human visual DOG model fused with random forest. *IEEE Transactions on Image Processing* **2015**, *24*, 3282–3292. 833
834
88. Liu, Y.; Gu, K.; Wang, S.; Zhao, D.; Gao, W. Blind quality assessment of camera images based on low-level and high-level statistical features. *IEEE Transactions on Multimedia* **2018**, *21*, 135–146. 835
836

-
89. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The elements of statistical learning: data mining, inference, and prediction*; Vol. 2, Springer, 2009. 837
90. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32. 838
91. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. In Proceedings of the ICML, 1996, Vol. 96, pp. 148–156. 839
92. Schapire, R.E.; Singer, Y. Improved boosting algorithms using confidence-rated predictions. In Proceedings of the eleventh annual conference on Computational learning theory, 1998, pp. 80–91. 840
93. Drucker, H. Improving regressors using boosting techniques. In Proceedings of the ICML, 1997, Vol. 97, pp. 107–115. 841
94. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-class adaboost. *Statistics and its Interface* **2009**, *2*, 349–360. 842
95. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *Journal of machine Learning research* **2011**, *12*, 2825–2830. 843
96. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Boca Raton, FL: CRC Press, 1984. 844